BILINGUAL PUBLISHING GROUP

Journal of Computer Science Research

Volume 5 | Issue 4 | October 2023 | ISSN 2630-5151（Online）

Volume 5 Issue 4 • October 2023 • ISSN 2630-5151 (Online)

Journal of Computer Science Research

BILINGUAL PUBLISHING GROUP

BILINGUAL PUBLISHING GROUP

Volume 5 | Issue 4 | October 2023 | Page1-57

Journal of Computer Science Research

Contents

Articles

Review

Short Communication

BILINGUAL PUBLISHING GROUP

Journal of Computer Science Research

https://journals.bilpubgroup.com/index.php/jcsr

ARTICLE

Play by Design: Developing Artificial Intelligence Literacy through Game-based Learning

***Xiaoxue Du*[1*], *Xi Wang*[2]**

[1] *MIT Media Lab, MIT, Cambridge, MA, 02139, USA*

[2] *Columbia University, New York, 10025, USA*

ABSTRACT

The paper proposes an innovative approach aimed at fostering AI literacy through interactive gaming experiences. This paper designs a game-based prototype for preparing pre-service teachers to innovate teaching practices across disciplines. The simulation, Color Conquest, serves as a strategic game to encourage educators to reconsider their pedagogical practices. It allows teachers to use and develop various scenarios by customizing maps, giving students agency to engage in the complex decision-making process. Additionally, this engagement process provides teachers with an opportunity to develop students' skills in artificial intelligence literacy as students actively develop strategic thinking, problem-solving, and critical reasoning skills.

Keywords: Game-based learning; Game-based assessment; Artificial intelligence literacy; Design thinking; Computational thinking; Teacher education

1. Introduction

Understanding AI is becoming essential in today's educational landscape. It equips students with knowledge to engage with the technology that is fundamentally reshaping our world. This comprehension empowers students to make informed decisions about how they interact with AI systems, both in their personal lives and future careers. It is critical to invite students to responsibly use AI and develop their abilities to apply what they have learned in solving authentic real-world challenges [1]. Moreover, a foundational understanding of AI fosters not only computational skills but also pivotal capabilities in

*CORRESPONDING AUTHOR:
Xiaoxue Du, MIT Media Lab, MIT, Cambridge, MA, 02139, USA; Email: xiaoxued@media.mit.edu

ARTICLE INFO
Received: 7 October2023 | Revised: 30 October 2023 | Accepted: 31 October 2023 | Published Online: 10 November 2023
DOI: https://doi.org/10.30564/jcsr.v5i4.5999

CITATION
Du, X.X., Wang, X., 2023. Play by Design: Developing Artificial Intelligence Literacy through Game-based Learning. Journal of Computer Science Research. 5(4): 1-12. DOI: https://doi.org/10.30564/jcsr.v5i4.5999

critical thinking and problem-solving, proving integral in navigating the increasingly dynamic technological landscape [2]. By integrating AI education into curricula, educational institutions should prepare students to be active participants in a future where AI is likely to play an even more significant role across various industries and aspects of society [3]. This ensures that they are not only consumers of AI-driven products and services but also informed contributors to the development and ethical implementation of these technologies.

At the same time, gamification provides an interactive learning experience in building the educational landscape. It not only imparts knowledge but also hones decision-making in an increasingly complex and AI-driven world. This innovative approach prepares students to not only be well-informed users of AI but also positions them as potential innovators and contributors in the field of artificial intelligence. This also allows teachers to innovative pedagogy in daily classroom teaching.

The paper proposes an innovative approach aimed at fostering AI literacy through interactive gaming experiences. By integrating AI concepts into multi-player games, participants are not only entertained but also empowered to grasp the fundamentals of AI in a practical and engaging manner. Through dynamic game-play scenarios, users navigate complex AI systems, make strategic decisions, and witness the impact of their choices. This hands-on approach not only demystifies AI but also cultivates a deeper appreciation for its potential and ethical considerations. Through Color Conquest, pre-service teachers could further develop pedagogical practice in classroom teaching.

2. Literature review

Game-based learning offers an innovative approach by infusing interactive and immersive experiences into traditional classrooms [4]. This approach captures students' attention and sustains their interest in learning within diverse simulations [5]. Through games, learners become active participants, making decisions, solving problems, and exploring complex scenarios, which leads to deeper comprehension and retention of concepts. Moreover, the experiential nature of game-based learning allows students to witness the practical implications of their studies, bridging the gap between abstract theory and real-world application [6]. Immediate feedback mechanisms and adaptive technologies provide tailored learning experiences, ensuring that students receive content at a pace aligned with their individual proficiency levels [7]. This fosters autonomy and self-directed learning, promoting a growth mindset and resilience in the face of challenges [8]. Collaborative elements within games encourage teamwork, communication, and a sense of community, enriching the learning environment. Additionally, games offer a safe space for experimentation and failure, instilling a culture of curiosity and exploration [9].

Studies have shown that using game-based learning for immersion learning enhances long-term knowledge retention, affirming its efficacy as a powerful educational tool [10]. Overall, game-based learning not only transforms education into a dynamic and engaging experience but also equips learners with critical thinking skills and a passion for lifelong learning [11].

In the field of teacher education, game-based learning provides an innovative approach as teachers could use diverse technology in classroom teaching [12]. Through interactive simulations and virtual environments, teachers could practice and refine their instructional techniques in a risk-free setting against dangerous settings in the real world [13,14]. This approach allows them to navigate various classroom scenarios, adapt to diverse student needs, and implement effective teaching strategies. By actively participating in these immersive experiences, educators develop an understanding of the complexities and nuances of teaching. Moreover, the dynamic nature of game-based learning encourages them to critically reflect on their teaching practices and make informed decisions in real time [15]. The incorporation of immediate feedback mechanisms ensures that they receive constructive input, enabling them to adjust and refine their approaches [16]. This iterative process

cultivates a growth mindset, preparing educators to adapt and innovate in the face of evolving educational landscapes. Additionally, collaborative elements within educational games simulate the teamwork and communication skills necessary for effective teaching [17]. This experiential approach not only makes the learning process more engaging and memorable but also instils in educators a deep sense of empathy and understanding for their future students [18].

Game-based learning for developing AI literacy is a growing field at the intersection of education and artificial intelligence [19]. Scholars have recognized games as immersive platforms to introduce complex AI concepts in an engaging and interactive manner. This approach capitalizes on the inherent appeal of games, which can explain abstract notions of algorithms, machine learning, and neural networks into tangible, hands-on experiences. Additionally, recent advances in technology have facilitated the creation of AI-driven educational games that simulate real-world scenarios, providing learners with a dynamic environment to experiment with AI algorithms and understand their implications in various contexts [20]. Moreover, the gamification of AI literacy serves to democratize access to this critical domain of knowledge. It allows learners of diverse backgrounds and ages to engage with complex AI concepts (e.g., decision trees) in a non-intimidating and inclusive manner. Several researchers have emphasized how multi-player environments in educational games can foster collaborative problem-solving and knowledge sharing, creating a community of learners dedicated to AI literacy. This communal aspect not only enhances comprehension but also nurtures a supportive learning environment where individuals can collectively grapple with the complexities of AI. As the demand for AI skills continues to grow across various industries, leveraging game-based learning approaches offers a promising avenue to equip a wider demographic with the foundational knowledge and critical thinking abilities necessary to navigate the evolving landscape of artificial intelligence [21].

3. Theoretical framework

Design thinking is a creative problem-solving approach that places a strong emphasis on empathy, ideation, and iterative development [22]. When applied to the realm of AI literacy, it introduces a structured framework that involves empathizing with end-users, defining problems, generating creative solutions, prototyping those solutions, and rigorously testing them. It gives students agency to engage in real-world problems and proposes solutions to solve them through a systematic process. Computational thinking is a problem-solving approach rooted in principles of computer science, aimed at dissecting intricate issues into more manageable components [23]. The benefits of integrating computational thinking into educational curricula, emphasise its role in fostering critical thinking and problem-solving skills among students [24].

The Five Big AI Ideas serve as a cornerstone in the understanding of artificial intelligence. These concepts encompass critical aspects of AI development and application [25]. The first idea, Perception, delves into enabling machines to comprehend and interpret the world, employing techniques like computer vision and natural language processing. Representation and reasoning involve instructing machines to not only hold knowledge but also make informed decisions based on that knowledge, often utilizing methods such as knowledge graphs and symbolic reasoning. Learning, the third idea, encompasses various machine learning techniques like supervised, unsupervised, and reinforcement learning, allowing machines to enhance their performance over time. Natural interaction explores the fusion of AI with physical systems, enabling machines to interact with the physical world. The fifth idea, social impact, highlights the critical demands of using AI to solve real-world problems.

In conjunction with these ideas, the AI Literacy Framework highlights the critical need to use AI responsively. It emphasizes the need for individuals to understand not only the capabilities but also the limitations and potential implications of AI. This awareness empowers people to make informed decisions

and take appropriate actions when interacting with AI systems or developing AI-based solutions. Incorporating responsible AI practices into the framework, ensures that AI is harnessed for the betterment of society while mitigating potential risks and ethical dilemmas. This approach aligns with the broader goal of promoting AI literacy as a means to navigate the evolving landscape of artificial intelligence effectively.

4. Game-based design

Design thinking is demonstrated in this game through several key aspects. Firstly, empathy plays a crucial role in the game-play as players engage in simultaneous decision-making, necessitating an understanding of their opponent's potential choices. This cultivates an empathetic perspective, a fundamental component of design thinking. Moreover, the game encourages ideation and iteration. Players are prompted to think strategically in both phases, requiring them to devise plans and adapt them in response to the changing dynamics of the board. This iterative process mirrors the cyclical nature of design thinking, emphasizing the importance of refinement and improvement. Additionally, the game adheres to a user-centred approach by preventing redundancy in choices. This design feature ensures that players are engaged in meaningful decision-making, contributing to their overarching strategic objectives. By incorporating these elements, the game embodies the principles of design thinking, promoting creative problem-solving and user-centric design.

Computational thinking principles are embedded within the mechanics of this game. Firstly, the concept of decomposition is illustrated as players systematically dissect the multifaceted task of territory claiming and connection into manageable, step-by-step actions. This mirrors the essence of computational thinking, which involves breaking down complex problems into smaller, more manageable components. Additionally, the game prompts players to engage in abstraction, compelling them to contemplate the conceptual notions of territory ownership and connection, which are visually represented on the board. Furthermore, players are challenged to think algorithmically, devising strategies for claiming territories and establishing connections. These strategies can be likened to algorithms, embodying computational thinking in practice.

Furthermore, the game encourages collaboration and community building through critical thinking, strategic planning, and adaptability, all of which are essential skills that align with the goals of AI literacy. Students should discuss their own strategies and the opponent's strategies. Students are encouraged to find the algorithm for gaining the highest score and assume the possibility of winning, and think in terms of both spatial reasoning and decision-making, which are pertinent to understanding AI concepts like perception and representation & reasoning.

5. Game-based mechanics

In the initial phase, Player 1 (Blue) and Player 2 (Green) are presented with the simultaneous task of strategically choosing an uncoloured area on the board (see Appendix). This decision-making process initiates the territorial claim, as chosen areas transition to the respective player's colour. However, in the event that both players opt for the same area, a strategic twist is introduced. The area is promptly marked in red, denoting its inaccessibility for subsequent selections (see **Figures 1 and 2**). This strategic mechanic prevents redundancy in choices, encouraging players to analyze their decisions with precision and foresight. The phase's iterative nature, defined by the parameter N set by the map creator, grants players the opportunity to strategically claim territories over multiple rounds, nurturing a dynamic environment that calls for adaptability and long-term planning(see **Figure 3**).

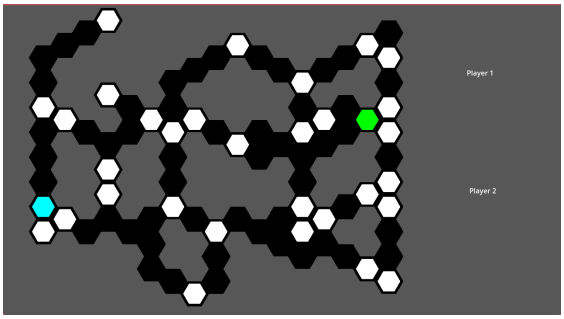


Figure 1. After two players simultaneously claim an uncoloured area on the board.

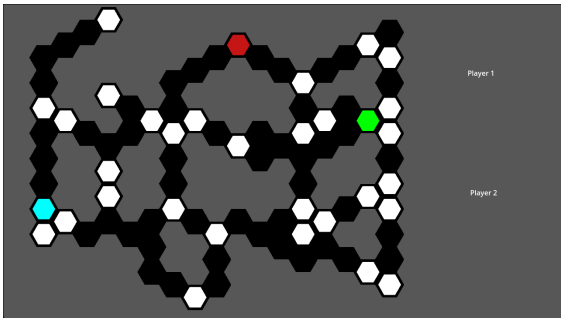


Figure 2. Once two players claim the same area.

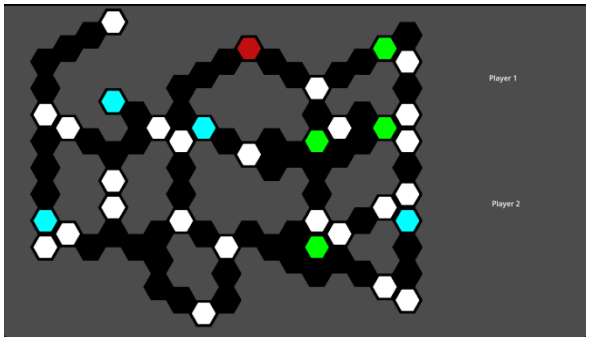


Figure 3. After N = 5 times claim area.

Transitioning into the second phase, players face the challenge of connecting two areas of their own colour with lines (see **Figure 4**). This phase introduces additional layers of complexity, as players must navigate through the board, factoring in the accessibility of black and white areas and the impassibility of red zones. Players must consider the implications of their routes while concealing their intentions from the opposing player. Once both players have successfully established their connections, the game proceeds to the scoring phase, a pivotal moment in determining the victor. The scoring system rewards strategic prowess and penalizes hasty decision-making. If the lines of the two players do not intersect, each player garners points commensurate with the length of their respective line (see **Figure 5**). Conversely, in the event of an intersection, players are prompted to strategically evaluate the length of their opponent's line (see **Figures 6 and 7**).

Furthermore, the visual cues of red markings serve as a feedback mechanism. The areas connected by players in this phase, as well as any intersection points, are distinctly marked in red. This visual representation reinforces the strategic implications of their choices.

This iterative game-play process persists until players are no longer able to connect further lines, marking the conclusion of the game. The player who emerges with the higher score attains victory, showcasing superior strategic acumen and territorial control. This culminating moment emphasizes the importance of thoughtful decision-making.

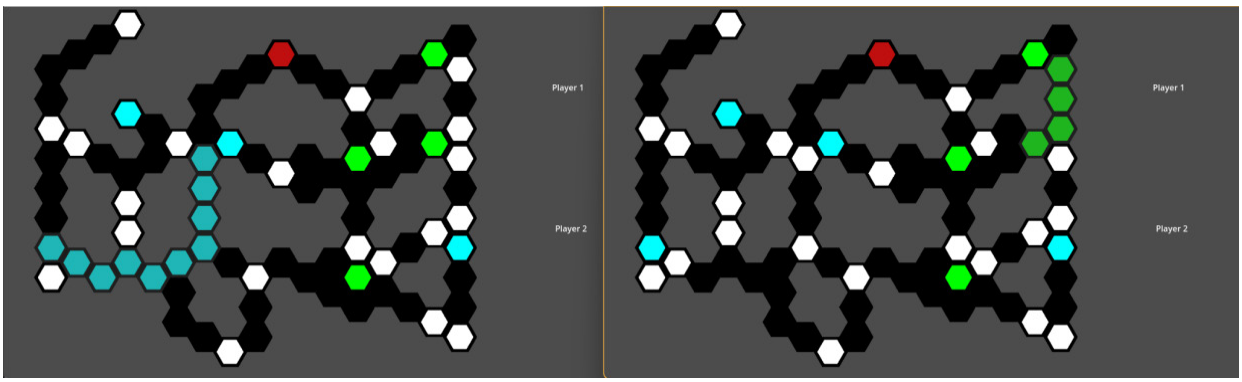


Figure 4. Two players try to connect their own colour area.

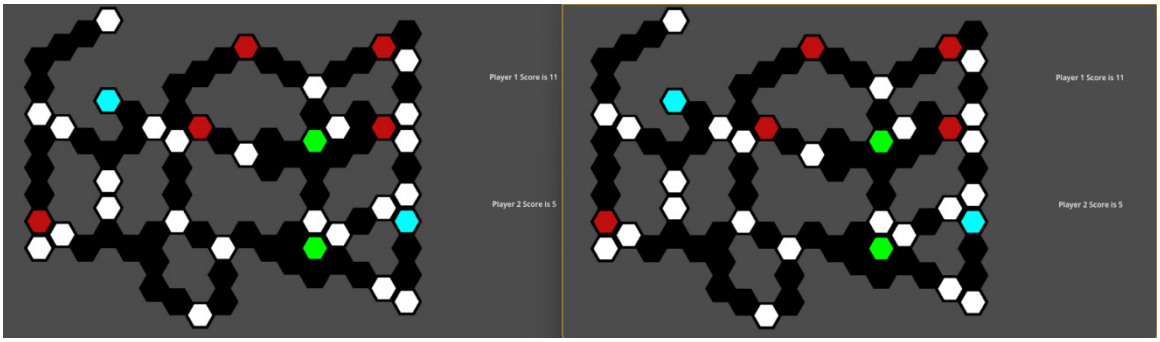


Figure 5. Players receive a score equal to their connecting range.

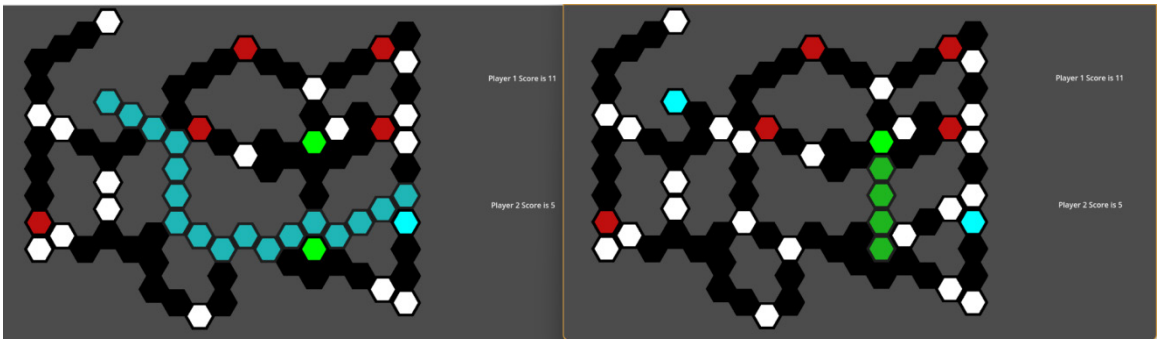


Figure 6. Players try to connect their own colour area again.

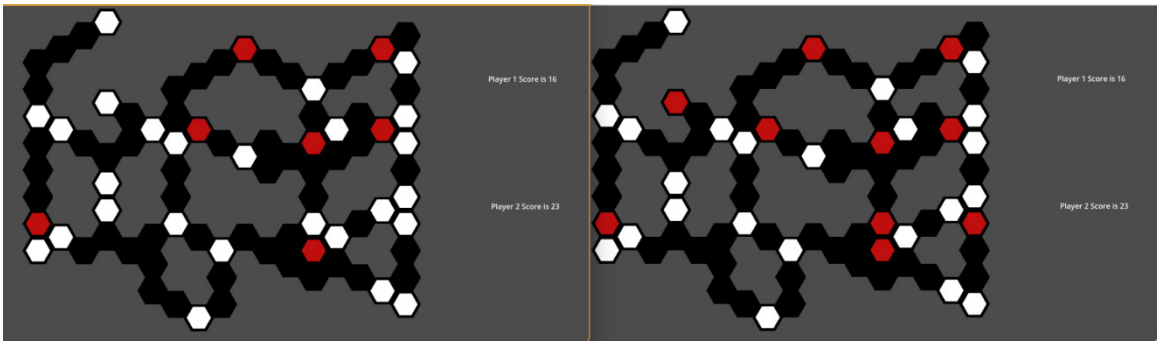


Figure 7. Players receive a score equal to their opponent's connecting range.

5.1 Customizable map design

One of the key innovations in the game is the integration of a map design feature, allowing players to take an active role in shaping their gaming experience. This feature empowers players with the ability to design their own maps, complete with uncolored areas, pathways, and potential obstacles. By providing players with creative agency, the game not only enhances engagement but also fosters strategic thinking. This customization element introduces a layer of personal investment, as players can craft environments that cater to their individual play styles and preferences (see **Figures 8-11**).

A science teacher could use the game's map design feature as a dynamic educational tool in the classroom. This innovative aspect enables students to delve into scientific concepts through interactive map creation. For instance, students might design maps representing ecosystems, complete with various biomes, habitats, and species. This hands-on activity encourages critical thinking as they strategically plan the layout and relationships within the ecosystem. Through this exercise, students gain a deeper understanding of ecological interactions, spatial relationships, and the interdependence of organisms within an ecosystem. This approach not only

makes science concepts tangible but also cultivates analytical thinking skills, providing a profound and memorable learning experience.

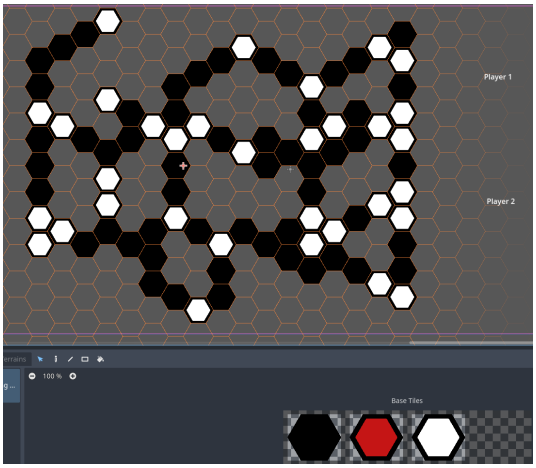


Figure 8. Customizable map design.

5.2 Community-driven interaction

The interactive aspect of map sharing amplifies the social dimension of game-play. Players can share their creations with their peers or the wider gaming community, creating an exchange of ideas and challenges. This not only builds community among players but also generates a diverse array of maps, each presenting unique challenges and strategic opportunities. Through this collective effort, the game's complexity expands exponentially, ensuring that game-play remains dynamic and engaging over time across disciplines.

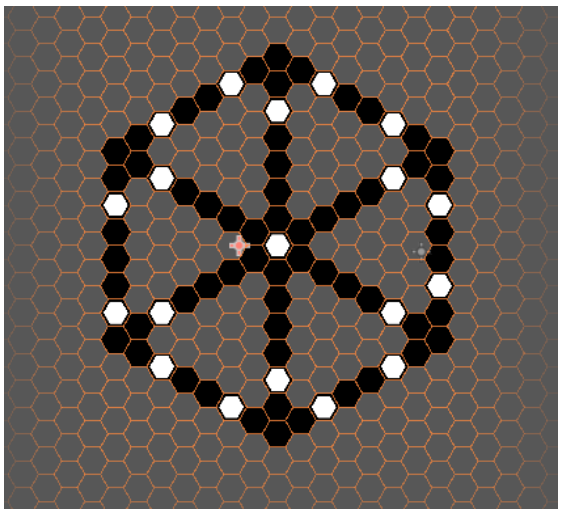


Figure 9. The idea of segregation in geometry.

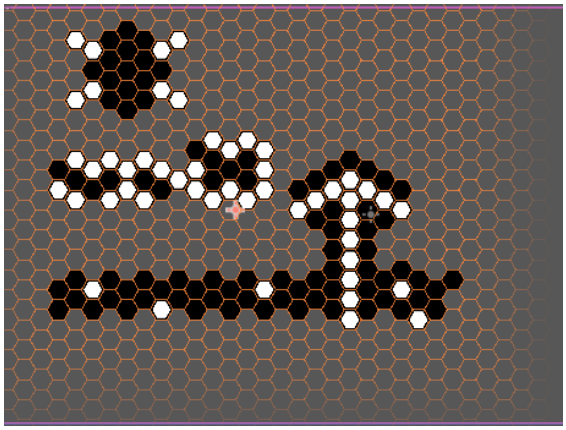


Figure 10. Analog ecosystem.

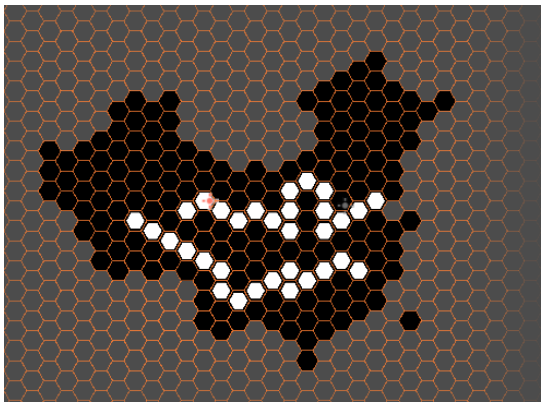


Figure 11. Analog China region with two main rivers.

5.3 Promoting creativity and collaboration

The game's map design feature encourages a culture of creativity and collaboration. Players are not only consumers of content but active creators within the game's ecosystem. This collaborative approach not only strengthens teamwork skills but also leads to the emergence of challenging maps that require collective problem-solving to conquer. Additionally, the ability to rate and provide feedback on user-generated maps helps identify high-quality designs and motivates creators to refine their work, further enriching the overall gaming experience across disciplines in classroom teaching.

6. Methods

The study involved a total of 20 pre-service teachers enrolled in a teacher education program. The primary material used in this study was the ed-

ucational game Color Conquest, designed to teach AI concepts in an interactive and gamified manner. The game incorporated elements of strategy, problem-solving, and critical thinking, all aimed at enhancing AI literacy skills. Participants completed a post survey to gather feedback on their experience with the game. The survey included Likert-scale items and open-ended questions to assess engagement, perceived learning, and overall satisfaction with the game. Quantitative data from the post-assessments was analyzed to understand the overall satisfaction rate (see **Table 1**). Qualitative data from the follow-up survey were analyzed thematically to extract key insights and feedback from participants.

7. Results

Responses from the follow-up survey provided valuable insights into participant perceptions of the Color Conquest game. Qualitative feedback highlighted the game's effectiveness in making complex AI concepts more accessible and enjoyable. A substantial majority of participants (87%) reported high levels of engagement with the "Color Conquest" game. Many expressed enthusiasm for the interactive nature of the game, noting that it provided a dynamic and immersive learning experience. A notable 92% of participants indicated that they perceived a significant increase in their understanding of AI concepts after engaging with the game. Respondents highlighted the game's ability to simplify complex topics and provide practical applications for theoretical knowledge. A significant proportion (85%) of participants expressed overall satisfaction with the "Color Conquest" game as an educational tool. Comments emphasized the game's effectiveness in making AI concepts more accessible, as well as its role in fostering a collaborative learning environment.

Several participants praised the game's feature that allowed them to customize scenarios. They appreciated having agency in their learning journey, as it enabled them to explore specific AI concepts

Table 1. Overall satisfaction questionnaire.

No.	Item
1	On a scale of 1 to 5, how would you rate your level of engagement with the "Color Conquest" game? (1 = Not Engaged at All, 5 = Highly Engaged)
2	How satisfied were you with your overall experience using the "Color Conquest" game as an educational tool? (1 = Very Dissatisfied, 5 = Very Satisfied)
3	To what extent did you feel that playing the "Color Conquest" game helped you learn and understand AI concepts?(1 = Very unhelpful, 5 = Very helpful)
4	Please rate the extent to which you feel your knowledge of AI concepts improved after playing the "Color Conquest" game. (1 = Not Improved, 5 = Significant Improvement)
5	Were you able to customize scenarios in the "Color Conquest" game to explore specific AI concepts of interest to you? (Yes / No)
6	To what extent did you feel that customized scenarios in the game helped you learn AI concepts? (1 = Very unhelpful, 5 = Very helpful)
7	How confident are you in applying the AI concepts you learned from the "Color Conquest" game in your future teaching practices in classroom teaching? (1 = Not Confident at All, 5 = Very Confident)
8	Do you believe that the "Color Conquest" game has positively influenced your long-term understanding and application of AI concepts? (Yes / No)
9	What specific aspects of the "Color Conquest" game did you find most effective in helping you grasp AI concepts?
10	Were there any challenges or areas of confusion you encountered while using the "Color Conquest" game for learning? Please describe.
11	How do you think the "Color Conquest" game could be further enhanced to provide a more engaging and educational experience for learners?
12	In what ways do you envision incorporating the knowledge gained from the "Color Conquest" game into your teaching or other professional endeavours?

of interest in greater depth. A small subset of participants suggested minor enhancements, such as incorporating additional levels or challenges to further reinforce specific AI principles. These suggestions were largely centered on expanding the depth and complexity of the game-play. A noteworthy proportion of participants expressed optimism regarding the long-term impact of the game on their AI literacy. They believed that the hands-on, gamified approach provided a solid foundation for continued learning and application in their future teaching endeavours.

8. Conclusions and future work

In conclusion, this paper presents an innovative approach to enhance AI literacy through immersive gaming. By introducing a game-based prototype, this approach provides pre-service teachers with the tools to rethink teaching strategies across diverse subjects [26]. The simulation, Color Conquest, serves as a catalyst for educators to re-evaluate their pedagogical approaches, encouraging them to create customized scenarios and giving students agency in their learning journey. Furthermore, this interactive process not only develops students' abilities in strategic thinking, problem-solving, and critical reasoning but also cultivates their proficiency in AI literacy in the 21st century.

Future work could explore the potential integration of AI-driven adaptive learning systems within gamified educational platforms to tailor content to meet individual student needs and learning styles. Also, future work could conduct assessments in evaluating students' cognitive, social and emotional learning outcomes as interacting with the games. The multivariate approach will contribute to the field of human-centered dynamic assessment in creating a student-learning environment that fosters not only academic skills but also emotional intelligence, teamwork, and problem-solving skills. This approach will advance the broader field of human-centered assessment, ultimately shaping a more dynamic and responsive learning environment for students.

Author Contributions

Both authors have made contributions to this article. Xiaoxue Du is designated as the corresponding author.

Conflict of Interest

There is no conflict of interest.

References

[1] Yang, Y.T.C., 2012. Building virtual cities, inspiring intelligent citizens: Digital games for developing students' problem solving and learning motivation. Computers & Education. 59(2), 365-377.

[2] Benvenuti, M., Cangelosi, A., Weinberger, A., et al., 2023. Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. Computers in Human Behavior. 148, 107903.

[3] Pedro, F., Subosa, M., Rivas, A., et al., 2019. Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development [Internet]. Available from: https://hdl.handle.net/20.500.12799/6533

[4] Welbers, K., Konijn, E.A., Burgers, C., et al., 2019. Gamification as a tool for engaging student learning: A field experiment with a gamified app. E-learning and Digital Media. 16(2), 92-109.

[5] Sánchez-Mena, A., Martí-Parreño, J., 2017. Drivers and barriers to adopting gamification: Teachers' perspectives. Electronic Journal of e-Learning. 15(5), 434-443.

[6] Wang, L.H., Chen, B., Hwang, G.J., et al., 2022. Effects of digital game-based STEM education on students' learning achievement: A meta-analysis. International Journal of STEM Education. 9(1), 1-13.

[7] Kalogiannakis, M., Papadakis, S., Zourmpakis, A.I., 2021. Gamification in science education. A systematic review of the literature. Education Sciences. 11(1), 22.

[8] Baca, T., Petrlik, M., Vrba, M., et al., 2021. The MRS UAV system: Pushing the frontiers of reproducible research, real-world deployment, and education with autonomous unmanned aerial vehicles. Journal of Intelligent & Robotic Systems. 102(1), 26.

[9] Zourmpakis, A.I., Papadakis, S., Kalogiannakis, M., 2022. Education of preschool and

elementary teachers on the use of adaptive gamification in science education. International Journal of Technology Enhanced Learning. 14(1), 1-16.

[10] Spiliotopoulos, D., Margaris, D., Vassilakis, C., et al. (editors), 2019. A mixed-reality interaction-driven game-based learning framework. Proceedings of the 11th International Conference on Management of Digital EcoSystems; 2019 Nov 12-14; Limassol, Cyprus. New York: Association for Computing Machinery. p. 229-236.

[11] Mao, W., Cui, Y., Chiu, M.M., et al., 2022. Effects of game-based learning on students' critical thinking: A meta-analysis. Journal of Educational Computing Research. 59(8), 1682-1708.

[12] Algayres, M., Triantafyllou, E., Werthmann, L., et al. (editors), 2021. Collaborative game design for learning: The challenges of adaptive game-based learning for the Flipped Classroom. Interactivity and Game Creation: 9th EAI International Conference, ArtsIT 2020; 2020 Dec 10-11; Aalborg, Denmark. p. 228-242.

[13] Nousiainen, T., Vesisenaho, M., Ahlstrom, E., et al. (editors), 2020. Gamifying teacher students' learning platform: Information and communication technology in teacher education courses. Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality; 2020 Oct 21-23; Salamanca, Spain. New York: Association for Computing Machinery. p. 688-693.

[14] Abdullah, N.M.A.F.N., Sharipuddin, A.H.A., Mustapha, S., et al. (editors), 2022. The development of driving simulator game-based learning in virtual reality. 2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA); 2022 May 12; Selangor, Malaysia. New York: IEEE. p. 325-328.

[15] Hasenbein, L., Stark, P., Trautwein, U., et al., 2022. Learning with simulated virtual classmates: Effects of social-related configurations on students' visual attention and learning experiences in an immersive virtual reality classroom. Computers in Human Behavior. 133, 107282.

[16] Toda, A.M., Klock, A.C., Oliveira, W., et al., 2019. Analysing gamification elements in educational environments using an existing Gamification taxonomy. Smart Learning Environments. 6(1), 1-14.

[17] Vidergor, H.E., 2021. Effects of digital escape room on gameful experience, collaboration, and motivation of elementary school students. Computers & Education. 166, 104156.

[18] Zourmpakis, A.I., Papadakis, S., Kalogiannakis, M., 2022. Education of preschool and elementary teachers on the use of adaptive gamification in science education. International Journal of Technology Enhanced Learning. 14(1), 1-16.

[19] Ng, D.T.K., Lee, M., Tan, R.J.Y., et al., 2023. A review of AI teaching and learning from 2000 to 2020. Education and Information Technologies. 28(7), 8445-8501.

[20] Du, X., Meier, E.B., 2023. Innovating pedagogical practices through professional development in computer science education. Journal of Computer Science Research. 5(3), 46-56.

[21] Skritsovali, K., 2023. Learning through playing: Appreciating the role of gamification in business management education during and after the COVID-19 pandemic. Journal of Management Development. 42(5), 388-398.

[22] Shé, C.N., Farrell, O., Brunton, J., et al., 2022. Integrating design thinking into instructional design: The# OpenTeach case study. Australasian Journal of Educational Technology. 38(1), 33-52.

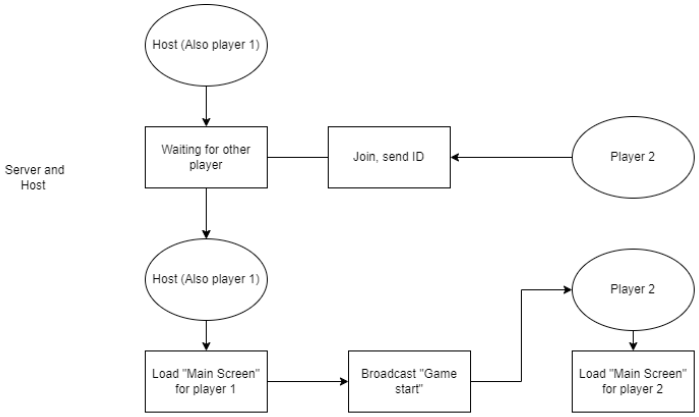
[23] Kafai, Y.B., Proctor, C., 2022. A revaluation of computational thinking in K-12 education: Moving toward computational literacies. Educational Researcher. 51(2), 146-151.

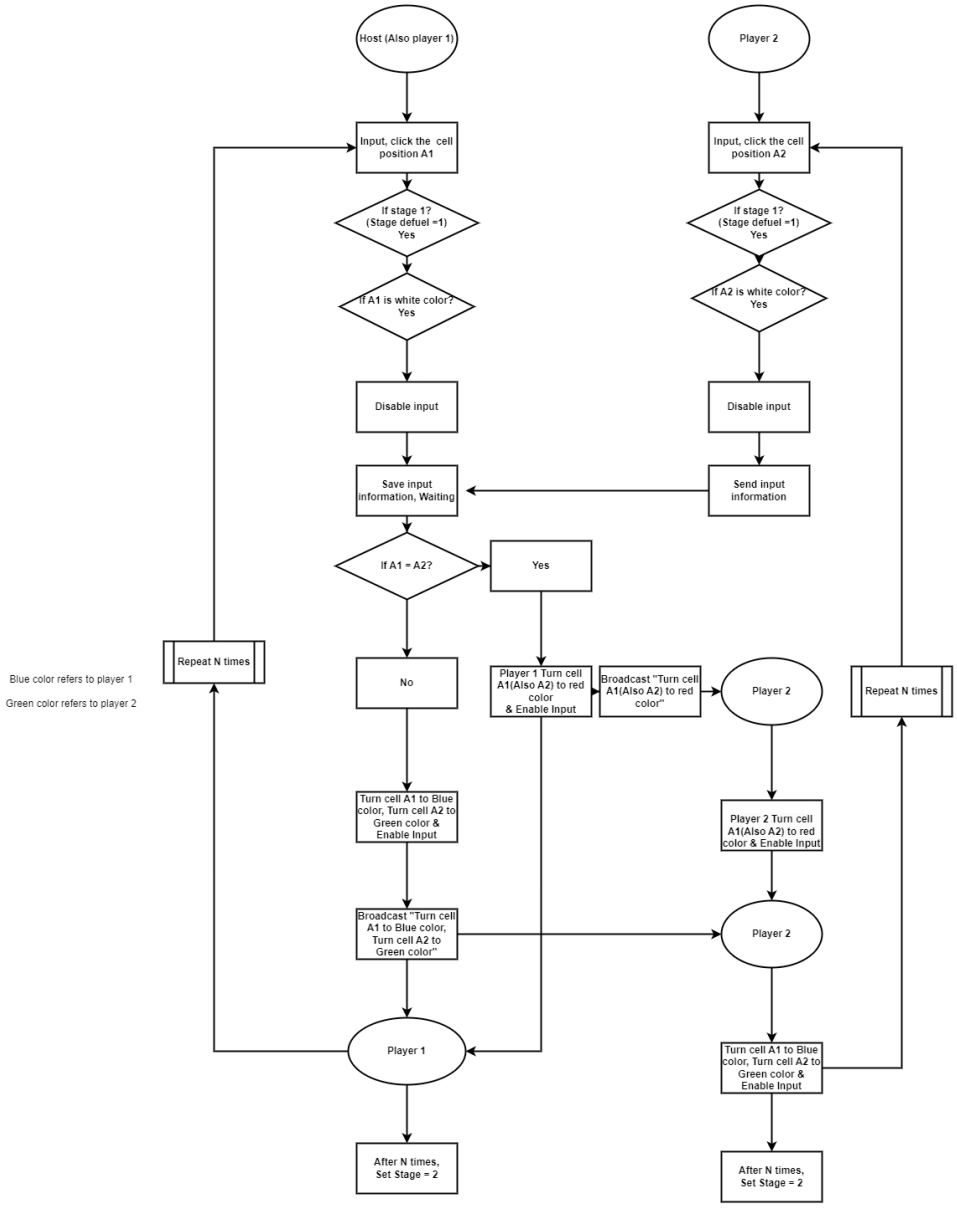
[24] Du, X., Taylor, M., Blumofe, N., et al. (editors), 2023. Widening the global access of artificial intelligence (AI) literacy curriculum through the participation of day of AI. Society for Information Technology & Teacher Education International Conference; 2023 Mar 13; New Orleans. Waynesville: Association for the Advancement of Computing in Education (AACE). p. 1896-1903.

[25] Hoffman, S.G., Joyce, K., Alegria, S., et al., 2022. Five big ideas about AI. Contexts. 21(3), 8-15.

[26] Lyublinskaya, I., Du, X., 2023. Annotated digital timelining: Interactive visual display for data analysis in mixed methods research. Methods in Psychology. 8, 100108.

Appendix


Server and Host
Host (Also player 1)
Waiting for other player
Join, send ID
Player 2
Host (Also player 1)
Player 2
Load "Main Screen" for player 1
Broadcast "Game start"
Load "Main Screen" for player 2


Host (Also player 1)
Player 2
Input, click the cell position A1
Input, click the cell position A2
If stage 1? (Stage defuel =1) Yes
If stage 1? (Stage defuel =1) Yes
If A1 is white color? Yes
If A2 is white color? Yes
Disable input
Disable input
Save input information, Waiting
Send input information
If A1 = A2?
Yes
Repeat N times
No
Player 1 Turn cell A1(Also A2) to red color & Enable Input
Broadcast "Turn cell A1(Also A2) to red color"
Player 2
Repeat N times
Blue color refers to player 1
Green color refers to player 2
Turn cell A1 to Blue color, Turn cell A2 to Green color & Enable Input
Player 2 Turn cell A1(Also A2) to red color & Enable Input
Broadcast "Turn cell A1 to Blue color, Turn cell A2 to Green color"
Player 2
Player 1
Turn cell A1 to Blue color, Turn cell A2 to Green color & Enable Input
After N times, Set Stage = 2
After N times, Set Stage = 2

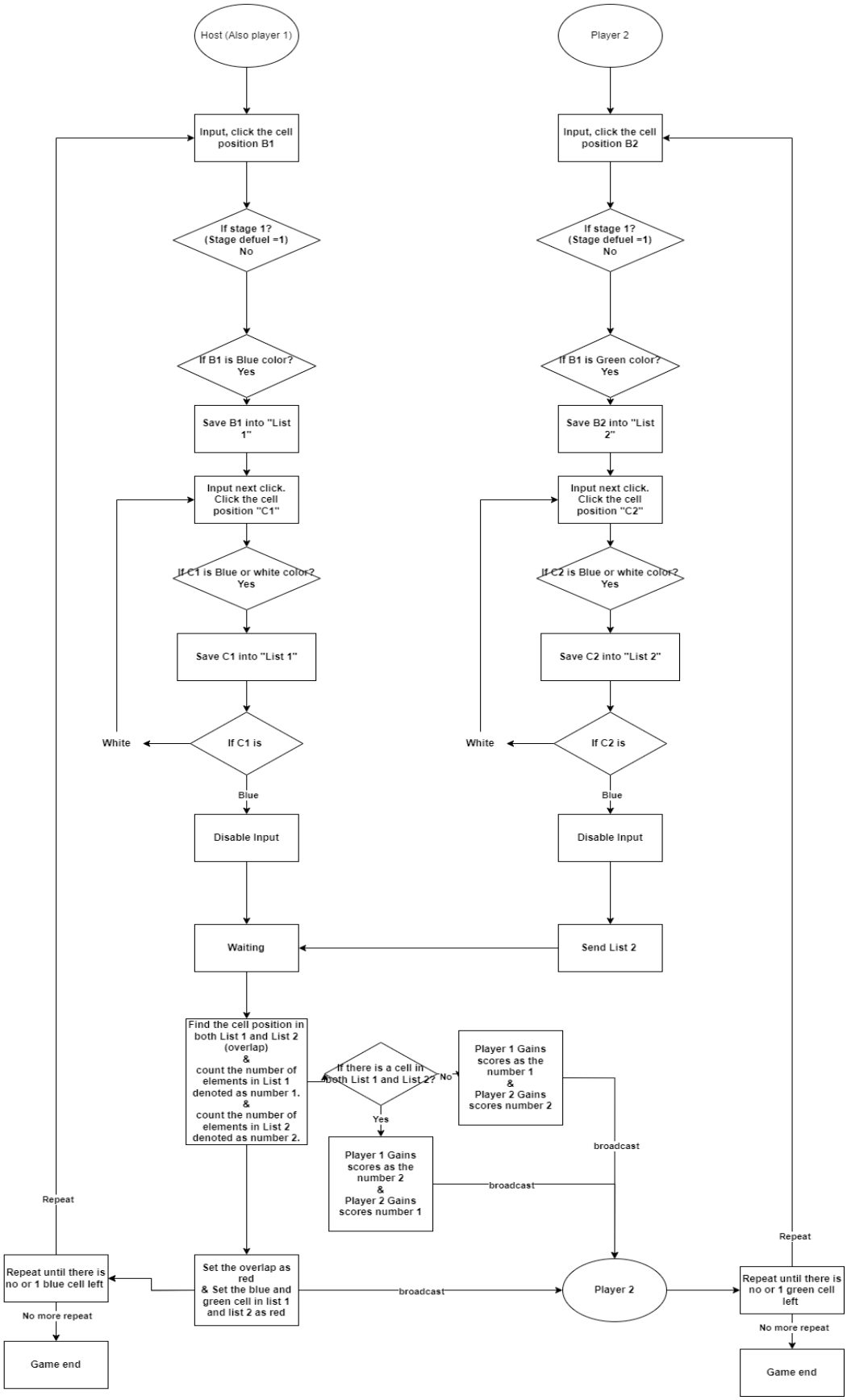


Figure A1. Flow chart of game logic.

BILINGUAL PUBLISHING GROUP

Journal of Computer Science Research

https://journals.bilpubgroup.com/index.php/jcsr

ARTICLE

Detection of Buffer Overflow Attacks with Memoization-based Rule Set

***Oğuz Özger*[1], *Halit ÖztekİN*[2*]**

[1] *Electrical-Electronics Engineering, Sakarya University of Applied Sciences, Sakarya, 54050, Turkey*

[2] *Computer Engineering, Sakarya University of Applied Sciences, Sakarya, 54050, Turkey*

ABSTRACT

Different abnormalities are commonly encountered in computer network systems. These types of abnormalities can lead to critical data losses or unauthorized access in the systems. Buffer overflow anomaly is a prominent issue among these abnormalities, posing a serious threat to network security. The primary objective of this study is to identify the potential risks of buffer overflow that can be caused by functions frequently used in the PHP programming language and to provide solutions to minimize these risks. Static code analyzers are used to detect security vulnerabilities, among which SonarQube stands out with its extensive library, flexible customization options, and reliability in the industry. In this context, a customized rule set aimed at automatically detecting buffer overflows has been developed on the SonarQube platform. The memoization optimization technique used while creating the customized rule set enhances the speed and efficiency of the code analysis process. As a result, the code analysis process is not repeatedly run for code snippets that have been analyzed before, significantly reducing processing time and resource utilization. In this study, a memoization-based rule set was utilized to detect critical security vulnerabilities that could lead to buffer overflow in source codes written in the PHP programming language. Thus, the analysis process is not repeatedly run for code snippets that have been analyzed before, leading to a significant reduction in processing time and resource utilization. In a case study conducted to assess the effectiveness of this method, a significant decrease in the source code analysis time was observed.

Keywords: Buffer overflow; Cybersecurity; Anomaly; SonarQube; Memoization

*CORRESPONDING AUTHOR:

Halit ÖztekİN, Computer Engineering, Sakarya University of Applied Sciences, Sakarya, 54050, Turkey; Email: halitoztekin@subu.edu.tr

ARTICLE INFO

Received: 28 October 2023 | Revised: 15 November 2023 | Accepted: 22 November 2023 | Published Online: 30 November 2023

DOI: https://doi.org/10.30564/jcsr.v5i4.6044

CITATION

Özger, O., ÖztekİN, H., 2023. Detection of Buffer Overflow Attacks with Memoization-based Rule Set. Journal of Computer Science Research. 5(4): 13-26. DOI: https://doi.org/10.30564/jcsr.v5i4.6044

1. Introduction

Internet networks are an indispensable component that facilitates information exchange between individuals and organizations; however, this infrastructure needs to be protected against threats such as network anomalies. One of these dangers is buffer overflows. Buffer overflow can lead to an attack that allows malicious codes to be executed on the target system. Therefore, the detection and prevention of buffer overflow attacks are of great importance in ensuring the security of computer systems. The attacks and impacts caused by buffer overflow errors hold a significant place in the history of technology. The Morris worm, one of the first major attacks in the history of the internet, originated from a buffer overflow error [1]. The worm named Code Red [2] on the other hand, affected millions of computers due to a buffer overflow error in Microsoft IIS servers, displaying a widespread distribution over the internet. The Heartbleed vulnerability, which emerged in the OpenSSL cryptography library, put at risk the SSL encryption protocol used across a large part of the internet [3]. Another incident stemming from a buffer overflow error is the attack on Equifax, a credit report service provider [4]. In addition, the security vulnerabilities named Spectre and Meltdown in Intel processors [5], the BlueKeep vulnerability in the Windows operating system [6], the attack on the popular messaging application WhatsApp [7] and the attack on the Microsoft Exchange email server [8] all share a common ground originating from structures susceptible to buffer overflow.

It is possible to protect against buffer overflow attacks by taking a series of precautionary measures. These include controlling the array size [9], using secure input/output functions, memory limiting [10], Address Space Layout Randomization (ASLR), stack canaries, software updates, conducting security testing, fuzzing, and sandboxing [11]. Thanks to these protection methods, many systems are known to have become more resilient [12,13]. Additionally, hardware-based measures are stronger compared to software-based measures, as they are based on the working logic of the hardware and are harder to override or bypass by software. However, these measures are only supported by modern CPUs and cannot be used in older systems. Hardware-based memory protection technologies are widely supported, particularly in modern processor architectures, such as Intel (VT-x, VT-d, etc.) and AMD (V, Vi, etc.) [14].

Hardware-based measures yield the most effective results when implemented alongside software-based solutions. Various software tools, such as dynamic and static code analysis tools, memory tracing utilities, and fuzzers, can be utilized to detect such types of cyber threats. Particularly, static code analysis is instrumental in identifying potential security weaknesses by conducting an analysis of the code without the necessity to execute it. In this realm, static code analysis tools like SonarQube offer distinctive advantages in the early detection of potential security vulnerabilities within source codes. The objective of this study is to ascertain the potential risks of buffer overflow incidents precipitated by functions commonly employed in the PHP programming language and to propose methodologies to mitigate these risks. The prominence of PHP as one of the globally most-utilized scripting languages today, coupled with the reality that a significant proportion of security vulnerabilities identified in web-based computer software in recent times are associated with PHP, underpins the rationale for selecting this language for our investigation. Static code analyzers are deployed to pinpoint security vulnerabilities, among which SonarQube is notable for its comprehensive library, adaptable customization options, and its industry-trusted reliability. Within this framework, a specialized, memoization-based rule set designed for the automatic detection of buffer overflows has been developed on the SonarQube platform.

This customized rule set, through extensive analysis executed on PHP source codes, can preemptively identify memory management faults, particularly dangerous function calls, and erroneous array operations. This facilitates the proactive correction of security vulnerabilities in the earlier stages of the development cycle. Integrated into SonarQube's rich plugin ecosystem, this rule set offers in-depth guid-

ance to developers, security analysis experts, and information security teams on fortifying PHP codes. Furthermore, the rule set can be calibrated to recognize specific security weaknesses prevalent in widely used PHP frameworks and libraries, significantly bolstering the security posture of PHP applications in the industry and elevating the standards of web application security.

2. Literature review

Aleph One [15] has presented a detailed report on the causes and effects of buffer overflow attacks, which hold a significant place among network anomalies, while Wagner and colleagues have addressed this threat with various protection methods and detection approaches [16]. Protection methods like StackGuard [17] and RAD [18] stand out as prominent approaches in the literature. Kuperman and colleagues have proposed detecting buffer overflow by combining static and dynamic analyses [19]. Le and Soffa have detected these attacks with demand-based and path-sensitive analysis [20]. Brooks has evaluated the effects of automatic vulnerability detection and exploitation techniques on buffer overflow [21]. A report [22] highlighting the critical importance of static analysis techniques for software security has been presented, and accordingly, tools like ARCHER [23] and Safe-C [24] introduced in the literature automatically detect memory access errors.

To detect and prevent memory errors at runtime, Valgrind performs dynamic program inspection to monitor potential memory errors [25], and Rinard and colleagues have conducted studies in this field using dynamic methods [26]. Additionally, FormatGuard [27] has been developed by Fen and colleagues [28], and Ruwase and Lam [29] have devised protection methods against memory overflow errors.

Buffer overflow attacks stem from memory management errors in software, providing attackers the opportunity to exploit these flaws for the execution of malicious code. Jha has conducted research on methods to close security vulnerabilities in software and enhance its security [30]. Emami, Ghiya, and Hendren have focused on the potential security challenges posed by the use of function pointers in C [31]. Liang and Sekar have generated automatic signatures against buffer overflow attacks using symbolic modeling [32]. In the same context, Newsome and Song have detected malware attacks using dynamic taint analysis [33]. Memory errors can pose a risk to the reliability of software. Qin and colleagues have introduced a method of monitoring ECC memory to detect these errors [34]. At the same time, Seward and Nethercote have outlined methods for detecting undefined value errors using the Valgrind tool [35]. Executable memory protection for Linux is a critical defence against attacks, and a comprehensive overview of this subject is provided on Wikipedia [36].

Nethercote and Seward have introduced methods for detecting software bugs by tracking dynamic features through the Valgrind framework [37]. Costa and colleagues have worked on enhancing software security by conducting dynamic input verification with Bouncer [38]. Song and his team have presented the BitBlaze method for the analysis of malicious software [39]. Liu and his colleagues have identified vulnerabilities in x86 programs using obfuscation methods and genetic algorithms [40]. Kroes and his team have provided automatic detection methods for memory management errors using Delta pointers [41]. Frantzen and Shuey have introduced hardware-assisted stack protection with StackGhost [42]. Novark and Berger have presented dynamic memory management approaches with the DieHarder tool [43]. Sayeed and colleagues have proposed protection against buffer overflow attacks through control flow integrity [44]. Andriesse and his team have offered methods for software integrity protection and blocking malicious code with Parallax [45].

In recent years, machine learning has emerged as a method used for attack detection from network traffic. Mukkamala and colleagues have conducted studies in this field [46], while Thottan and Ji have performed anomaly detection by examining the characteristics of network traffic data [47].

For the detection of attacks, tools such as dynamic and static code analysis tools, memory tracing tools, and fuzzers can be utilized. In particular, static code

analysis helps identify potential security vulnerabilities through the method of analyzing the code without executing it. To detect security vulnerabilities in software, Cova and colleagues have presented methods combining flow graphs and static analysis [48]. Lanzi and colleagues have also proposed a vulnerability detection tool for the x86 architecture [49]. At the same time, Miller and colleagues have tested the resilience of applications for MacOS X [50]. In this context, static code analysis tools like SonarQube possess unique advantages in detecting potential security vulnerabilities in software at an early stage. Its multi-language support for various programming languages allows for language-independent analysis of projects. Its customizability feature allows for the addition of specific rules. It can incorporate customized rule creation in SonarQube as well as the integration of an optimization algorithm with machine learning.

3. Speed up static code analysis

WERTYMemoization is an optimization technique used to store the results of computationally expensive operations, preventing the need for repeated execution of the same operations. This method is particularly employed in dynamic programming problems to minimize redundant calculations. In the fields of data mining and big data analysis, memoization contributes to resolving the time issues encountered when algorithms process large data sets. A HashMap function can be created for the source code, storing information about previously seen function names and whether these functions produce a "buffer overflow". Consequently, when a function listed is encountered, the code does not have to recheck if the function produces a "buffer overflow"; instead, it quickly retrieves the result from the HashMap function.

Static code analysis tools aim to identify errors, assess code quality, and pinpoint security vulnerabilities in codes written in various programming languages. However, in large-scale projects, completing each analysis can take a considerable amount of time. The integration of the memoization technique into static code analysis is targeted to quickly analyze repeating functions or method calls, thereby reducing the total analysis time. A functional comparison of prominent and widely accepted static code analysis tools in the literature is provided in **Table 1**.

The static code analysis tool SonarQube stands out among other analysis tools due to its support for numerous programming languages, visual reporting and user-friendly interface, offering broader customization options, and having a large user and developer community.

Table 1. Functional comparison of code analysis tools.

Feature/Criterion	SonarQube [51]	ESLint	Checkmzrx	Coverity
Supported languages	20+ (Java, C#, PHP, JS vb.)	Firstly, JS ve TypeScript	20+	20+
Web based interface	Yes	No (CLI tabanlı)	Yes	No
CI/CD integration	Extensive (Jenkins, Travis, Azure Pipelines vb.)	Limited	Extensive	Extensive
Customizable rules	Yes	Yes	Yes	Yes
Community support	Strong	Strong	Medium	Medium
Code quality metrics	Yes	No	No	No
Licensing and cost	Free and commercial versions	Free	Commercial	Commercial
Multi-language project support	Yes	No	Yes	Yes
Plugin and extension support	Yes (Numerous plugins available on the Marketplace)	Yes (npm packages)	Limited	Limited

4. Case study

Methods used for automatic detection of security vulnerabilities causing buffer overflow in static code analysis are presented in this section. The interface named Sonar-Scanner, which can work integrated with SonarQube, enables the visualization of analysis results and the execution of management operations. On the SonarQube platform, a special security rule has been established with the aim of detecting PHP functions that can create a security vulnerability (buffer overflow).

4.1 Memoization-based custom rule integration

The steps for incorporating the memoization principle into the rule are given below. Integration of the rule with the SonarQube analysis tool not only reduces analysis time, but also enables quicker detection of critical security risks.

Step-1: Creating a custom security rule: Leveraging the customizable structure of SonarQube analysis tool for PHP, a security rule incorporating the memoization technique is established.

Step-2: Cache management: Implementing a caching mechanism within the custom rule to store the results of functions that pose a security vulnerability risk.

Step-3: Risk analysis: Evaluating potential security risks when functions or methods are called for the first time, and storing the result in the cache.

Step-4: Utilizing the cache: When a function of the same name is called again, use the information in the cache to instantly perform a security assessment.

In the first step, key files and functions for the specific rules determined for static code analysis are utilized. The file named BufferOverflowCheck.java examines function calls, checking for the usage of specific functions such as strcpy, strcat, and fwrite. When the use of these functions is detected, a security alert is generated. The MyPhpRules.java file stores the list of existing custom rules, and these rules are added to the SonarQube rule repository. The PHPCustomRulesPlugin.java file defines the Sonar Plugin and adds the MyPhpRules class. The pom.xml is used as a Maven configuration file for building the project and managing its dependencies. The necessary modules for creating a custom rule in the SonarQube analysis tool are provided in **Figure 1**.

As shown in **Figure 2**, the SonarQube tool uses the "@Rule" annotation to customize rule definitions. There are different priority levels in these checks, which are listed as INFO, MINOR, MAJOR, CRITICAL, and BLOCKER. For the rule created

```
package org.sonar.samples.php.checks;

import org.sonar.check.Priority;
import org.sonar.check.Rule;
import org.sonar.plugins.php.api.tree.expression.FunctionCallTree;
import org.sonar.plugins.php.api.visitors.PHPVisitorCheck;

import java.util.Map;
import java.util.HashMap;
import java.util.List;
import java.util.ArrayList;
import java.nio.file.*;
import java.io.IOException;

import java.io.BufferedWriter;
import java.io.FileWriter;
```

Figure 1. Loading of modules.

```
@Rule(
  key = BufferOverflowCheck.KEY,
  priority = Priority.CRITICAL,
  name = "Buffer overflow issues should be fixed.",
  tags = {"security"}
)
```

Figure 2. Determination of rule level.

in this study, the CRITICAL priority level has been selected because buffer overflow can lead to security breaches, and such a violation can pose a critical threat to the entire system.

As shown in **Figure 3**, the constructor (Buffer OverflowCheck()) creates a HashMap named functionResultsCache. This structure is later used to quickly detect risky functions. By calling the loadCache() function, a previously created cache is loaded.

As illustrated in **Figure 4**, the loadCache() function reads the previously saved cache data from the disk and places it in the HashMap. Each line read from the disk contains a function name and a boolean value. The function name serves as the key, and the boolean value is processed into the map.

In every situation where the cache is updated, the saveCache() function is invoked as depicted in **Figure 5**. This function saves the functionResultsCache HashMap to a file.

As illustrated in **Figure 6**, the visitFunctionCall() function is triggered for each function call in the source code. After retrieving the function name, it is checked whether it has been previously added to the cache or not.

If the function name exists in the cache, the stored boolean value is used. If the value is true, a "Buffer overflow issue detected" warning is generated. If the function name has not been seen before, it is checked whether it is risky. If it is in the list of risky functions (such as strcpy, strcat, gets, etc.), it is added to the cache as true and subsequently a warning is generated. When the cache is updated, the saveCache() function is called, and the process continues for other potential SonarQube tool inspections with super. visitFunctionCall(tree) (**Figure 7**).

```
public class BufferOverflowCheck extends PHPVisitorCheck {

  public static final String KEY = "S3";
  private final Map<String, Boolean> functionResultsCache;

  public BufferOverflowCheck() {
    functionResultsCache = new HashMap<>();
    loadCache();
  }
```

Figure 3. Creation the BufferOverflowCheck class.

```
  private void loadCache() {
    try {
      List<String> lines = Files.readAllLines(Paths.get("/cache_dosyasi/cache.txt"));
      for (String line : lines) {
        String[] parts = line.split(",");
        functionResultsCache.put(parts[0], Boolean.parseBoolean(parts[1]));
      }
    } catch (IOException e) {
      e.printStackTrace();
    }
  }
```

Figure 4. LoadCahce function.

```
  private void saveCache() {
    try {
      List<String> lines = new ArrayList<>();
      for (Map.Entry<String, Boolean> entry : functionResultsCache.entrySet()) {
        lines.add(entry.getKey() + "," + entry.getValue());
      }
      Files.write(Paths.get("/cache_dosyasi/cache.txt"), lines);
    } catch (IOException e) {
      e.printStackTrace();
    }
  }
```

Figure 5. SaveCahce function.

```
@Override
public void visitFunctionCall(FunctionCallTree tree) {
  String functionName = tree.callee().toString();

  if (functionResultsCache.containsKey(functionName)) {
    if (functionResultsCache.get(functionName)) {
      context().newIssue(this, tree, "Buffer overflow issue detected.");
    }
    return; // Cache'den bilgi alındı, işlemi sonlandır
  }
```

Figure 6. visitFunctionCall function.

```
  boolean isOverflow = functionName.equals("strcpy") ||
                       functionName.equals("strcat") ||
                       functionName.equals("strncpy") ||
                       functionName.equals("sprintf") ||
                       functionName.equals("unpack") ||
                       functionName.equals("addslashes") ||
                       functionName.equals("fread") ||
                       functionName.equals("fwrite") ||
                       functionName.equals("strtr") ||
                       functionName.equals("gzuncompress") ||
                       functionName.equals("getimagesize") ||
                       functionName.equals("iptcembed") ||
                       functionName.equals("mb_parse_str") ||
                       functionName.equals("move_uploaded_file") ||
                       functionName.equals("array_splice");

  functionResultsCache.put(functionName, isOverflow);

  if (isOverflow) {
    context().newIssue(this, tree, "Buffer overflow issue detected.");
  }

  saveCache(); // Cache değiştiği için kaydet

  super.visitFunctionCall(tree);
 }
}
```

Figure 7. Identifying functions that cause buffer overflow.

4.2 Memoization-based rule for attack detection

This structure also employs a caching mechanism to enhance performance. Specifically, the saveCache and loadCache methods are used for loading the cache file from the disk and saving it back to the disk. This custom rule can effectively detect buffer overflow issues in PHP projects, assisting in the prevention of such security vulnerabilities. The workflow of the rule added to the Sonar Qube analysis tool is provided in the steps below.

Step-1: The BufferOverflowCheck class stores function names and whether these functions may cause a buffer overflow in a HashMap.

Step-2: When any function call is seen in the PHP code, the visitFunctionCall function is triggered by the rule.

Step-3: If the function name is in the cache (HashMap), the rule can immediately generate a warning. Otherwise, the function name and whether it creates a buffer overflow is added to the cache.

In the example application below, a custom rule has been added to the SonarQube analysis tool to detect buffer overflow attacks. When dangerous PHP functions (such as strncpy, strcat, addslashes, fwrite, array_splice, etc.) are called, the application issues a security vulnerability warning (**Figure 8**).

```
        printf("<form class='frmsource' method='post'>
                        <textarea id='sourcefocus' name='sourcecode' rows='25' cols='100'>%s</textarea>
                        <input type='Submit' value='Save file' name='save'/>
                        <label>%s</label>
                </form>",$source,$status);

        if(any("status",$_SESSION)) unset($_SESSION['status']);

        if(any("save",$_REQUEST))
        {
                $new_source=$_REQUEST['sourcecode'];
                if(function_exists("chmod")) chmod($file,0755);
                $source_edit=fopen($file,'w+');
                $tulis=fwrite($source_edit,$new_source);
                fclose($source_edit);
                if($tulis)
                {
                        $_SESSION['status']="File Saved ! ".GetFileTime($file,"modify")." | ".GetFileSize(filesize($file));
                }
                else
                {
                        $_SESSION['status']="Whoops, something went wrong...";
                }
                header("location:".php_self."?a=e&r=".urle($file));
        }
}

if($_REQUEST['a']=='r')
{
        printf("<form class='new' method='post'>
                <input type='text' name='name' value='%s'/>
                <input type='Submit' value='Rename' name='rename'/>
                <label>%s</label>
        </form>",basename($file),$status);

        if(any("status",$_SESSION)) unset($_SESSION['status']);
```

Figure 8. Code block in bat.php file where the “fwrite” function is used.

Additionally, the effectiveness of the memoization method has been measured by applying the custom rule created to a PHP code found in the bat.php repository located in the “b4tm4n” repository on GitHub [45]. The source code selected for analysis consists of a total of 3962 lines. After the SonarQube analysis tool successfully analyzes the code, it displays the identified risky functions and security vulnerabilities on the interface. **Figure 9** shows the results of the analysis summary of the bat.php file available on Github.

Figure 10 shows the lines of code in the bat.php file that are potentially vulnerable to buffer overflow attacks.

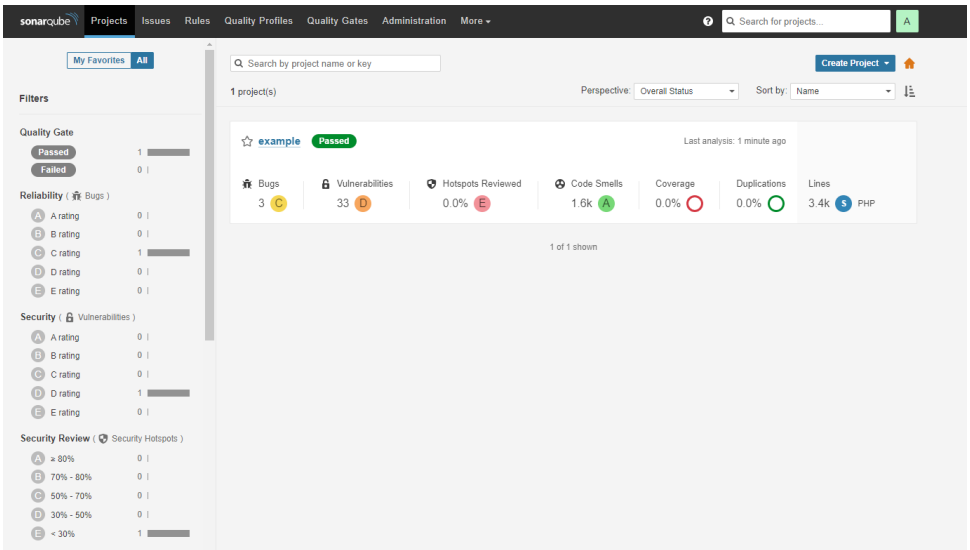


Figure 9. Analysis summary results of the bat.php file located on Github.

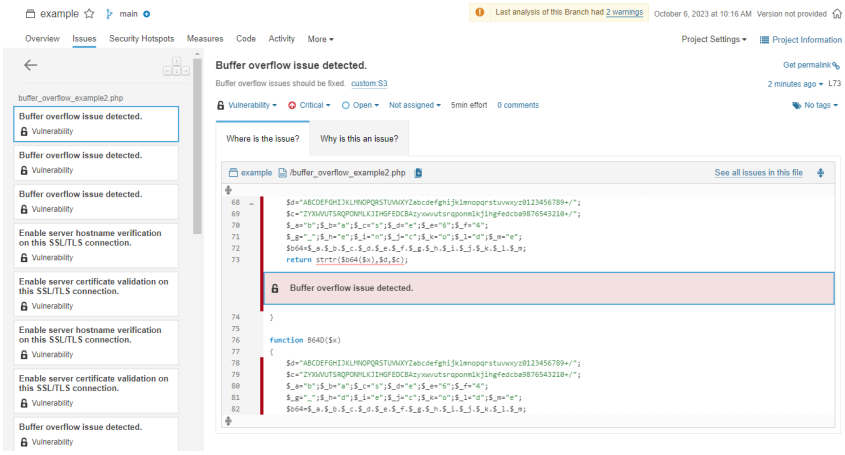


Figure 10. Code lines in the bat.php file that are potentially vulnerable to buffer overflow attacks.

During the analysis with the custom rule created, the information of the dangerous and safe functions detected thanks to the loadCache function is saved to a file named cache.txt. An example of a cache.txt file is shown below:

As shown in **Figure 11**, there are functions marked as true or false. Here, true represents the functions that may pose a security vulnerability risk, while false represents the functions that will not create a security vulnerability. The existence of the cache.txt file significantly increases the performance during the next run of the custom rule. The loadCache function reads this file at the beginning of the analysis and loads the function information into memory (RAM). Thus, when re-analyzing the same code, instead of analyzing the security risks of the same functions again, the custom rule directly retrieves the information from the memory. This is an optimization technique known as memoization. This approach shortens the analysis time and can quickly identify previously detected security vulnerabilities in each new analysis. **Figures 12-14** show the analysis times of the lines in the bat.php file analyzed above against buffer overflow attacks. While the total analysis time is 29.118 s when the cache is empty, it is 22.048 s when the cache is full. It can be seen that the use of a cache significantly reduces the analysis time.

```
sprintf,true
set_error_handler,false
strlen,false
filter_var,false
http_build_query,false
move_uploaded_file,true
GetUser,false
```

Figure 11. Sample content is taken from the cache.txt file.

```
private void loadCache() {
  try {
    List<String> lines = Files.readAllLines(Paths.get("/cache_dosyasi/cache.txt"));
    for (String line : lines) {
      String[] parts = line.split(",");
      functionResultsCache.put(parts[0], Boolean.parseBoolean(parts[1]));
    }
  } catch (IOException e) {
    e.printStackTrace();
  }
}
```

Figure 12. Memoization.

```
INFO: Analysis total time: 19.351 s
INFO: ------------------------------------------------------------------------
INFO: EXECUTION SUCCESS
INFO: ------------------------------------------------------------------------
INFO: Total time: 22.048s
INFO: Final Memory: 23M/87M
INFO: ------------------------------------------------------------------------
```

Figure 13. Analysis duration with cache usage.

```
INFO: Analysis total time: 26.699 s
INFO: ------------------------------------------------------------------------
INFO: EXECUTION SUCCESS
INFO: ------------------------------------------------------------------------
INFO: Total time: 29.118s
INFO: Final Memory: 23M/87M
INFO: ------------------------------------------------------------------------
```

Figure 14. Analysis duration without cache usage.

5. Results and discussion

In this study, we examine how a customized rule can be added to the SonarQube analysis tool to identify potential buffer overflow security vulnerabilities in codes written in PHP. The main goal is to develop a system that not only performs static code analysis but also enhances analysis performance with a caching mechanism. Specific functions that could pose a risk in PHP codes have been identified, and a custom SonarQube rule has been designed for these functions. The analysis process has been optimized using the memoization technique, reducing repeated analyses on the same code. This approach saves time and resources, particularly for large codebases.

The current work also records significant advancements in the effective detection of commonly used functions that carry the risk of buffer overflow. The increase in analysis performance has been made possible by the caching mechanism. The flexibility and extensibility of this custom rule mean it can be applied to different functions and methods. However, the study has limitations, such as being specific to the PHP language and covering only certain functions. While the initial version is capable of static code analysis only, the integration of dynamic analysis and adaptation to different programming languages represent significant potential for future work.

This research makes a notable contribution to the field of web security, especially regarding the security of PHP applications. In the future, the automatic detection and resolution of such security vulnerabilities are expected to facilitate the creation of more effective and secure applications in software development and cybersecurity. Particularly, the integration of the custom rule with dynamic analysis, the incorporation of machine learning techniques, and improvements to the user interface will push the innovations in this field further. In addition, advanced automation approaches for strengthening defence mechanisms against security vulnerabilities and making software development processes more robust could be developed, contributing to risk assessment and management strategies. Designing an extended security framework applicable in various software languages to deal with a wide spectrum of security threats, beyond specific challenges such as buffer overflow, will be an important goal for future researchers and applications. Such a framework will play a crucial role in protecting critical systems and securing sensitive data.

Author Contributions

Halit Oztekin and Oguz Ozger—conceptualization, methodology, formal analysis, investigation, supervision, validation, visualization, writing—original draft, and writing—review and editing.

Conflict of Interest

The author declares that there are no conflicts of interest.

Funding

This research received no external funding.

Acknowledgement

There is no acknowledgement for this article.

References

[1] Spafford, E.H., 1989. The Internet worm program: An analysis. ACM SIGCOMM Computer Communication Review. 19(1), 17-57. DOI: https://doi.org/10.1145/66093.66095

[2] Moore, D., Paxson, V., Savage, S., et al., 2003.

Inside the slammer worm. IEEE Security & Privacy. 1(4), 33-39.
DOI: https://doi.org/10.1109/MSECP.2003.1219056

[3] Springall, D., Durumeric, Z., Halderman, J.A. (editors), 2016. Measuring the security harm of TLS crypto shortcuts. IMC'16: Proceedings of the 2016 Internet Measurement Conference; 2016 Nov 14-16; Santa Monica California USA. New York: Association for Computing Machinery. p. 33-47.
DOI: https://doi.org/10.1145/2987443.2987480

[4] Oversight and Government Reform [Internet]. Available from: https://oversight.house.gov/wp-content/uploads/2018/12/Equifax-Report.pdf

[5] Kocher, P., Horn, J., Fogh, A., et al. (editors), 2019. Spectre attacks: Exploiting speculative execution. 2019 IEEE Symposium on Security and Privacy (SP); 2019 May 19-23; San Francisco, CA, USA. New York: IEEE.
DOI: https://doi.org/10.1109/SP.2019.00002

[6] Remote Desktop Services Remote Code Execution Vulnerability [Internet]. Available from: https://msrc.microsoft.com/update-guide/en-US/vulnerability/CVE-2019-0708

[7] Bhutan Built a Bitcoin Mine on the Site of Its Failed 'Education City' [Internet]. Available from: https://www.forbes.com/sites/zakdoffman/2019/05/15/whatsapp-hack-an-effortless-way-to-steal-your-account/?sh=678fe50a1051

[8] On-Premises Exchange Server Vulnerabilities Resource Center—updated March 25, 2021 [Internet]. Available from: https://msrc.microsoft.com/blog/2021/03/multiple-security-updates-released-for-exchange-server/

[9] Dowd, M., McDonald, J., Schuh, J., 2006. The art of software security assessment: Identifying and preventing software vulnerabilities. Pearson Education: London.

[10] SEI CERT Oracle Coding Standard for Java [Internet]. Available from: https://wiki.sei.cmu.edu/confluence/display/java/SEI+CERT+Oracle+Coding+Standard+for+Java

[11] Conklin, W.A., White, G., Cothren, C., et al., 2022. Principles of Computer Security: CompTIA Security+™ and Beyond [Internet]. Available from: https://sisis.rz.htw-berlin.de/inh2010/12389366.pdf

[12] Xu, J., Patel, S., Iyer, R., et al., 2002. Architecture Support for Defending Against Buffer Overflow Attacks [Internet]. Available from: https://www.ideals.illinois.edu/items/100089/bitstreams/319746/data.pdf

[13] Anley, C., Heasman, J., Lindner, F., et al., 2007. The shellcoder's handbook: Discovering and exploiting security holes. Wiley: Hoboken.

[14] Intel® 64 and IA-32 Architectures Software Developer Manuals [Internet]. Available from: https://www.intel.com/content/www/us/en/developer/articles/technical/intel-sdm.html

[15] One, A., 1996. Smashing the stack for fun and profit. Phrack Magazine. 7(49), 14-16.

[16] Wagner, D., Foster, J.S., Brewer, E.A., et al., 2000. A First Step Towards Automated Detection of Buffer Overrun Vulnerabilities [Internet]. Available from: https://www.cs.umd.edu/class/spring2021/cmsc614/papers/automated-buffer.pdf

[17] Baratloo, A., Singh, N., Tsai, T. (editors), 2000. Transparent run-time defense against stack smashing attacks. Proceedings of the 2000 USENIX Annual Technical Conference; 2000 Jun 18-23; San Diego, California, USA.

[18] Chiueh, T.C., Hsu, F.H. (editors), 2001. RAD: A compile-time solution to buffer overflow attacks. Proceedings of the 21st International Conference on Distributed Computing Systems; 2001 Apr 16-19; Mesa, AZ, USA. New York: IEEE. p. 409-417.
DOI: https://doi.org/10.1109/ICDSC.2001.918971

[19] Kuperman, B., Brodley, C., Ozdoganoglu, H., et al., 2005. Detection and prevention of stack buffer overflow attacks. Communications of the ACM. 48(11), 50-56.
DOI: https://doi.org/10.1145/1096000.1096004

[20] Le, W., Soffa, M.L. (editors), 2008. Marple: A demand-driven path-sensitive buffer overflow detector. Proceedings of the ACM SIGSOFT

Symposium on the Foundations of Software Engineering; 2008 Nov 9-14; Atlanta Georgia. New York: Association for Computing Machinery. p. 272-282.
DOI: https://doi.org/10.1145/1453101.1453137

[21] Brooks, T.N., 2017. Survey of automated vulnerability detection and exploit generation techniques in cyber reasoning systems. Advances in intelligent systems and computing. Springer: Cham. pp. 1083-1102.
DOI: https://doi.org/10.1007/978-3-030-01177-2_79

[22] Chess, W., 1998. Secure programming with static analysis. Pearson Education: London.

[23] Cowan, C., Beattie, S., Johansen, J., et al. (editors), 2003. Pointguard TM: Protecting pointers from buffer overflow vulnerabilities. Proceedings of the 12th USENIX Security Symposium; 2003 Aug 4-8; Washington D.C., USA.

[24] Yong, S.H., Horwitz, S. (editors), 2003. Protecting C programs from attacks via invalid pointer dereferences. Proceedings of the 9th European Software Engineering Conference Held Jointly with 11th ACM SIGSOFT International Symposium on Foundations of Software Engineering; 2003 Sep 1-5; Helsinki Finland. New York: Association for Computing Machinery. p. 307-316.
DOI: https://doi.org/10.1145/940071.940113

[25] Nethercote, N., Seward, J., 2003. Valgrind: A program supervision framework. Electronic Notes in Theoretical Computer Science. 89(2), 44-66.
DOI: https://doi.org/10.1016/S1571-0661(04)81042-9

[26] Rinard, M., Cadar, C., Dumitran, D., et al. (editors), 2004. A dynamic technique for eliminating buffer overflow vulnerabilities (and other memory errors). 20th Annual Computer Security Applications Conference; 2004 Dec 6-10; Tucson, AZ, USA. New York: IEEE. p. 82-90.
DOI: https://doi.org/10.1109/CSAC.2004.2

[27] Cowan, C., Barringer, M., Beattie, S., et al. (editors), 2001. FormatGuard: Automatic protection from printf format string vulnerabilities. Proceedings of the 10th USENIX Security Symposium; 2001 Aug 13-17; Washington D.C., USA. p. 191-200.

[28] Fen, Y., Fuchao, Y., Xiaobing, S., et al., 2012. A new data randomization method to defend buffer overflow attacks. Physics Procedia. 24, 1757-1764.
DOI: https://doi.org/10.1016/j.phpro.2012.02.259

[29] Ruwase, O., Lam, M.S., 2003. A Practical Dynamic Buffer Overflow Detector [Internet]. Available from: http://www.cs.cmu.edu/afs/cs.cmu.edu/Web/People/oor/papers/cred.pdf

[30] Jha, S. (editor), 2010. Retrofitting legacy code for security. Computer Aided Verification, 22nd International Conference, CAV 2010; 2010 Jul 15-19; Edinburgh, UK. Berlin: Springer.
DOI: https://doi.org/10.1007/978-3-642-14295-6_2

[31] Emami, M., Ghiya, R., Hendren, L. (editors), 1994. Context-sensitive interprocedural points-to analysis in the presence of function pointers. Proceedings of the ACM SIGPLAN 1994 Conference on Programming Language Design and Implementation; 1994 Jun 20-24; Orlando Florida USA. New York: Association for Computing Machinery.
DOI: https://doi.org/10.1145/178243.178264

[32] Liang, Z., Sekar, R. (editors), 2005. Automatic generation of buffer overflow attack signatures: An approach based on program behavior models. 21st Annual Computer Security Applications Conference (ACSAC'05); 2005 Dec 5-9; Tucson, AZ, USA. New York: IEEE. p. 10-224.
DOI: https://doi.org/10.1109/CSAC.2005.12

[33] Newsome, J., Song, D., 2005. Dynamic Taint Analysis for Automatic Detection, Analysis, and Signature Generation of Exploits on Commodity Software [Internet]. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5803709632cf010d3923e8f85416bb95db0dd8ea

[34] Qin, F., Lu, S., Zhou, Y. (editors), 2005. Safe-

Mem: Exploiting ECC-memory for detecting memory leaks and memory corruption during production runs. 11th International Symposium on High-Performance Computer Architecture; 2005 Feb 12-16; San Francisco, CA, USA. New York: IEEE. p. 291-302.
DOI: https://doi.org/10.1109/HPCA.2005.29

[35] Seward, J., Nethercote, N. (editors), 2005. Using Valgrind to detect undefined value errors with bit-precision. Proceedings of the USENIX'05 Annual Technical Conference; 2005 Apr 10-15; Anaheim, California, USA.

[36] Executable-space Protection [Internet]. Available from: https://en.wikipedia.org/wiki/Executable-space_protection

[37] Nethercote, N., Seward, J., 2007. Valgrind: A framework for heavyweight dynamic binary instrumentation. ACM Sigplan Notices. 42(6), 89-100.
DOI: https://doi.org/10.1145/1273442.1250746

[38] Costa, M. (editor), 2008. Bouncer: Securing software by blocking bad input. Proceedings of the 2nd Workshop on Recent Advances on Intrusiton-tolerant Systems; 2008 Apr 1; Glasgow United Kingdom. New York: Association for Computing Machinery.
DOI: https://doi.org/10.1145/1413901.1413902

[39] Song, D., Brumley, D., Yin, H., et al. (editors), 2008. BitBlaze: A new approach to computer security via binary analysis. Information Systems Security, 4th International Conference, ICISS 2008; 2008 Dec 16-20; Hyderabad, India. Berlin: Springer. p. 1-25.
DOI: https://doi.org/10.1007/978-3-540-89862-7_1

[40] Liu, G.H., Wu, G., Tao, Z., et al. (editors), 2008. Vulnerability analysis for x86 executables using genetic algorithm and fuzzing. 2008 Third International Conference on Convergence and Hybrid Information Technology; 2008 Nov 11-13; Busan, Korea (South). New York: IEEE. p. 491-497.
DOI: https://doi.org/10.1109/ICCIT.2008.9

[41] Kroes, T., Koning, K., Kouwe, E., et al. (editors), 2018. Delta pointers: Buffer overflow checks without the checks. EuroSys'18: Proceedings of the Thirteenth EuroSys Conference; 2018 Apr 23-26; Porto Portugal. New York: Association for Computing Machinery. p. 1-14.
DOI: https://doi.org/10.1145/3190508.3190553

[42] Frantzen, M., Shuey, M. (editors), 2001. StackGhost: Hardware facilitated stack protection. 10th USENIX Security Symposium; 2001 Aug 13-17; Washington, D.C., USA.

[43] Novark, G., Berger, E. (editors), 2010. DieHarder: Securing the heap. Proceedings of the 17th ACM Conference on Computer and Communications Security; 2010 Oct 4-8; Chicago, Illinois, USA. New York: Association for Computing Machinery. p. 573-584.
DOI: https://doi.org/10.1145/1866307.1866371

[44] Sayeed, S., Marco-Gisbert, H., Ripoll, I., et al., 2019. Control-flow integrity: Attacks and protections. Applied Sciences. 9(20), 4229.
DOI: https://doi.org/10.3390/app9204229

[45] Andriesse, D., Bos, H., Slowinska, A. (editors), 2015. Parallax: Implicit code integrity verification using return-oriented programming. 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks; 2015 Jun 22-25; Rio de Janeiro, Brazil. New York: IEEE. p. 125-135.
DOI: https://doi.org/10.1109/DSN.2015.12

[46] Mukkamala, S., Janoski, G., Sung, A. (editors), 2002. Intrusion detection using neural networks and support vector machines. Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290); 2002 May 12-17; Honolulu, HI, USA. New York: IEEE. p. 1702-1707.
DOI: https://doi.org/10.1109/IJCNN.2002.1007774

[47] Thottan, M., Ji, C., 2003. Anomaly detection in IP networks. IEEE Transactions on Signal Processing. 51(8), 2191-2204.
DOI: https://doi.org/10.1109/TSP.2003.814797

[48] Cova, M., Felmetsger, V., Banks, G., et al. (editors), 2006. Static detection of vulnerabilities in x86 executables. 2006 22nd Annual Computer Security Applications Conference

(ACSAC'06); 2006 Dec 11-15; Miami Beach, FL, USA. New York: IEEE. p. 269-278.
DOI: https://doi.org/10.1109/ACSAC.2006.50

[49] Lanzi, A., Martignoni, L., Monga, M., et al. (editors), 2007. A smart fuzzer for x86 executables. Third International Workshop on Software Engineering for Secure Systems (SESS'07: ICSE Workshops 2007); 2007 May 20-26; Minneapolis, MN, USA. New York: IEEE.
DOI: https://doi.org/10.1109/SESS.2007.1

[50] Miller, B.P., Cooksey, G., Moore, F., 2007. An empirical study of the robustness of MacOS applications using random testing. Operating Systems Review. 41(1), 78-86.
DOI: https://doi.org/10.1145/1228291.1228308

[51] SonarQube Homepage [Internet]. Available from: https://www.sonarqube.org/

BILINGUAL PUBLISHING GROUP

Journal of Computer Science Research

https://journals.bilpubgroup.com/index.php/jcsr

ARTICLE

A Natural Language Generation Algorithm for Greek by Using Hole Semantics and a Systemic Grammatical Formalism

Ioannis Giachos[1], Eleni Batzaki[1], Evangelos C. Papakitsos[1], Stavros Kaminaris[2], Nikolaos Laskaris[1]*

[1] *Department of Industrial Design & Production Engineering, University of West Attica, Egaleo, Athens, 12241, Greece*

[2] *Department of Electrical and Electronics Engineering, University of West Attica, Egaleo, Athens, 12241, Greece*

ABSTRACT

This work is about the progress of previous related work based on an experiment to improve the intelligence of robotic systems, with the aim of achieving more linguistic communication capabilities between humans and robots. In this paper, the authors attempt an algorithmic approach to natural language generation through hole semantics and by applying the OMAS-III computational model as a grammatical formalism. In the original work, a technical language is used, while in the later works, this has been replaced by a limited Greek natural language dictionary. This particular effort was made to give the evolving system the ability to ask questions, as well as the authors developed an initial dialogue system using these techniques. The results show that the use of these techniques the authors apply can give us a more sophisticated dialogue system in the future.

Keywords: Natural language processing; Natural language generation; Natural language understanding; Dialog system; Systemic grammar formalism; OMAS-III; HRI; Virtual assistant; Hole semantics

1. Introduction

The purpose of this work is to develop a computational Natural Language Generation (NLG) algorithm, for the Greek language, which will serve the human-machine communication process. An integrated HRI (Human-Robot Interaction) system [1] includes the dialogue process [2] and upgrades it to a role in relation to a Virtual Assistant [3]. In general, the need to generate sentences from an engine exists for two reasons:

*CORRESPONDING AUTHOR:

Evangelos C. Papakitsos, Department of Industrial Design & Production Engineering, University of West Attica, Egaleo, Athens, 12241, Greece; Email: papakitsev@uniwa.gr

ARTICLE INFO

Received: 5 November 2023 | Revised: 26 November 2023 | Accepted: 27 November 2023 | Published Online: 5 December 2023

DOI: https://doi.org/10.30564/jcsr.v5i4.6067

CITATION

Giachos, I., Batzaki, E., Papakitsos, E.C., et al., 2023. A Natural Language Generation Algorithm for Greek by Using Hole Semantics and a Systemic Grammatical Formalism. Journal of Computer Science Research. 5(4): 27-37. DOI: https://doi.org/10.30564/jcsr.v5i4.6067

1) for instructions or information to the user,

2) for questions about ambiguities (from the system side) of incoming suggestions from the user.

The first case, is the most widely applicable case, since machines give instructions to people all the time and everywhere, whether in the form of navigators or virtual assistants in homes and services. The second case is more crucial for dialogue development, since the system itself requests the information it needs and can then use it even in the same dialogue. The information that can be sought by a machine during the dialogue process is equivalent to that which a human would seek, and a large part of it is included in the following cases:

- location information
- time information
- attribute information (color, height, distance, etc.)
- issues of ambiguity
- unknown words
- issues of grammatical structure problems, due to idioms or replacement of sentence parts by expressions or movements.

As is known, a person receives much of this information from his wider interaction with the interlocutor, as well as from his intelligence that comes from a healthy brain. Parenthetically here, it is noteworthy to mention that, in general, the mechanism by which a human brain learns, perceives, synthesizes and uses knowledge through speech is complex and much research [4] is being done in many directions in modern science. In continuing, this raises the question: What happens when it comes to a machine, which by definition lacks the intelligence of a human brain but also the ability to perceive implied expressions and movements to understand ambiguous sentences? The answer is that in the case of the machine we can define a context in which, when it does not receive from its interlocutor the required information, since it will not be able to combine it with its current knowledge, it can directly target questions to it, until the full clarification.

According to the above, an attempt is made using the theory and hole semantics [5], the computational model OMAS-III [6] and the graduate thesis "Implementation of OMAS-III as a Grammatical Formalism for Robotic Applications" [7] and its related work [8], in order to create the algorithm that will compose the queries to the human/user.

The work is structured in four chapters. In the first chapter, reference is made to the theory and semantics of holes (Hole Semantics), to OMAS-III and how the combination of all of them can work in the synthesis of natural language. In the second chapter, reference is made to the use of OMAS-III as a grammatical formalism. In the third chapter, the construction of the natural language generation algorithm is done, as examples of its use. Chapter 4 presents the conclusions and suggestions for future research.

2. Theories and models

In this chapter, a brief presentation of the theory of holes as well as the OMAS-III systemic model and their connection for the further development of the study is made.

2.1 Theory and hole semantics

Hole theory, also known as "multilevel hole theory", is an approach to linguistics that focuses on the idea that language is structured at different levels of linguistic analysis, and each level operates independently. In this theory, the holes represent the different levels of language, such as phonetic, morphological, syntactic, and semantic. Each layer operates independently, but there is cooperation between them to create the overall meaning. That is, this theory emphasizes the interaction between these levels during language processing. Thus, by understanding the structure and function of each level, we can analyze how language is created and interpreted. So we can say that this approach helps to understand language as a complex system with various levels that interact, while at the same time maintaining their own autonomy. Hole semantics [9,10] is a framework that defines underdefined representations in arbitrary object languages [11]. Hole semantics constructs types of

an object language① (such as FOL② [12,13] or DRT③ [14]) with holes, into which other types can be attached. Each hole with a type (named by its label) and a connection is acceptable if it respects certain constraints.

2.2 OMAS-III

The Organizational Method of Analyzing Systems (OMAS) [6] is a diagrammatic technique of systems analysis and belongs to the category of general description techniques. The diagrammatic techniques of systems analysis were developed as tools of systemic thinking and visualization, providing a more complete and flexible way of describing the relevant concepts for each foreseeable field of application, which emphasizes the supervisory representation with the use of diagrams. OMAS-III is a designed process to achieve the best possible determination of the organization (structure and function/behavior) of an object or phenomenon (system), according to the application of basic organizational rules, adapted to specific conditions. OMAS belongs to the family of SADT④ and IDEFx techniques [15,16], being their design evolution. OMAS-III is the third improved version of the original method. A complete understanding of a system through this particular method requires answers to the unique seven fundamental questions concerning it:

- Why does it exist and work?
- What results and conclusions does it give?
- How much means (resources) does it need?
- How does it work?
- Who monitors or guides its operation?
- Where does it work?
- When does it work?

Understanding the system leads to its complete description or conversely, its complete description leads to the understanding of its structure (the structure of a whole) and its organization and operation (the arrangement concerning the relationships between its entities). These seven questions ("journalistic questions") constitute the basic assumption of a system, while the basic description of the system is made with the help of notation, implementing this assumption.

2.3 Connection of OMAS-III with hole semantic

According to the paper "Implementation of OMAS-III as a Grammatical Formalism for Robotic Applications" [7], in a minimalist language the word order should be SVO (subject-verb-object) and AN. "AN" means that words qualifying a noun (adjectives or adverbs), as well as its complements, should precede (the noun). In general, all words that qualify or complement any word, including relative noun clauses, must precede the main word. However, the Greek definite article/pronoun "TO" as well as the Greek indicative "AYTO" can be used as relative pronouns to introduce a relative clause after the verb. The reason why we have to follow the SVO and AN structure is that otherwise the language will not be minimalistic. The SVO structure is recommended for use in this language, but requires some indication to distinguish the subject from the object. Every language syntax is based on the concept "the first is the second", or "the first has the second", that is, the second word is the property of the first. Therefore, when we say "task easy", it should mean "task is easy". So we use the minimal meaning, without the need for conjunctions or articles. If we say "easy task", based on the same principle it should mean "this easy thing is a task", but now this information is not necessary. Thus, we understand "that easy thing which is a work" and more simply "an easy work".

3. OMAS as grammatical formalism

In this chapter, we will see how OMAS-III implements a grammatical formalism, so that the computer can understand sentences that arrive at the system.

① Object language is a language that is the object of study in various fields, such as logic, linguistics, mathematics, and theoretical computer science.

② FOL (First-order logic): Refers to logic in which the predicate of a sentence or statement can refer to only one subject. It is also known as first-order predicate calculus or first-order functional calculus.

③ In formal linguistics, Discourse Representation Theory (DRT) is a framework for investigating meaning under a formal semantics approach.

④ https://en.wikipedia.org/wiki/Structured_Analysis_and_Design_Technique

The following sections briefly present the study and application made in the work "Implementation of OMAS-III as a Grammatical Formalism for Robotic Applications" [7]. The flow diagram of the whole system is shown in **Figure 1**. The heart of the grammar formalism is in the box titled "Hypothesis 2", in the upper right of the diagram.

3.1 The grammatical formalism

In order to understand the OMAS-III model as a formalism tool, it is necessary for the system to answer the seven key questions, also known as journalist questions. The questions answer:

- the causality of the system (Why?);
- to the result including feedback (What?);
- in the introduction of included feedback (Which?);
- in the operating regulation conditions (How?);
- to those who oversee and guide operations (Who?);
- to the spatial aspects of functionality (Where?) and finally;
- on the temporal aspects of functionality (When?).

By answering all seven questions, our system receives all the information given to it. So, its ultimate goal is to get all the answers and if it can't do it in its "brain", then he externalizes its questions to get answers and understand the situation it is in, with the result that its intelligence also raises. More specifically:

- The question "Why" contains causal and explanatory factors (because, to). They are subordinate clauses and their answer is a supplementary clause with an explanation. It should be noted that it is not always given as a question, because knowledge is not required from the robotic system, as the robot only needs to recognize and accept it when it is given.
- The question "What" is recognized as an output of the system, and the answer is the verb used in the sentence where it is executed or was executed or will be executed, provided that the robot has detected the verb of the in-

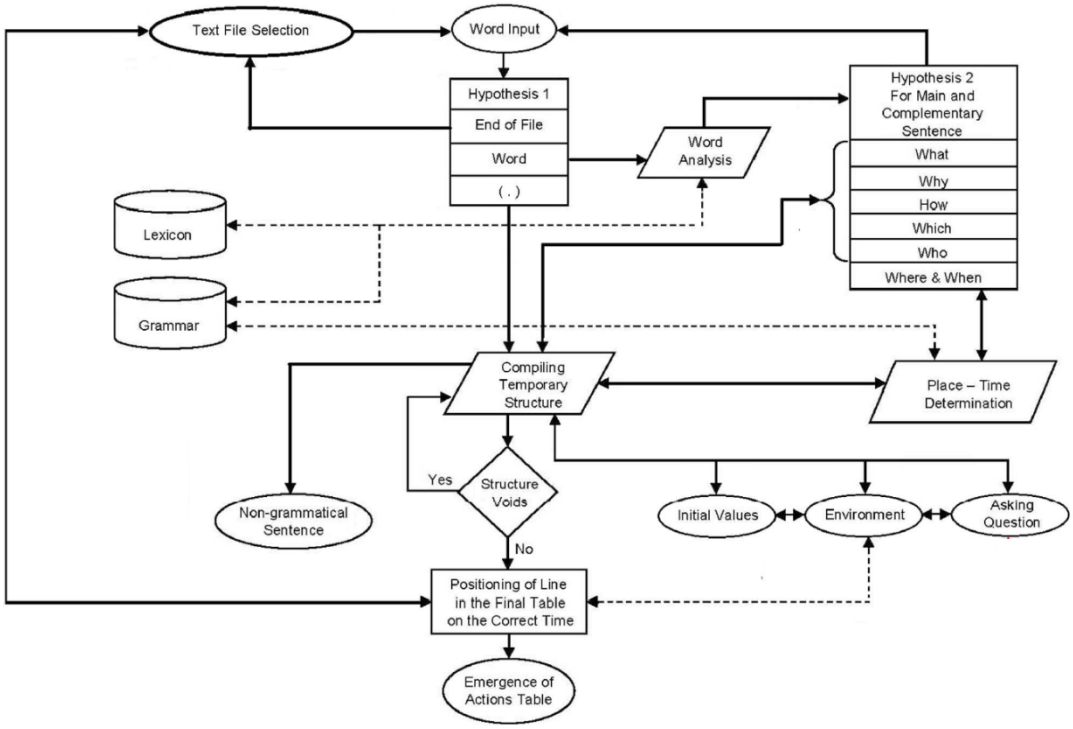


Figure 1. Basic algorithm.

Source: [8], after adaptation.

coming sentence. The purpose of the "What" question is to give the machine the ability to strip the words of the extras it may have received within the sentence and thus detect or not the verb there. If the verb does not locate it, it mainly finds the subjects or objects in the sentence and then asks what action it should do or what it is supposed to do.

- The question "Which/How much" contains all quantifiers and generally the objects of the sentence (adverbs are excluded). They are semantically placed in this position even though they do not indicate quantity, even though they are looking for what the verb should have.
- With the question "How", the action that will be used in the verb is given, with the help of modal adverbs, participles and general determinations that indicate manner.
- In the question "Who" the answer is the subject, provided the verb and object are clear in the sentence. In some cases the subject is omitted, as it may be embedded within the verb or implied. But if none of these cases applies, then the engine asks the question "Who". We usually create databases for cases like object recognition and hold the point in space and time it is meant to be found.
- When the question "Where" is asked and the place is not specified, then the robot will take for granted the current location, as it may have been defined in a previous sentence or a subsequent one of the current text. There is a chance that the machine perceives that it is a static point, with the result that the beginning and the end are identical. However, if the starting point is not given, then the robot will take as its location the previous location it was in the earlier sentence. If we don't give points or movement but ask it to be placed where someone else is, then if it recognizes someone, it takes that information and acts accordingly. It goes without saying that when the location is required but not given, the robot should not determine and ask.
- When "When" is asked, the time and moment when an action will take place is sought. Its most general form is indicated by the verb tenses and time determiners. This form is most often presented for the extended present, future and past. Along with the use of time, day and generally specific timing, it makes it easier to identify on the machine. By integrating a Real Time Clock (RTC) into a robotic system and at the same time with time management software, the machine would also experience the moments virtually. Except it would require more absolute time values than are given. For current needs, the time control given is predefined. If not declared by the data, then the robot approximates the time from the verb.
- Finally, there is the question "Why". In this particular case the machine will not have the mental capacity to give an answer, so it will not return the question again if it is not satisfied. Along with "Why" go the causal and explanatory words "because" and "to", with "to" having the role of purpose in language, but all three words are presented in subordinate clauses, where they will carry out the process retrospective.

In general, OMAS-III is a tool where the system thinks and visualizes, comprehensively, descriptions of concepts for predictable applications. It is the basic framework for building applications where robots are able to ask questions and provide information based on their existing and acquired knowledge.

3.2 Semantic grammars

Semantic grammars [17] consist of three steps:

- The primary is the semantics of the sentence accepted by the machine. Through a tree diagram, called a semantic interpreter, it derives the interpretation. It also contains conceptual dependency, where it represents language-independent concepts.
- The second step is the grammar features that have the pattern of frames and the act of unification to handle the semantic information.

The above two steps introduce artificial intelligence that makes the robot-machine capable of asking questions beyond statically recording language structures and relationships. The last step is constraint-based grammars [18], which contain hundreds of rules and are applied to multiple languages with a systematic success rate of over 99%.

3.3 Software design

The logic of the software design consisted of computational functions, where it would be permissible for the system to manage the data it would receive in the form of propositions, to perform detailed questions where necessary, and finally to expose lists of actions, ordered in time order, containing all details of who, where, when and how. All words found in the sentences are parsed and the results are stored in a temporary table. Nevertheless, the sentences that show deficiencies in objects and subjects, through a mechanism of clarification, are separated into necessary data. Clarifications are sought in already existing data of the text, however, in case no answer is found, then queries are executed. A necessary condition is the use of initial values, where they occupy the position of grammatical elements. By going back and completing the sentences, the temporary table is finalized, and all the data are transferred in time order to the time list of actions. The ultimate goal is for our system to offer the appropriate action that is requested and at the same time to determine in time the moment of its performance. The choice is made for the present, past and future tenses. In addition, in case the sentence consists of more specific temporal data, the action will be further characterized. The incoming data is passed through a six-option filter that sorts and characterizes it into words, according to the question it has to answer. It should be noted that if it is related to an explanation determination or is a supplementary proposal, then a corresponding process is activated in order to accept the new proposal. In this case, if the sentence does not contain all the features of grammar, based on the syntax of the language, then an analogous message is externalized to the outside world. On the other hand, if there are deficiencies in the structure of the sentence, then there is a two-way communication with the outside world. This has the effect of creating a temporary one-dimensional array, where the data of a sentence are expanded to construct a data line (i.e. a standardization of input data), with the ultimate goal of organizing the system to cope with what is asked of it. If there are any misses, then a temporary structure process takes place to search for data until the answer is negative, to place the line into a 2D output array, and finally give it to the outside world.

3.4 Complications

Through experiments and processing, some malfunctions appeared. The first was the involvement in endless processes, where a review was made of sentences that presented syntactic errors, and where it was possible to correct them under various conditions. The second difficulty was system-wide problems that could not be completely fixed. In general, the types of problems presented are either morphological (handling complex words), syntactic (determining the part of speech of words) or semantic, because the dictionary used each time is considered finite.

4. The NLG algorithm for Greek

In this chapter, we will see the algorithms based on which Natural Language Generation is done. In many modern methods [19], it is proposed that this process be carried out using neural networks through known techniques and their variants. In our developing system, the process of natural language generation is achieved by creating simple algorithms based on Greek grammar. These algorithms are in flowchart form. First, however, we will show how the perception and recognition of a word takes place, given a natural Greek language dictionary, as presented in the paper titled "Systemic and Whole Semantics in Human-Machine Language Interfaces" [20].

4.1 Creating word perception in the system

In a previous referenced work [7], an artificial language SostiMatiko was used. The concrete language

has a great advantage in relation to the morphology of the words. The root in each is grammatically invariant, for any part of speech and for any tense. Thus, a specific ending that is the same for all words determines whether we have a verb, an adverb, a noun or an adjective, singular or plural, subjunctive or imperative, and so on. In the case of natural language, however, this does not happen, at least for most of the cases. So the Greek word "ελα" for example (**Figure 2**: come_IMPERATIVE), is a verb in imperative and becomes "έρθω" in future (**Figure 2**: I will come), "έρχομαι" in present continue (**Figure 2**: to come), "ήρθα" in past (**Figure 2**: I came), etc. In English language the corresponding single word is "come". In the case of the Greek word, it becomes clear that in natural language the process becomes more difficult and the limited dictionary is imposed, since we need to have more information for the development of a source root. For the above reason, in the database each root that gives a series of words, changing only the endings, should have its own position. Any group of such roots that show the same root meaning should be linked to that meaning (**Figure 2:** Linking meaning-root-endings to form words). This way we find its tense and grammatical position and can change it accordingly to return a sentence. That is, if a command comes: "έλα εδώ τώρα" which means in English "come here now", the answer should be given: "έρχομαι εκεί τώρα" meaning "I come there now". Although this whole process goes beyond the scope of this work, let's make a small and simple report about the mechanism during the formation of words, in such a system. In the scheme of **Figure 2:** Linking concept-roots-ends to create words, we essentially observe three levels. Above is the general meaning of the word. Then we have three roots (ελ-, -ρθ-, ερχ-) that belong to this concept. For the root "-ρθ-" we have the development of two new roots, based on a phoneme "**ε**-" or "η-" placed before it. At the third level, there are all the endings that are attached to the roots to form a final word that defines a verb at a time to some person or persons, and so on.

This does not stop here, because the same roots, or others connected with the same meaning, give us adverbs and other parts of speech.

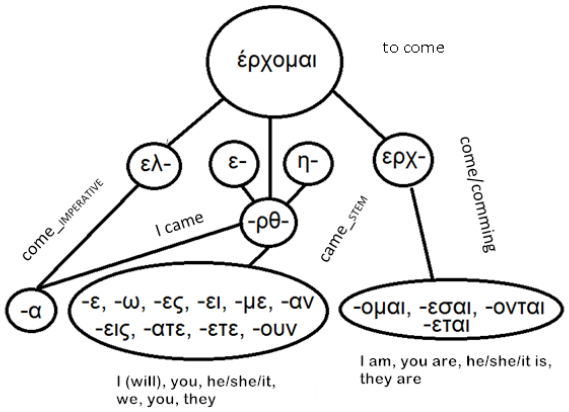


Figure 2. Linking meaning-root-endings to form words.

4.2 Word recognition algorithm

The word recognition algorithm is shown in **Figure 3**. The word recognition algorithm is described as follows:

- Each word enters the beginning of the algorithm.
- An object is initially created in which the grammatical characteristics of the word will be registered. They are registered in two variables, the length of the word and the length of the base with the roots, which is a dynamic element and can change during the operation of the system, learning new words [20].
- It looks to find which roots are shorter in length than the word. If it is not found, then the process stops and the sentence is not correct.
- If roots are found that are shorter than the length of the word, then a root that is contained in the word is searched for among them. If not found, then it will go back two steps and be rejected. If it is found, then the attributes of the word will be written to the object that was originally created, and the process will stop.

Based on the algorithm above, we check and identify all the words one by one. As long as all the words in the sentence are correct, their corresponding objects have been created. Thus, we have a com-

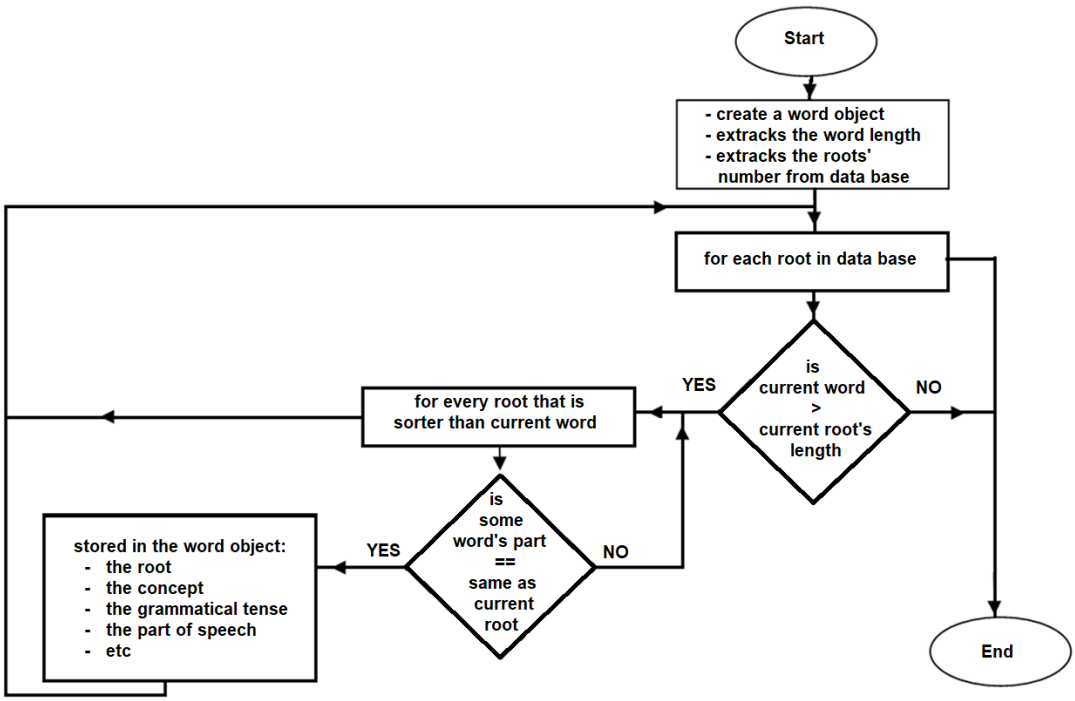


Figure 3. Word recognition algorithm.

plete mapping of the sentence, both morphologically and conceptually.

4.3 NLG algorithm for Greek

The creation of speech with a composition of natural Greek language is done after determining some basic elements. For example, the system must know all the grammatical features that will characterize the sentence, such as its person, verb and tense, as well as the object. The attributes combined with the predefined concept create the extracted sentence. Because it's easier to understand this with a fairly simple example, we'll use the queries generated by the system when it detects gaps in a sentence. If we only have the sentence "Πήγαινε και περίμενε", which means "Go and wait", then we detect two points of ambiguity. While the proposition is correct, the system needs to know where and when. The system has detected these two gaps and needs to ask questions. What else does it know? It knows that it has been given an order, i.e. imperative in the present tense. It begins to compose the questions. Since the second person becomes first, the imperative will become passive. So the meaning of "πάω" ("go" in English) from the verb form "Πήγαινε" (imperative of "go" in English) will become "πηγαίνω" ("I'm going", in English). The questions of where and when will come in front, and will be followed by "να" ("to" in English), because it is something I will do. Thus arise the questions "Πού να πάω;" ("Where should I go?" in English) and "Πότε να πάω;" ("When should I go?" in English).

This is exactly what is described in the algorithm that follows in **Figure 4** and is considered basic for natural language generation in this work.

In the case where there is no imperative, then there is no change of person and there is no addition of "NA" (to). Of course, if the word "θA" (will) exists, then it will remain. Such examples are the sentences: "O A περιμένει" (A is waiting) and "O B θα πάει" (B will go), which lead to the questions: "Πού περιμένει;" (Where is he waiting?) and "Πού θα πάει;" (Where will he go?) respectively.

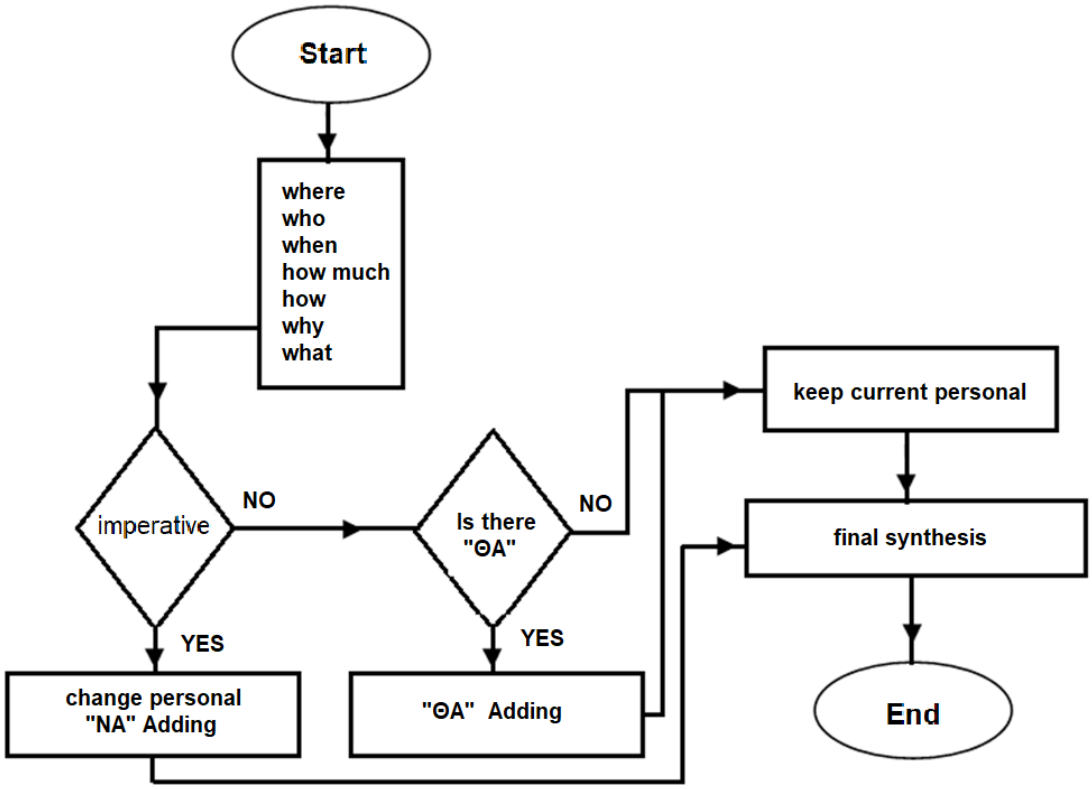


Figure 4. Basic NLG algorithm for Greek.

5. Conclusions

The following conclusions were drawn from this study:

- OMAS-III as a grammatical formalism can semantically and morphologically describe a sentence, which has been described again in the referenced work we used. In addition, with the same flexibility it is possible to compose a sentence that helps human-machine interaction through dialogue.
- This method of formalism enables us to easily intervene and add algorithms, enriching the already existing study system.

In all the studies so far on the specific techniques of OMAS-III and hole semantics, it seems that our developing system can easily accept upgrades by adding, relating or upgrading modules, such as speech synthesis, or understanding or learning a new word in the form of dialogue, etc. For this reason, we can simulate this system with various diseases of the brain, where missing a module can mean some encephalopathy. So, it is proposed as a topic of further study, the study of patients (with speech disorders), by studying problems that we will create in the OMAS-III formalism system, as it was examined in the referenced work and in the present work.

Author Contributions

All authors contributed equally to this work.

Conflict of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

References

[1] Giachos, I., Piromalis, D., Papoutsidakis, M., et al., 2020. A contemporary survey on intelligent human-robot interfaces focused on natural

language processing. International Journal of Research in Computer Applications and Robotics. 8(7), 1-20.

[2] Tellex, S., Gopalan, N., Kress-Gazit, H., et al., 2020. Robots that use language. Annual Review of Control, Robotics, and Autonomous Systems. 3, 25-55.
DOI: https://doi.org/10.1146/annurev-control-101119-071628

[3] Giachos, I., Papakitsos, E.C., Savvidis, P., et al., 2023. Inquiring natural language processing capabilities on robotic systems through virtual assistants: A systemic approach. Journal of Computer Science Research. 5(2), 28-36.
DOI: https://doi.org/10.30564/jcsr.v5i2.5537

[4] Lidz, J., 2023. Parser-grammar transparency and the development of syntactic dependencies. Language Acquisition. 30(3-4), 311-322.
DOI: https://doi.org/10.1080/10489223.2022.2147840

[5] Koller, A., Niehren, J., Thater, S. (editors), 2003. Bridging the gap between underspecification formalisms: Hole semantics as dominance constraints. Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics; 2003 Apr 12-17; Budapest Hungary. Stroudsburg: Association for Computational Linguistics. p. 195-202.
DOI: https://doi.org/10.3115/1067807.1067834

[6] Papakitsos, E., 2013. The systemic modeling via military practice at the service of any operational planning. International Journal of Academic Research in Business and Social Sciences. 3(9), 176.

[7] Giachos, I., 2015. Υλοποίηση της ΟΜΑΣ-ΙΙΙ ως Γραμματικού Φορμαλισμού για Ρομποτικές εφαρμογές (in Greek) [Implementation of OMAS-III as a grammar formalism for robotic applications] [Master's thesis]. Athens: National & Kapodistrian University of Athens and National Technical University of Athens.

[8] Giachos, I., Papakitsos, E.C., Chorozoglou, G., 2017. Exploring natural language understanding in robotic interfaces. International Journal of Advances in Intelligent Informatics. 3(1), 10-19.

[9] Bos, J., 1996. Predicate Logic Unplugged [Internet]. Available from: https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/25244/1/report_103_96.pdf

[10] Bos, J., 2002. Underspecification and resolution in discourse semantics [Ph.D. thesis]. Saarbrücken: Saarland University.

[11] Jumanto, J., Rizal, S.S., Asmarani, R., et al. (editors), 2022. The discrepancies of online translation-machine performances: A mini-test on object language and metalanguage. 2022 International Seminar on Application for Technology of Information and Communication (iSemantic); 2022 Sep 17-18; Semarang, Indonesia. New York: IEEE. p. 27-35.
DOI: https://doi.org/10.1109/iSemantic55962.2022.9920394

[12] Georgi, G., 2020. Demonstratives in first-order logic. The architecture of context and context-sensitivity. Springer: Cham. pp. 125-148.
DOI: https://doi.org/10.1007/978-3-030-34485-6_8

[13] Michalczenia, P., 2022. First-order modal semantics and existence predicate. Bulletin of the Section of Logic. 51(3), 317-327.

[14] Bos, J., 2021. Variable-free Discourse Representation Structures [Internet]. Available from: https://semanticsarchive.net/Archive/jQzMzJlY/compact.pdf

[15] Ross, D.T., 1977. Structured analysis (SA): A language for communicating ideas. IEEE Transactions on Software Engineering. (1), 16-34.
DOI: https://doi.org/10.1109/TSE.1977.229900

[16] Grover, V., Kettinger, W.J., 2000. Process think: Winning perspectives for business change in the information age. IGI Global: Hershey.

[17] Pereira, J., Franco, N., Fidalgo, R., 2020. A semantic grammar for augmentative and alternative communication systems. The 23rd International Conference on Text, Speech, and

Dialogue. Springer: Cham. pp. 257-264.
DOI: https://doi.org/10.1007/978-3-030-58323-1_28

[18] Bîlbîie, G., 2020. A constraint-based approach to linguistic interfaces. Lingvisticæ Investigationes. 43(1), 1-22.
DOI: https://doi.org/10.1075/li.00038.bil

[19] Chen, W., Chen, J., Su, Y., et al., 2020. Logical natural language generation from open-domain tables. arXiv preprint arXiv:2004.10404.
DOI: https://doi.org/10.48550/arXiv.2004.10404

[20] Giachos, I., Papakitsos, E.C., Antonopoulos, I., et al. (editors), 2023. Systemic and hole semantics in human-machine language interfaces. 2023 17th International Conference on Engineering of Modern Electric Systems (EMES); 2023 Jun 9-10; Oradea, Romania. New York: IEEE. p. 1-4.
DOI: https://doi.org/10.1109/EMES58375.2023.10171635

BILINGUAL PUBLISHING GROUP

Journal of Computer Science Research

https://journals.bilpubgroup.com/index.php/jcsr

SHORT COMMUNICATION

An Integrated Software Application for the Ancient Coptic Language

***Argyro Kontogianni*[1], *Evangelos C. Papakitsos*[2*], *Theodoros Ganetsos*[1]**

[1] *Department of Industrial Design & Production Engineering, University of West Attica, Egaleo, Athens, 12241, Greece*

[2] *Research Laboratory of Electronic Automation, Telematics and Cyber-Physical Systems, University of West Attica, Egaleo, Athens, 12241, Greece*

ABSTRACT

Coptic language was an important period of the Egyptian language, coinciding with a period of social and cultural changes. Coptic is also associated with the Greek language, as its alphabet is used for the transcription of Coptic. Despite the fact that the Coptic element is strong in Greece, the theoretical background is rather weak. For this reason, it is considered necessary to create a software tool that aims to help in the translation of Coptic into Greek and at the same time to overcome various obstacles that the researcher may encounter while processing the various corpora or artifacts, such as processing issuer, diacritics etc. This tool consists of a database, a search engine and an interface.

Keywords: Coptic software tools; Computer-assisted translation; Digital heritage

1. Introduction

Ancient Egypt has been a cradle of culture and Egyptian script is one of its best accomplishments. As language is a living and evolving organism, the Egyptian language was no exception to this rule and created a transitional continuum as it was affected by various political, economic and social changes. The last stage of these linguistic changes is the Coptic script, which is the last phase of a language that remained active from 3200 BCE until almost 1500 CE. When Alexander the Great conquered Egypt (332 BCE), the Greek language started supplanting Egyptian in documents of public administration. Thus, an impetus was given to abolish hieroglyphics and to incorporate into the script phonetic elements that aided the understanding of written texts. This transition from the Egyptian to Greek alphabet was

*CORRESPONDING AUTHOR:

Evangelos C. Papakitsos, Research Laboratory of Electronic Automation, Telematics and Cyber-Physical Systems, University of West Attica, Egaleo, Athens, 12241, Greece; Email: papakitsev@uniwa.gr

ARTICLE INFO

Received: 5 November 2023 | Revised: 22 November 2023 | Accepted: 30 November 2023 | Published Online: 8 December 2023

DOI: https://doi.org/10.30564/jcsr.v5i4.6068

CITATION

Kontogianni, A., Papakitsos, E.C., Ganetsos, T., 2023. An Integrated Software Application for the Ancient Coptic Language. Journal of Computer Science Research. 5(4): 38-42. DOI: https://doi.org/10.30564/jcsr.v5i4.6068

also prompted by the spread of the Christian religion in the Middle East, which desired a new script for its sacred texts rather than Demotic Egyptian which was directly associated with the existing pagan religion. So, Coptic, by the 5th century CE, gradually became the dominant script for secular and sacred texts. The six main dialects of Coptic (there are also many sub-dialects) are Sahidic, Bohairic, Fayyumic, Akhmimic, Lycopolitan, and Oxyrhynchite. Although each dialect has some separate linguistic features, some common features help researchers place a dialect in specific geographical areas [1,2]. When Egypt was conquered by the Arabs in 640 CE, its Islamization gradually began, naturally affecting the language as well. Nowadays the Christian communities of Egypt and the expatriate Coptic communities around the world use Coptic as a functional language [3].

2. Methodology

The Greek language had a decisive influence on the formation of Coptic. First of all, Egyptians adopted the Greek alphabet and used the Greek language for documents and administrative affairs. Furthermore, they transliterated their language by using the Greek alphabet. Thus, Egyptians enriched their alphabet with additional 8 signs, to represent the consonants that do not exist in the Greek language and created a script of 32 signs and 26 distinctive sounds. According to some researchers, Coptic language was created by Pantainos, director of the Catechetical School of Alexandria [4]. It is worth mentioning that the creation of Coptic was a way for the Egyptians to read their language as it happened to other populations (e.g. Slavic languages). Due to the strong influence of Greek on Coptic, there are various loanwords and Greek words that have been incorporated into the Coptic language: in the Biblical, Ecclesiastical, Liturgical, dogmatic, monastic and ascetic traditions, even in everyday speech. In this aspect, Coptic proves the transition from pagan Egypt to Christianity, as is the language of Christian religious texts, the language of the gospels, and the language of letters.

In Greece, Coptic element survives in museums and institutions. Manuscripts and artifacts in various materials of different durability, such as ivory, limestone, fabric, wood and clay be found on display or in archives at:

- the Byzantine and Christian Museum of Athens,
- the Benaki Museum in Athens,
- the Museum of Modern Greek Culture in Athens,
- the Peloponnesian Folklore Foundation "V. Papantoniou" in Nafplion,
- the Holy Monastery of Iveron on Mount Athos,
- the National Library of Greece in Athens.

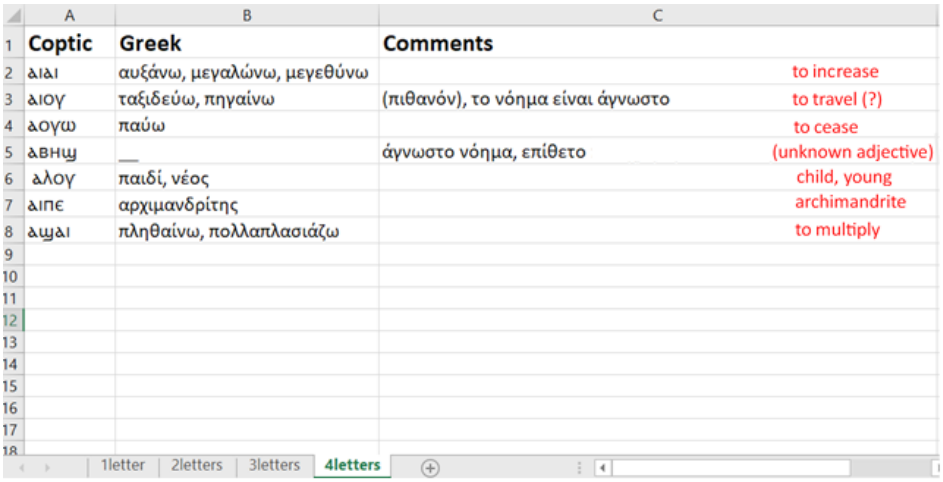
Despite these findings, the digital tools for Greek researchers are absent [1].

3. Results

Worldwide, there is some interest in creating tools for the Coptic language. One of the most remarkable works is Coptic SCRIPTORIUM, (created by Caroline T. Schroeder and Amir Zeldes [5]) where researchers can search Coptic dictionaries with translations to English, French and German, corpora, or use NLP service and tools for annotation etc. However, in regards to Greek scholarship, we have concluded that researchers are in need of a tool that will allow them to recognize the Coptic script and will be a useful aid for interpreting the texts and studying the language, and this is the aim of this software application presented herein. Moreover, Coptic writings appear on various artifacts, so this is meant to be a useful tool for Greek museums and heritage curators, with no or limited knowledge of Coptic. The intention of the development of this tool is also to overcome various obstacles that may occur, like processing issues, absence of spaces between words, use of diacritics, punctuation, abbreviations etc. So, the semi-automated approach allows interfering with the artifacts without effort and risk of damaging them.

This software tool consists of three parts:

i) *The database*, is practically the Coptic-Greek digital dictionary and one of the major parts of our tool. The database is an Excel file, instructed on a single spreadsheet (**Figure 1**) to be modified or enriched



	A	B	C	
1	**Coptic**	**Greek**	**Comments**	
2	ⲁⲓⲁⲓ	αυξάνω, μεγαλώνω, μεγεθύνω		to increase
3	ⲁⲓⲟⲩ	ταξιδεύω, πηγαίνω	(πιθανόν), το νόημα είναι άγνωστο	to travel (?)
4	ⲁⲟⲩⲱ	παύω		to cease
5	ⲁⲃⲏⲱ	__	άγνωστο νόημα, επίθετο	(unknown adjective)
6	ⲁⲗⲟⲩ	παιδί, νέος		child, young
7	ⲁⲓⲡⲉ	αρχιμανδρίτης		archimandrite
8	ⲁⲱⲁⲓ	πληθαίνω, πολλαπλασιάζω		to multiply
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				

1letter | 2letters | 3letters | 4letters

Figure 1. A sample of the Coptic-Greek digital dictionary.

easily. Coptic dictionaries were used, available both in printed form and online (such as Crum [6] and Coptic Dictionary Online [7]). The Coptic words are sorted into lists, by size (according to the number of their letters) and then alphabetically in each separate list. Each spreadsheet includes three columns: The first column is used for the word in Coptic, the second one for translation into Greek and the third one for comments considered necessary (e.g., original source, dialect, or anything we consider important for the user).

ii) *The search process*, will be done in two ways. First, there is a Cartesian dictionary in which the words were arranged in size and alphabetically doing a linear search. Secondly, another base will be created in order to be used for a weighted linear search, a process based on Zipf's law [8]. It is also sequential, like the previous one, but it is executed in tables with the data in terms of their probability of occurrence, in descending order (preloading) [9]. For creating the frequency tables, Scriptorium corpora will be used. It will provide us with 39 separate texts for reading, analysis and complex searches.

Their subject matter is quite long like magic papyri, the *Book of Ruth*, manuscripts, the *Gospel of Mark*, the *Assumption of John*, various biographies, etc. The benefits of dual search are great. It will achieve faster and more accurate search results, while at the same time, it will provide the necessary conclusions on how to rank databases for faster search with the most reliable results.

iii) *The interface*, is very simple and user friendly (**Figure 2**). On the left side, we have the Coptic al-

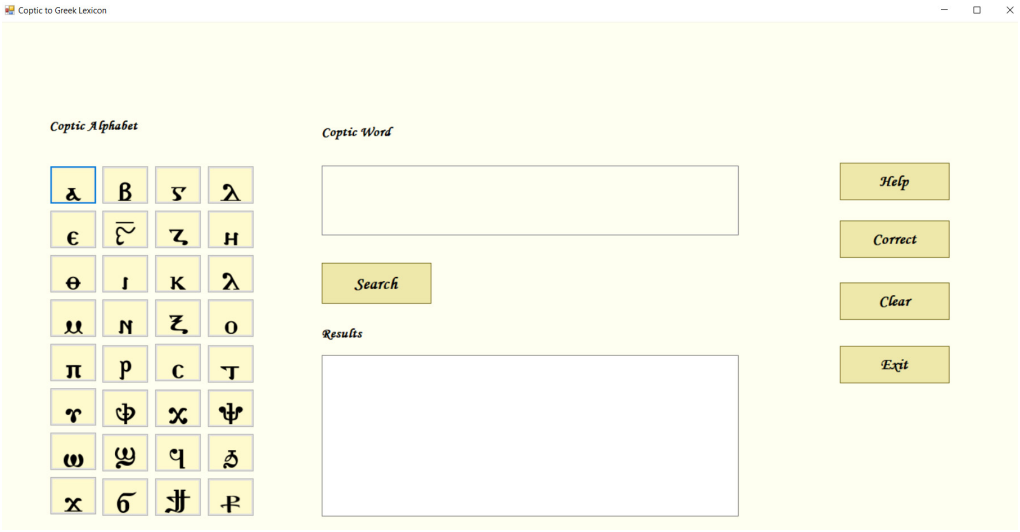


Figure 2. The interface of the computer-assisted translation software.

phabet. The users choose the letters that they see on their artifact or on their papyrus. Then they click the "Search" button or the "Correct" button in the case of a mistake. The "Results" text-box, returns the Greek translation and the corresponding comments, if the formed word exists in the dictionary. Otherwise, a failure message will be displayed. When the user clicks the box "Clear" he can move on to the next word search, or the box "Help" for further information. Finally, with the "Exit" button, the user can save all the work in a "txt" file.

4. Conclusions

Although the Coptic community in Greece is particularly active [10], the Greeks seem to ignore the Coptic as an ethno-religious community and the same applies to the plethora of their artifacts which have not received the attention of the majority of Greeks. Furthermore, the theoretical background for Coptic is limited and the software tools related to this language for Greek researchers are, currently, non-existent. This tool is an excellent way of digitizing the Coptic heritage, since the Coptic script is written on such fragile materials and artifacts, that could be potentially destroyed, and human intervention is imperative. Although there are other tools for the Coptic language, none of them translate Coptic to Greek. Moreover, this tool could help to avoid the physical interference of human-artifact and diminish the possibility of damage. Finally, it is part of a wider range of software tools, which have been developed for processing ancient languages [11,12] and are still being developed under the auspices of the University of West Attica [13-15], in order to study the ancient languages and digitize cultural heritage.

Author Contributions

All authors contributed equally to this work.

Conflict of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

References

[1] Kontogianni, A., Ganetsos, T., Kousoulis, P., et al., 2021. Computer-assisted translation of Egyptian-Coptic into Greek. Journal of Integrated Information Management. 5(2), 26-31. DOI: https://doi.org/10.26265/jiim.v5i2.4470

[2] Kontogiani, A., Ganetsos, Th., Papakitsos, E.C. (editors), 2022. A software tool for Egyptian-Coptic language. 8th Balkan Symposium on Archaeometry; 2022 Oct 3-6; Belgrade, Serbia.

[3] Kontogianni, A., Ganetsos, T., Zacharis, N., et al., 2021. A detailed study about Egyptian-Coptic and software engineering. Archaeology. 9(1), 28-33. DOI: https://doi.org/10.5923/j.archaeology.20210901.06

[4] Fourtounas, I., 2012. Η Ελληνικότης της Κοπτικής Γλώσσας—Πάνταινος, ο Έλληνας Δημιουργός της Κοπτικής (Greek) [The Greekness of the Coptic language—Pantainos, the Greek Creator of Coptic]. Athens.

[5] Schroeder, C., Zeldes, A., n.d. University of Oklahoma & Georgetown University [Internet] [cited 2021 Jun 20]. Available from: https://copticscriptorium.org/

[6] Crum, W.E., 2005. A Coptic dictionary. Wipf and Stock: Eugene, OR.

[7] Coptic Dictionary Online [Internet] [cited 2021 Mar 17]. Available from: https://coptic-dictionary.org/

[8] Papakitsos, E.C., 2013. Φυσική Γλώσσα & Υπολογιστικά Μαθηματικά (Greek) [Natural Language & Computational Mathematics] [Master's thesis]. Athens: Linguistic Computing of the National & Kapodistrian University of Athens and the National Technical University of Athens.

[9] Papakitsos, E.C., 2014. Γλωσσική Τεχνολογία

Λογισμικού: ΙΙ. Πραγμάτωση (Greek) [Linguistic software engineering: II. Realization]. National Library of Greece: Athens.

[10] Daskaloudis, N., 2023. Η κοπτική κοινότητα στην Ελλάδα, το ζήτημα της Θρησκευτικής ταυτότητας σε ένα νέο περιβάλλον (Greek) [The Coptic community in Greece: The issue of religious identity in a new environment]. Thessaloniki: Aristotle University of Thessaloniki.

[11] Kontogianni, A., 2014. Ερευνητικό Λογισμικό: Ανάπτυξη Ηλεκτρονικού Λεξικού Γραμμικής Β΄ (Greek) [Research software: Development of an electronic dictionary of linear B] [Master's thesis]. Athens: Linguistic Computing of the National & Kapodistrian University of Athens and National Technical University of Athens.

[12] Kontogianni, A., Papamichail, C., Papakitsos, E.C., 2017. Εκπαιδευτικό Λογισμικό για την Εκμάθηση της Γραμμικής Β΄ (Greek) [Educational Software for Learning Linear B] [Internet]. Available from: http://events.di.ionio.gr/cie/images/documents17/cie2017_Proc_OnLine/new/custom/pdf/4.08.CIE2017_259_Kontog_Final%20(1)_P.pdf

[13] Kontogianni, A., Gkanetsos, T., Papakitsos, E.C. (editors), 2019. Ψηφιοποίηση Έργων Πολιτιστικής Κληρονομιάς: Η Περίπτωση της Γραμμικής Β΄ (Greek) [Digitization of cultural heritage works: The case of Linear B]. The Proceedings of the 3rd Panhellenic Conference on Digitization of Cultural Heritage (EuroMed 2019); 2019 Sep 25-27; Egaleo, Greece. Available from: https://www.euromed-dch.eu/wp-content/uploads/2020/11/%CE%A0%CE%A1%CE%91%CE%9A%CE%A4%CE%99%CE%9A%CE%91-2019-v5.pdf

[14] Mavridaki, A., Galiotou, E., Papakitsos, E.C. (editors), 2020. Designing a software application for the multilingual processing of the linear a script. Proceedings of the 24th Pan-Hellenic Conference on Informatics (PCI 2020); 2020 Nov 20-22; Athens, Greece. New York: ACM.
DOI: https://doi.org/10.1145/3437120.3437299

[15] Mavridaki, A., Galiotou, E., Papakitsos, E.C., 2021. Developing a software application for the study and learning of linear a script. Review of Computer Engineering Research. 8(1), 8-13.
DOI: https://doi.org/10.18488/journal.76.2021.81.8.13

BILINGUAL PUBLISHING GROUP

Journal of Computer Science Research
https://journals.bilpubgroup.com/index.php/jcsr

REVIEW

Data, Analytics, and Intelligence

Zhaohao Sun

Department of Business Studies, PNG University of Technology, Lae 411, Papua New Guinea

ABSTRACT

We are living in an age of big data, analytics, and artificial intelligence (AI). After reviewing a dozen different books on big data, data analytics, data science, AI, and business intelligence (BI), there are the current questions: 1) What are the relationships between data, analytics, and intelligence? 2) What are the relationships between big data and big data analytics? 3) What is the relationship between BI and data analytics? This article first discusses the heuristics of the Greek philosopher Plato and French mathematician Descartes and how to reshape the world. Then it addresses the above questions based on a Boolean structure, which destructs big data, data analytics, data science, and AI into data, analytics, and intelligence as the Boolean atoms. Data, analytics, and intelligence are reorganized and reassembled, based on the Boolean structure, to data analytics, analytics intelligence, data intelligence, and data analytics intelligence. The research will analyse each of them after examining the system intelligence. The proposed approach in this research might facilitate the research and development of big data, data analytics, AI, and data science.

Keywords: Big data; Big analytics; Business intelligence; Artificial intelligence; Data science

1. Introduction

Big data, artificial intelligence (AI), and data science have been critical for academia and industries. We are living in an age of big data, data analytics, and AI. Big data has become one of the most important frontiers for innovation, research, and development in the computer industry and business [1-4]. Big data has also been a key enabler in exploring business insights, business intelligence (BI), and the economics of services. This has drawn an unprecedented interest in industries, universities, governments, and organizations [5,6]. Data analytics has also played a vital role in BI and management activities [7,8]. Market-oriented AI, big data-based AI, and BI have

*CORRESPONDING AUTHOR:
Zhaohao Sun, Department of Business Studies, PNG University of Technology, Lae 411, Papua New Guinea; Email: zhaohao.sun@pnguot.ac.pg; zhaohao.sun@gmail.com

ARTICLE INFO
Received: 8 November 2023 | Revised: 3 December 2023 | Accepted: 7 December 2023 | Published Online: 14 December 2023
DOI: https://doi.org/10.30564/jcsr.v5i4.6072

CITATION
Sun, Zh.H., 2023. Data, Analytics, and Intelligence. Journal of Computer Science Research. 5(4): 43-57. DOI: https://doi.org/10.30564/jcsr.v5i4.6072

become the fiercest competition in the world [9,10], such as chips, 5G and 6G. ChatGPT, driverless cars, TikTok, the Internet of Things (IoT), and artificial drones have made us immerse in the era of big data and AI. However, the following fundamentals of data, analytics, and intelligence are still open for comprehending and exploring big data, AI, and data science:

1) What are the relationships between data, analytics, and intelligence?

2) What are the relationships between big data, big knowledge, and big intelligence?

3) What is the relationship between business intelligence and analytics, as well as data intelligence?

This research will use the Boolean structure to address each of the issues based on the graduality of data, analytics, and intelligence as well as a systematic analysis. The Boolean structure destructs big data, data analytics, data science, and AI into data, analytics, and intelligence as the Boolean atoms. Data, analytics, and intelligence are then reorganized and reassembled, based on the Boolean structure, to data analytics, analytics intelligence, data intelligence, and data analytics intelligence.

The rest of this article is structured as follows: Section 2 reviews a dozen different books on big data, analytics, data science, and artificial intelligence. Section 3 looks at the heuristics of Aristotle and Descartes. Section 4 discusses how to reshape the world by shaping the wood based on a story. Section 5 explores data, analytics, and intelligence using a Boolean structure. Section 6 provides related work and discussion. The final section ends this article with some concluding remarks and suggestions for future work.

2. Book reviews for data, analytics, and intelligence

There are many books on each of the topics, using the research of Google gooks and Amazon's books. A dozen different books the author has used for his teaching and research recently are illustrated in **Table 1**.

Table 1. Book reviews for data, analytics, and intelligence.

Items	Books
Data	[6], [10], [11], [12]
Analytics	[6], [11], [12], [13], [14], [15], [16], [17], [18], [19]
Intelligence	[6], [10], [11], [13], [14], [15], [20], [21]
Data analytics	[6], [11], [12], [17], [18], [22], [23]
Data intelligence	[9], [11], [22], [23]
Analytics intelligence	[6], [13], [14], [15].
Data analytics intelligence	1. Analytics: Business Intelligence, Algorithms and Statistical Analysis [19]. 2. Artificial Intelligence, Analytics and Data Science: Volume 1 Core Concepts and Models [24].

In what follows, we will examine each of them:

Database Systems: Design, Implementation, and Management is a classic textbook in the world that focuses on data (including big data), database systems, and management [6]. However, it lacks the investigation into analytics and intelligence, data intelligence, and data analytics intelligence.

Laudon & Laudons look at management information systems by investigating data and intelligence [10]. However, they lack investigation on relationships between data intelligence and data analytics intelligence from data, analytics and intelligence.

Artificial Intelligence: A Modern Approach is a classic textbook [9]. In the past AI was a knowledge-based intelligence, it is also a data-driven intelligence technique. However, it investigates the AI using an agent approach. The book has not investigated analytics and intelligence, as well as analytics intelligence and much more. Further, the agent approach in the book can be also replaced by other approaches like a multi-industry approach.

Aroraa, et al. investigate data and analytics [18]. From a structural viewpoint, Aroraa, et al, have not classified the principles, tools, and practices for mentioned topics such as database management

systems, data warehouse, BI, data visulization, big data, machine learning, and data analytics. In other words, their book has not really discussed data analytics, although the book is titled *Principles, Tools, and Practices for Data Analytics*. This book is similar to the book of Sharda, et al. [6] and the book of Weber [11], because Chapters 4 and 5 of the book are similar to Chapter 2 of the book of Sharda, et al. [6]. Chapter 2 of this book is similar to Chapter 3 of the book of Sharda, et al. [6]. The feature of this book is that it provided more techniques for big data; Chapters 6 and 7 look at the introduction to big data and Hadoop, NoSQL and MaPreduce. It also provided more applications of big data and machine learning. Even so, Aroraa, et al. have not detailed an investigation into data analytics, data intelligence, and analytics intelligence.

Brooks investigates BI and analytics from concepts, techniques, and applications [13]. This book consists of five chapters: Chapter 1 Introduction to BI; Chapter 2 Understanding of Business Analytics; Chapter 3 Data for BI and Analytics; Chapter 4 Operational Intelligence, and Chapter 5 Role of Analytics for business decision making. However, this book lacks a detailed investigation into BI and analytics. This book also lacks related topics on various kinds of BI and business analytics at a deep level. Even so, Brooks investigates BI but also more intelligence on business, that is, market intelligence and location intelligence are important. This leads to significant questions:

1) What is the difference between BI and other intelligence such as market intelligence?

2) What is the interrelationship between intelligence and analytics?

3) What is the minimum intelligence and maximum intelligence?

Thompson & Rogers focus on analytics based on their understanding of how to win with intelligence [15]. They first provide the competitive advantage stemming from analytics in Chapter 1, understanding advanced analytics in Chapter 2, the age of the algorithm economy in Chapter 3, the modern data systems in Chapter 4, and then study a few business cases. Their discussion on operational analytics (in Chapter 7) is similar to Chapter 4, Operational Intelligence of Brooks's book [13]. However, they have not investigated BI and analytics, as well as data intelligence. The challenging question for these two books is: What is the relationship between analytics and intelligence?

EMC provides big data analytics, data analytics lifecycle, and advanced analytical theory and methods using R [17]. From a research methodological perspective, this book classifies the advanced analytical theory and methods into clustering, association rules, regression, classification, visualization and report of data and more; the latter has been considered as the functions of data mining [25]. However, EMC has no interest in implementing data intelligence and analytics intelligence and their interrelationships.

Sharda, et al. focus on BI, analytics, and data science [6,14]. They look at data, big data, BI, and analytics. In particular, they investigated techniques and tools in analytics (descriptive, predictive, and prescriptive analytics). The good strength of the book is that the techniques and tools in analytics have been structured excellently. This is the first book for processing analytics using this structure. However, they have not processed diagnostic analytics [26]. They also have not investigated the relationship between BI and analytics.

Ghavami investigated big data analytics methods: analytics techniques in data mining, deep learning, and natural language processing [16]. Ghavami first looks at Part 1: big data analytics, consisting of a data analytics overview in Chapter 1, basic data analysis in Chapter 2, and data analytics process in Chapter 3, and then Part 2: advanced analytics methods, consisting of natural language processing in Chapter 4, qualitative analysis in Chapter 5, advanced analytics and predictive modelling in Chapter 6, ensemble of models in Chapter 7, machine learning and deep learning in Chapter 8, and model and optimization in Chapter 9. The book also provides case studies. Therefore, from a methodological perspective, Ghavami used models and methods to discuss data and big data. He also classifies big data

analytics and then advanced analytics methods. This book is a detailed data/analytics method for BI and analytics [14,6]. Even so, this book has a good structure from a structuralist and modelling viewpoint.

From a technique and application viewpoint, Wade looks at advanced analytics in Power BI with R and Python [22], and Pittman uses Google Analytics and GA4 to improve online sales by better understanding customer data [23]. Both can be considered a technique and applications for understanding data, analytics, and intelligence. Both have an impact on the techniques of data analytics and BI.

There are two books on intelligence. Hawkins provides a new theory on intelligence [21]. Bostrom looks at paths, dangers, and strategies towards superintelligence, which is a kind of meta intelligence [20]. Both books will be useful for investigating data, analytics, and intelligence, for example, Hawkins's human intelligence [21], Bostrom's collective superintelligence and digital intelligence [20] are vital for intelligence.

Artificial Intelligence (AI), Analytics, and Data Science written by Chee Chew Hua in 2020 focuses on real data and real problems instead of purely mathematical constructs, although AI, analytics and data science use mathematics to solve real problems [24]. The main chapters of this book encompass The main chapters of this book encompass chapter 3: data exploration and summaries, chapter 4: data structures and visualisation, chapter 5: data cleaning and preparation, chapter 6: linear regression, chapter 7: logistic regression, chapter 8: classification and regression Tree (CART), chapter 9: neural network, chapter 10: strings and text mining. In other words, this book mainly covers data structures, data cleaning and preparation, and visualization as the first part: data organization and visualization. The book covers linear regression, logistic regression, classification and regression tree as the second part: statistics, and this book covers neural network and text mining as a part of machine learning. Relatively, this book focuses on computations over mathematical proof, statistics are the basis for this book, because it is a textbook for students in the areas, maybe not for students of AI, overall, this book is a theory-flavored different from other books mentioned next [11,18,19]. All these mentioned books are not related to statistics, because they are free from mathematics and mathematical formulas.

This book, titled *Big Data Analytics: Applications in Business and Marketing*, covers 4 parts below after its introduction [27]:

1) Applications of business analytics consists of chapter 2: big data analytics and algorithms, chapter 3: market basket analysis: An effective data mining technique for anticipating consumer purchase behavior, chapter 4: customer view variation in shopping patterns, chapter 5: big data analytics for market intelligence, chapter 6: advancements and challenges in business applications of SAR images, and chapter 7: exploring quantum computing to revolutionize big data analytics for various industrial sectors.

2) Business intelligence consists of 2.Business Intelligence consists of chapter 8: evaluation of green degree of reverse logistic of waste electrical appliances, chapter 9: nonparametric approach of comparing company performance: a Grey relational analysis, chapter 10: applications of big data analytics in supply chain management, chapter 11: evaluation study of churn prediction models for business intelligence.

3) Analytics for marketing decision-making consists of Analytics for marketing decision making consists of chapter 12: big data analytics for market intelligence, chapter 13: data analytics and consumer behavior, chapter 14: marketing mode and survival of the entrepreneurial activities of nascent entrepreneurs, chapter 15: the responsibility of big data analytics in organization decision-making, chapter 16: decision making model for medical diagnosis based on some new interval neutrosophic Hamacher power choquet integral operators.

4) Digital marketing consists of Digital Marketing consists of chapter 17: prediction of marketing by consumer analytics; chapter 18: web analytics for digital marketing., chapter 19: smart retailing: a novel approach for retailing business, chapter 20: leveraging web analytics for optimizing digital mar-

keting strategies, chapter 21: smart retailing in digital business, and chapter 22: business analytics and performance management in India.

The last two books can be considered as a new type of publishing, open free publishing, different from traditional publishers. Weber looks at data and big data, data science, intelligence, AI, advanced AI, machine learning, analytics and cyber security [11], although he does not discuss analytics intelligence. His book is an introduction to all the mentioned topics. There are also a lot of good ideas in his discussion, although he contributed a number of good ideas in this book on big data, data science, AI and cyber security. However, the lack of references damaged the quality of this book.

Blatt provides each of the elements for analysis, BI, algorithms, and statistical analysis [19]. He understands that analysis has become an extremely important aspect to consider when you are thinking of starting any new line of business or even when it comes to purchasing a new house in today's world, although his book focuses on analytics rather than analysis: BI, Algorithms, and Statistical Analysis. At least, what is the relationship between analysis and analytics? It is still an issue for the author and many other readers. There are not any references in the book, which also damaged this book [19]. Even so, his discussion on descriptive, predictive, and prescriptive analytics [19] can provide a bitter understanding of data analytics as a classification of analytics.

Overall, all these books have not processed data, analytics, intelligence and their relationships properly or logically. For example, how to generalize and specialize their topics based on the research methodology is still a topic. All these books ignore the relationships between data, information, knowledge, intelligence and wisdom. They also ignore analytics and intelligence, and intelligent analytics with applications in business and other fields including management and decision-making. All these books' thoughts, technologies, and methods need a systematic integration based on a Boolean structure (see later) and our existing research sections. Therefore, the following research issues should be addressed:

1) What is the relationship between data, analytics, and intelligence?

2) What is the relationship between BI and Analytics?

3) What is the relationship between BI, Analytics, and data intelligence?

4) How to generalize and specialize data science, big data, BI and AI based on the Boolean research methodology?

3. The heuristics of Plato and Descartes

The Republic was written by an ancient Greek philosopher Plato [28] about justice, order, character and the man of the republic. A few years ago, my friend told the author that he would like to build a republic and become a president. As a scholar, the author cannot create a republic. However, the author can write an article and publish a book. In fact, writing an article and publishing a book, similar to creating a republic, must follow a set of rules, referencing rules, writing rules, publishing rules, formatting rules, templating rules, and communicating rules. Some articles and books are also required to have a research methodology consisting of a set of research rules.

Descartes is a great French mathematician. It is he who introduced analytical geometry that let the author know how to integrate algebra and geometry. It is he who gave the author a better understanding of analytics. Descartes is a great man not only because of his profound knowledge of analytical geometry, but his book on his research methodology titled *Discourse on the Methods* [29]. The author does not have a lot of knowledge and skill in data science, AI, or computer science although he has been working in these areas for a few decades. However, this research tries to use a new research methodology and ideas, just like Descartes, to create a new research methodology for a new article and a new book.

4. Reshaping the world

When the author was very young. His dream was to become a carpenter. He bought a very expensive

saw, plane, chisel, nail, axe and ruler to make a table, similar to the existing table around him. His idea was to reshape wood although he could not reshape the world. Yes, he was very happy that he made a table, which made his parents also very happy. His neighbors and villagers helped him to become a master of reshaping wood. However, he did not continue this way, and instead, he continued to study and took the national examination for universities and changed my way completely.

Now, he became a scholar and drove to an Australian furniture shop to buy a box of tables made in China. After he came back to his home, he opened the box and reassembled the table based on the instruction: how to reassemble the table using a provided screwdriver.

The author asked the boss of the table factory how to make a table. The boss told him that this was a reshaping of wood by:

1) Design a table and draw a table blueprint.

2) Disassemble the table into a table plate, table leg, nails, furniture cam lock and nut.

3) Procure a table plate, table leg, furniture cam lock and nut, nails, and booklet for reassembling the table from different factories using the Internet and putting them into a box.

4) Advertise the information of the table to all the world using the Internet.

5) Sell all the boxes of tables to the world using the Internet.

Therefore, disassembling, procuring, and selling are important tasks of the table factory.

This is a kind of reshaping wood towards mass production based on big data, business analytics, and the Internet. In fact, more than 40% of the furniture is made in China in 2022 by these furniture factories. The Internet and big data have been playing an important role in meeting the furniture requirements of the world.

We cannot destroy the existing world. However, we must smash and reorganize human living conditions and everything such as commodities, organizational structures, design and art, education and development in the world. The rules of the world and new algorithms are about to change the world completely, to reorganize things accordingly based on advanced technologies such as AI, big data, and digital technology. They will basically overturn the foundation of the whole world that has been established for half a century since the inception of digital computers in 1946. The computer infrastructure is based on chips for CPUs (a central processing unit) and AI GCUs of NVIDIA (https://www.wired.co.uk/article/nvidia-ai-chips), and big data as the core of AI computing. This article aims to smash and reorganize AI, big data, data analytics, and business intelligence and reorganize them using first a Boolean structure and then reorganize the new atoms and internal components of this Boolean structure using a new algorithm to smash and reorganize computing such as data computing, information computing, knowledge computing, intelligence computing, and wisdom computing.

5. Overview on data, analytics, and intelligence

What is intelligence? It is related to patterns, knowledge discoveries, and insights for decision-making. Even so, wisdom is still more important than intelligence. We are at the trinity age of data, analytics, and intelligence.

We are in the age of AI and big data. AI including its natural language processing (NLP) and big data has been applied to almost every sector and has been revolutionizing our work, lives and societies [30]. ChatGPT is an example of NLP.

Big data does not have very big value without big data analytics, just as oil without the significant progress of the petrochemical industry [31]. However, the commercial value of big data becomes bigger and bigger with the processing, deep processing, smart processing, and intelligent processing of big data. Big data analytics is behind processing, deep processing, second-time processing, and multi-processing... of big data. Therefore, big data analytics is more important than big data, and intelligent big data analytics is at the core of this age of trinity and will become a disruptive technology for the age of Trinity

in terms of healthcare, web services, service computing, cloud computing, IoT (the Internet of Things) computing, and social networking computing [2].

6. Big data, analytics, and intelligence: A Boolean structure

This research will use the Boolean structure to address each of the issues based on the graduality of data, analytics, and intelligence as well as a systematic analysis of related books and journal articles and the research methodology such as a meta-processing, systematic generalization and specialization of existing publications, as illustrated in **Figure 1**. It will provide multi-industry applications in business, management and decision-making based on the Boolean structure. All these are treated using an integrated approach. This Boolean structure destructs the existing world of the books mentioned in section 2. For example, AI, BI, big data analytics, and data science [14,6] are destructed into data, analytics and intelligence as the three atoms of Boolean structure. The data, analytics, and intelligence are reorganized and reassembled based on Boolean structure to data analytics, analytics intelligence, and data intelligence and then data analytics intelligence. In such a way, some of the mentioned books have ignored some of them, for example, analytics intelligence and data analytics intelligence, technologically. Accordingly, big data analytics intelligence has not been discussed in the mentioned books.

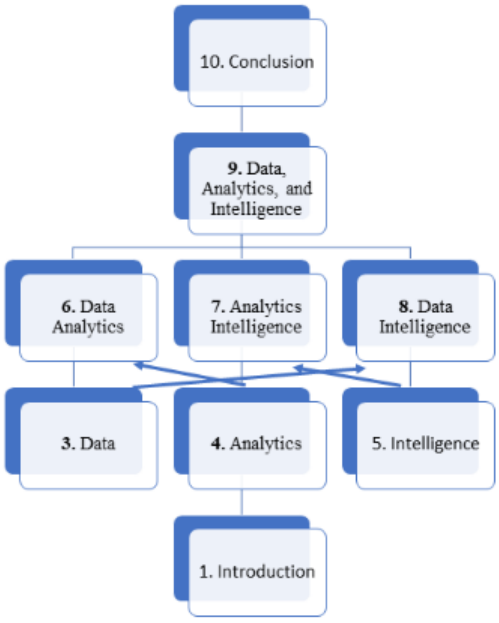


Figure 1. Data, analytics and intelligence: A Boolean structure.

Based on the Boolean structure, the introduction is the basis. The three atoms of the Boolean structure are composed of data, analytics, and intelligence. Data, analytics, and intelligence will be represented as three composite Boolean expressions, that is, data analytics, analytics intelligence, and data intelligence. Finally, data, analytics and intelligence have been represented as the Boolean expression: data analytics intelligence. In other words, this research (as a book under contract with CRC Press and Tayler Francis in the USA), based on a Boolean structure, is composed of the following chapters [32]:

1) Introduction
2) Data
3) Analytics
4) Intelligence
5) Data analytics
6) Analytics intelligence
7) Data intelligence
8) Data, analytics and intelligence
9) Conclusions

In what follows, we will explore each of them (except for the introduction and conclusion), corresponding to a chapter in the book, based on an integrated approach.

This research uses computing, science, technology, systems, management, services, and applications to explore each of the above terms listed in the Boolean structure. For example, it will explore data computing, data science, data engineering, data technology, data systems, data management, data services, and data applications [33]. It will demonstrate that data engineering aims to use data science and data technology to develop and manage data systems to provide data intelligence with data system products and services [33].

6.1 Data

All data and big data are important for ruling the world. Data have become an important element

of the economy [32]. One hundred years ago, big companies dominated steel, oil, and manufacturing companies. Recently, big data companies have ruled the whole world. Apple, Alphabet, Meta (formerly Facebook), Amazon, Microsoft, Tencent, and Alibaba have dominated the whole world. Big data as a disruptive technology is transforming how we live, study, work, and think [32]. Therefore, this research first looks at data, and then big data. It identifies and examines ten big characteristics of big data with an example for each: big volume, big velocity, big variety, big veracity, big data technology, big data systems, big Infrastructure, big value, big services, and big market [31]. It also explores a service-oriented foundation of big data and calculus of big data. Besides data and big data, this research also explores not only data, but also information, knowledge, and intelligence (DIKI) are important for ruling the world. Then it explores DIKI computing, science, Engineering, technology, systems, management, and services [33]. For example, data computing = data science + data engineering + data technology + data systems + data Management + data service. Finally, this research provided big trends in the era of big data, that is, we are embracing the era of big data. Countries around the world have drawn increasing attention to the research and development of big data since 2012 [26]. Big data industries have been booming in the world. These six big trends consist of the informatization of big data, mining big data for big knowledge, mining big data for big intelligence, networking of big data, socialization of big data, and commercialization of big data [32]. This research also discusses the interrelationships of data from three different viewpoints: data science, big data and artificial intelligence (AI). The research demonstrates that big data is the raw material that will be transformed into information, knowledge, intelligence, networking, society and big market using ICT and digital technology, data science, and AI. These six big trends will bring about big industries, smarter cities, smarter societies and smarter countries. Overall, data in general and big data in particular are a foundation of big information, big knowledge, big intelligence and big analytics. Therefore, we are still at the foundational stage, and enter the emerging age of big information, big knowledge, big intelligence, and big analytics.

6.2 Analytics

We are in the age of analytics although it has been around us for about a century [15,34]. Analytics is a science and technology of using mathematics, AI, computer science, data science and operations research to provide practical applications to business, management, research and development, economic and societal problems. Analytics can be defined as a process of understanding and exploring data by creating meaningful patterns and insights [19]. Analytics is one area that requires complex algorithms [15,9]. One of the most important parts of analytics is data visualization [19,18]. After discussing the evolution of analytics, this research looks at analysis ≠ analytics, and examines mathematical analytics, a system process of analytics. This research provided types of analytics based on a few perspectives, one of them is that analytics can be classified into descriptive statistics, diagnostic statistics, predictive statistics, and predictive statistics (DDPP analytics) based on cyclic business operations. This research explores analytics algorithms and models, and analytics computing, for example, analytics computing = analytics science + analytics engineering + analytics technology + analytics systems + analytics management + analytics services + analytics intelligence. Analytics engineering aims to use analytics science and analytics technology to create and manage analytics systems to provide analytics services with analytics intelligence [33]. Finally, this research analyzes three major trends of analytics, that is, the rise of analytics scientists, and algorithms becoming more and more important for the Algorithm industry, and the blossoming analytics industry and then all analytics including analytical approach, analytical modelling, analytical analysis, analytical metrics are central for ruling the world.

6.3 Intelligence

All data, big data, and analytics are all for intelligence [32]. Intelligence is not only a lasting topic for computer science, AI, intelligence computing, BI, and intelligent analytics, but also an exciting topic for industries, organizations, and businesses [35]. AI has facilitated the development of intelligent services, intelligent manufacturing, intelligent systems [9], and intelligent analytics [34]. BI has promoted the improvement of competitiveness of business and marketing performance, supported management decision-making of organisations, and produced trillion level enterprises such as Google, Amazon, and Meta [3,36]. However, the current AI is a very mixed intelligence, and very market-driven. A lot of companies brand their products as AI products. Many social media brands do a lot of things as AI products and services. Even so, this research looks at the fundamentals of intelligence including basic intelligence, how can calculate intelligence? Then it explores Intelligence 1.0, Intelligence 2.0, and Meta AI with six levels of intelligence. It examines multi-intelligence and intelligence of the five senses. This research provides a meta-approach to a hierarchy of data and intelligence including meta (DIKI) and a meta-approach to intelligent systems. This research demonstrates that, meta (data) = information; meta (information) = knowledge, meta (knowledge) = Meta^3 (data) = intelligence; Meta^4 (data) = meta (intelligence) = mind, Meta^5 (data) = meta (mind) = wisdom. After overviewing intelligence in AI, this research explores wisdom and mind, from AI to artificial mind, cloud intelligence, data intelligence, and similarity intelligence. It provides an integrated framework of intelligence to a DIKW Intelligence where DIKW is the abbreviated form of data, information, knowledge and wisdom. This research also examines the age of meta intelligence as competing in the digital world.

6.4 System intelligence

This leads to a new question, that is, how can we calculate system intelligence? This research explores temporality, expectability, and relativity of intelligence as three system intelligences [35].

Temporality of intelligence

There are two meanings for temporal intelligence. 1) Temporal intelligence is the ability to adapt to change. This has motivated the development of temporal logic and evolutionary computing including genetic algorithms [9]. 2) Temporality of intelligence means that intelligence is related or limited to a time interval [35]. For example, at the time of writing this section, few people consider floppy disks as intelligent storage devices. However, a few decades ago floppy disks were considered intelligent in comparison to paper tape for data storage. In what follows, we limit ourselves to the meaning of item 2.

Expectability of intelligence

Intelligence can be referred to as a substitution for easier, faster, smarter, friendlier, more efficient, more satisfactory. This is the expectability of intelligence. We denote them using the degree of satisfaction. All these related concepts are a set of expectations of humans, as parts of human intelligence. Some aim to become billionaires. Some like to become the president of a country to provide services to the people of the country. Others aim to become a CEO of top companies in the world. Different people have really different expectations. We denote these expectations for a system or product P, as $E_P = \{e_i \,|e_i \textit{ is an expected performance for function}_i \textit{ of a product}\} = \{e_i \,|i \in \{1, 2, ..., n-1, n\}$, where n is a given integer. For every $i \in \{1, 2, ..., n-1, n\}$, there is a perceived performance of the customer for *function*$_i$, p_i, then a product P is intelligent if and only if there exists at least one $i \in \{1, 2, ..., n-1, n\}$ such that [35]:

$$S_i = \frac{p_i}{e_i} \geq 0 \tag{1}$$

where s_i is the satisfaction degree of the customer to the i^{th} function of system P.

For example, a Huawei Mate 60 Pro smartphone, cost CN¥6,599 ($923.99) as a launch price in August, 2023, with BDS satellite calling and message and 5G telecommunication is smarter. "Smarter" is what the

customer perceived, $p_1 = 1.5$, while "smart" is an expected performance, $e_1 = 1$, for Huawei Mate 60 Pro from a customer, based on Equation (1), we have the satisfaction degree of the customer $s_i = 1.5 > 0$. Then a Huawei Mate 60 Pro smartphone is intelligent.

Relativity of intelligence

If one lives in the data world, one will find information = metadata (i.e., meta (data)) is a result of meta intelligence. However, if one lives in the information world, one cannot have such knowledge. Further, if one lives in the information world, one will find knowledge = meta (information) (i.e., meta (information)) is a result of meta intelligence. However, if one lives in the knowledge world, one cannot have such a vision. Therefore, one has relativity of intelligence and relativity of meta intelligence. This also means that meta is relative.

Generally speaking, let X and Y be two systems. X is intelligent if X is better than Y with respect to E, where E is a set of human expectations. "Better" is a relativity concept. For example, a new microwave is intelligent because it displays the temperature when microwaving food. A user believes that displaying the temperature is better than not displaying it. This example reflects the relativity of intelligence. Displaying temperature belongs to the set of expectations E. The set of human expectations can be considered as a set of demands. The expectation of human beings and society promotes intelligence and social development. Therefore, it is significant to define the intelligence of systems with respect to the set of human expectations or demands.

In summary, system intelligence can be measured through three dimensions: temporality, expectability, and relativity. In other words, there are three characteristics of system intelligence: temporality, expectability, and relativity. The degree of intelligence of a system product or service can be measured using this triad, that is:

$$\begin{aligned}\text{degree of system intelligence} = {} & \text{temporality} \\ & + \text{expectability} \\ & + \text{relativity}\end{aligned} \quad (2)$$

Equation (2) is more useful for system intelligence and big data intelligence, because they are based on performance, business advantages, competitive advantages of systems products or services. All of these are closely associated with temporality, expectability, and relativity of system intelligence. This formula can be realized by using big data analytics and big data, in other words, big data and big data analytics can generate big data intelligence, for short,

$$\text{big data intelligence} = \text{big data} + \text{big data analytics} \quad (3)$$

Equation (3) indicates that increases in either big data or big data analytics can improve big data intelligence. This is partially proved by what Professor Peter Norvig, Google's Director of Research, said: "We don't have better algorithms; we just have big data" [37].

Temporality, expectability, and relativity of system intelligence can be considered fundamental for BI including organization intelligence, enterprise intelligence, marketing intelligence, big data intelligence, analytics intelligence, and data analytics intelligence [13]. We will explore them in the next subsection.

6.5 Data analytics

Data analytics might be the oldest among all types of analytics. Data have become the new oil and gold of the 21st century. Data analytics mines the data from data sources such as data warehouses and data lakes for new knowledge and meaningful insights [16]. Data analytics is at the heart of business and decision-making [6], just as data analysis is at the heart of decision-making in almost real-world problem-solving [38]. This research first discusses the fundamentals of data analytics. Then it explores the classification of data analytics. It explores the fundamentals of big data analytics and advanced analytics platforms. The research examines big analytics covering big information analytics, big knowledge analytics, big wisdom analytics, and big intelligence analytics. Then this research discusses data science covering database systems, data warehousing, data mining, data computing and data analytics computing. Finally, this subsection will explore data analytics and big data analytics with applications in busi-

ness, management, and decision-making.

6.6 Analytics intelligence

Analytics intelligence is about how to use analytics to win intelligence [15]. Strategically, analytics intelligence is an intelligence that is derived from analytics systems. This research looks at analytics intelligence and intelligence analytics, as well as DIKW analytics intelligence and big DIKW analytics intelligence. It overviews generative intelligence. The research explores analytical intelligence as the core of AI and generative intelligence not only in academia and the market. The research demonstrates that the earlier analytical intelligence was from logical AI, and then symbolic AI. This period has lasted till the inception of the Internet. However, big data has been booming from 2012 and onwards. Data analytics and big data analytics have become an important part of business analytics, BI, and intelligent analytics. What is the key to data analytics? It is analytic. How can we use analytic methods and techniques to process data and big data, information and big information, knowledge and big knowledge intelligently? This is the analytical intelligence underpinned by data analytics, big data analytics, and big analytics. Therefore, we can work on intelligent data analytics and intelligent big data analytics, both are used to develop analytical intelligence in terms of business and society.

Mathematically [35],

$$\text{Analytics} = \text{analysis} + \text{SM} + \text{DM} + \text{DW} + \text{ML} + \text{visualization} \quad (4)$$

where SM, DM, DW, and ML are abbreviated forms of statistical modeling, data mining, data warehouse, and machine learning. Therefore, using intelligence as a right operation to both sides of the above equation, we have:

$$\begin{aligned}\text{Analytics intelligence} = {} & \text{analysis intelligence} \\ & + \text{SM intelligence} \\ & + \text{DM intelligence} \\ & + \text{DW intelligence} \\ & + \text{ML intelligence} \\ & + \text{visualization intelligence}\end{aligned} \quad (5)$$

The research examines big data analytics intelligence with applications. Finally, the research discusses the spectrum of intelligent analytics.

6.7 Data intelligence

Data intelligence is the analysis of various forms of data in such a way that it can be used by companies to expand their services or investments future [39]. This research introduces data intelligence by addressing the following research questions: What is the fundamental of data intelligence? What are the applications of data intelligence? After reviewing backgrounds and related work, this research analyzes data as an element of intelligence, and looks at data and knowledge perspective on intelligence including information intelligence and knowledge intelligence. It examines big intelligence not only with data, but DIKW intelligence through proposing an integrated framework of intelligence. This research presents the fundamentals, impacts, challenges, and opportunities of data intelligence in the age of big data, AI, and data science. This research also presents a unified framework for not only data intelligence. This research also looks at the age of meta intelligence for competing in the digital world. Finally, this research explores big data 4.0 as the era of big intelligence we are living in. There are at least two contributions to the academic communities. 1) The research demonstrates that data intelligence is the basis for knowledge intelligence, which is a core of artificial intelligence. 2) Big data 4.0 = big intelligence will play a critical role in our organisations, economies, and societies.

6.8 Data, analytics, and intelligence

Google Web and Google Scholar search and summarize their popularity for each of analytics on data, information, knowledge, and wisdom. This is a kind of analytics on data, information, knowledge, and wisdom (retrieved on July 26, 2023).

From **Table 2**, Google Scholar implies that data analytics and information analytics are similar in academia, while Google Web means that information analytics plays a more important role than data

analytics in academia and industry although data analytics are very popular in the big data and business intelligence world [6,26]. Knowledge analytics and wisdom analytics have also played an important role in academia and industry although they are not popular in the business intelligence world [10,13]. This implies that not only data analytics but also information analytics, knowledge analytics, and wisdom Analytics (DIKW analytics) have played a critical role in computer science, AI, and data science as well as business and management. Therefore, this research looks at data analytics intelligence and big data analytics intelligence. It explores DIKW computing, analytics, and intelligence. The research presents a cyclic model for big data analytics and intelligence. The research provides the calculus of intelligent data analytics: elements, principles, techniques, and tools for business analytics and business intelligence. Finally, the research provides fundamentals of business analytics and discusses the relationships of business analytics, digital analytics, and business intelligence and applications of big data analytics intelligence.

Table 2. Data analytics, information analytics, knowledge analytics, and wisdom analytics.

Analytics	Google Web	Google Scholar
Data analytics	1,970,000,000	4,390,000
Information analytics	2,630,000,000	4,560,000
knowledge analytics	894,000,000	3,170,000
Wisdom analytics	36,400,000	114,000

7. Discussion

This section will discuss the related work on data, analytics, and intelligence. It also examines the limitations of this research.

A Google Scholar search for "data, analytics, and intelligence" found about 11,000,000, 4,980,000, and 4,500,000 results, respectively (retrieved on December 2, 2023). This implies that data, analytics, and intelligence have become significant topics for the research of scholars and researchers. A Google search for "data, analytics, and intelligence" found about 18,600,000,000, 5,650,000,000, and 2,780,000,000 results respectively (retrieved on December 2, 2023). This means that data, analytics, and intelligence have significantly influenced our lives, communities, economies, and societies. Therefore, data, analytics, and intelligence are still a topic for us to explore and develop in the age of digital technology.

This article first reviewed a dozen different books on big data, data analytics, data science, artificial intelligence, and business intelligence that have been used for the author's teaching and research in the past few years. Some authors use textbooks such as *Database Systems: Design, Implementation, and Management* [40], *Management Information Systems: Managing the Digital Firm* [10], *Business Intelligence, Analytics, and Data Science: A Managerial Perspective* [14], and *Artificial Intelligence: A Modern Approach* [9] to apprehend data, analytics and intelligence. Other authors use *Big Data and Artificial Intelligence: Complete Guide to Data Science, AI, Big Data, and Machine Learning* [11]; *Business Intelligence and Analytics: Concepts, Techniques and Applications* [13]; *Analytics: Business Intelligence, Algorithms and Statistical Analysis* [19] and *Data Analytics: Principles, Tools and Practices* [18] to look at the relationships of data, analytics, and intelligence. Some authors also provide case studies for how to data and analytics to win with intelligence [15]. Other authors use *Google Analytics and GA4: Improve Your Online Sales by Better Understanding Customer Data and How Customers Interact with Your Website* in order to understand data, analytics, and intelligence [23]. Different from above mentioned books and related publications, this article uses the motivation of the heuristics of Greek philosopher Plato and French mathematician Descartes, a story for reshaping the world and the Boolean structure to destroy big data, data analytics, data science, AI into data, analytics, and intelligence as the Boolean atoms. Then data, analytics, and intelligence are reorganized and reassembled, based on the Boolean structure, to data analytics, analytics intelligence, data intelligence, and data analytics intelligence. Then this article examines each of the eight Boolean elements in depth.

A limitation of this research is that it should pro-

vide a deeper investigation into data, analytics, and intelligence in order to provide more rationales for each of them with multi-industry applications. Another limitation of this research is that connecting (or connection) should be another component of human intelligence. Connectivity should be another element of system intelligence. Advanced communication technologies and tools such as mail, telephone, email, social media, and information sharing on the Web aim to develop the skill of connecting as a form of system intelligence [32]. For example, the current advanced ICT technology and systems (have brought about social networking services such as Facebook, LinkedIn, and WeChat [10]). All these have developed the skill of connecting as a part of system intelligence. we will add connectivity intelligence as the fourth of human intelligence and system intelligence in future work.

8. Conclusions

This research first reviews a dozen different books on big data, data analytics, data science, artificial intelligence, and business intelligence and discusses the heuristics of Greek philosopher Plato and French mathematician Descartes and how to reshape the world. The key scientific methodology and tool is destructing, reorganizing, and reassembling to reshape the world of big data, data analytics, data science, artificial intelligence, and business intelligence. This article uses the Boolean structure as a scientific tool to destruct big data, data analytics, data science, AI into data, analytics, and intelligence as the three Boolean atoms. Then data, analytics, and intelligence are reorganized and reassembled, based on the Boolean structure, to data analytics, analytics intelligence, data intelligence, and data analytics intelligence. The article analyses each of them after examining the system intelligence. Corresponding to the above key scientific methodology and tool, this article and book will use engineering method to discuss each of the chapters. For example, one of the key contributions of this article is that data (analytics) engineering aims to use data science and data technology to develop and manage data systems to provide data intelligence with data system products and services [32].

This research is based on the early book publication proposal submitted to CRC Press Florida and Tayler Francis Delaware, USA. The book has been under contract with the mentioned company. This article and this book hope that ideas, methods, and techniques with applications help readers prepare critical data, analytics, and intelligence for the changing uncertainties ahead, not only to understand the knowledge of them.

In future work, as an extension of future research directions and our research of this article, we will delve into real world cases such as the IoT including the Internet of People (IoP) and the Internet of Services (IoS), and ChatGPT to further verify big intelligence where we are living in. We will also develop a unified framework for analytics thinking and analytics intelligence to support big intelligence.

Conflict of Interest

There is no conflict of interest.

Acknowledgement

This research is supported partially by the Papua New Guinea Science and Technology Secretariat (PNGSTS) under the project grant No. 1-3962 PNGSTS.

References

[1] Chen, C.P., Zhang, C.Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on big data. Information Sciences. 275, 314-347.
DOI: https://doi.org/10.1016/j.ins.2014.01.015

[2] Laney, D., Jain, A., 2017. 100 Data and Analytics Predictions through 2021 [Internet] [cited 2018 Aug 4]. Available from: https://www.scribd.com/document/569847512/100-data-and-analytics-predictions-through-2021

[3] Howarth, J., 2023. 30+ Incredible Big Data Statistics (2023) [Internet] [cited 2023 May 3].

Available from: https://explodingtopics.com/blog/big-data-stats#top-big-data-stats

[4] Kumar, B., 2015. An encyclopedic overview of 'big data' analytics. International Journal of Applied Engineering Research. 10(3), 5681-5705.

[5] Products in Analytics and Business Intelligence Platforms Market [Internet]. Gartner [cited 2023 Jun 12]. Available from: https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms

[6] Sharda, R., Delen, D., Turban, E., 2018. Business intelligence and analytics: Systems for decision support (10th edition). Pearson: Boston, MA.

[7] The Evolution of Big Data Analytics Market [Internet]. 3PILLARGLOBAL [cited 2021 Aug 29]. Available from: https://www.3pillarglobal.com/insights/the-evolution-of-big-data-analytics-market/

[8] Kronz, A., Schlegel, K., Sun, J., et al., 2022. Magic Quadrant for Analytics and Business Intelligence Platforms [Internet]. Gartner. Available from: https://www.gartner.com/en/documents/4012759

[9] Russell, S.J., Norvig, P., 2020. Artificial intelligence: A modern approach (4th edition). Prentice Hall: Upper Saddle River.

[10] Laudon, K.C., Laudon, J.P., 2020. Management information systems: Managing the digital firm (16th edition). Pearson: Harlow, England.

[11] Weber, H., 2020. Big data and artificial intelligence: Complete guide to data science, AI, big data, and machine learning. Independently Published.

[12] Hurley, R., 2019. Data science: A comprehensive guide to data science, data analytics, data mining. Artificial intelligence, machine learning, and big data. Independently Published.

[13] Brooks, S., 2022. Business intelligence and analytics: Concepts, techniques and applications. Murphy & Moore Publishing: New York, NY.

[14] Sharda, R., Delen, D., Turban, E., et al., 2018. Business intelligence, analytics, and data science: A managerial perspective (4th edition). Pearson: London.

[15] Thompson, J.K., Rogers, S.P., 2017. Analytics: How to win with intelligence. Technics Publications: Basking Ridge, NJ, USA.

[16] Ghavami, P., 2020. Big data analytics methods: Analytics techniques in data mining, deep learning and natural language processing (2nd edition). de Gruyter: Boston/Berlin.

[17] EMC Education Services, 2015. Data science and big data analytics: Discovering, analyzing, visualizing and presenting data. Wiley: Hoboken.

[18] Aroraa, G., Lele, C., Jindal, M., 2022. Data analytics: Principles, tools and practices. BPB: New Dalhi, India.

[19] Blatt, T.J., 2017. Analytics: Business intelligence, algorithms and statistical analysis (5th edition). Todd Blatt: Orlando, FL, USA.

[20] Bostrom, N., 2014. Superintelligence: Paths, dangers, strategies. Oxford: London, UK.

[21] Hawkins, J., 2021. A thousand brains: A new theory of intelligence. Basic Books: New York.

[22] Wade, R., 2020. Advanded analytics in Power BI with R and Python. Apress: New York.

[23] Pittman, C., 2022. Google Analytics and GA4: Improve your online sales by better understanding customer data and how customers interact with your website. SMP Publishing: Kingsport, Tennessee.

[24] Hua, C.C., 2020. Artificial intelligence, analytics, and data science. Blue Kingfisher Ltd.: Hong Kong.

[25] Kantardzic, M., 2011. Data mining: Concepts, models, methods, and algorithms. Wiley & IEEE Press: Hoboken, NJ.

[26] Sun, Z., Stranieri, A., 2021. The nature of intelligent analytics. Intelligent analytics with advanced multi-industry applications. IGI-Global: Hershey. pp. 1-22.

[27] Chaudhary, K., Alam, M., 2022. Big data analytics: Applications in business and marketing. Auerbach Publications: New York.

[28] Plato, Jowett, B., (translator), 2022. The Republic. Independently published.

[29] Descartes, R., 1637. Discourse on the methods. GlobalGrey: London.

[30] Strang, K.D., Sun, Z., 2022. ERP Staff versus AI recruitment with employment real-time big data. Discover Artificial Intelligence. 2(1), 21. DOI: https://doi.org/10.1007/s44163-022-00037-1

[31] Sun, Z., Strang, K., Li, R. (editors), 2018. Big data with ten big characteristics. Proceedings of the 2nd International Conference on Big Data Research; 2018 Oct 27-29; Weihai, China. New York: Association for Computing Machinery. p. 56-61. DOI: https://doi.org/10.1145/3291801.3291822

[32] Sun, Z., Wang, P.P., 2021. The scientification of China. Cambridge Scholars Publishing: Cambridge.

[33] Sun, Z., 2022. Problem-based computing and analytics. International Journal of Future Computer and Communication. 11(3), 52-60.

[34] Sun, Z., 2019. Managerial perspectives on intelligent big data analytics. IGI Global: Hershey.

[35] Sun, Z., Sun, L., Strang, K., 2018. Big data analytics services for enhancing business intelligence. Journal of Computer Information Systems. 58(2), 162-169. DOI: https://doi.org/10.1080/08874417.2016.1220239

[36] Sabherwal, R., Becerra-Fernandez, I., 2011. Business intelligence: Practices, technologies, and management. John Wiley & Sons, Inc.: Hoboken, NJ.

[37] McAfee, A., Brynjolfsson, E., Davenport, T.H., et al., 2012. Big data: The management revolution. Harvard Business Review. 90(10), 60-68.

[38] Azvine, B., Nauck, D., Ho, C., 2003. Intelligent business analytics—a tool to build decision-support systems for eBusinesses. BT Technology Journal. 21, 65-71. DOI: https://doi.org/10.1023/A:1027379403688

[39] Data Intelligence [Internet]. Techopedia [cited 2021 Oct 26]. Available from: https://www.techopedia.com/definition/28799/data-intelligence

[40] Coronel, C., Morris, S., Rob, P., 2020. Database systems: Design, implementation, and management (14th edition). Course Technology, Cengage Learning: Boston.

BILINGUAL PUBLISHING GROUP
Tel:+65 65881289
E-mail: contact@bilpublishing.com
Website:https://journals.bilpubgroup.com

2630-5151 04
9 772630 515230