A New Approach for Visual Analytics Applying to Multivariate Data of Student Intakes in the University

Dang Van Pham[1,2,3(✉)], Vinh Cong Phan[1], and Nam Hoang Do[1]

[1] Faculty of Information Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
pvdang.tps@gmail.com, {pvdang,pcvinh,namdh}@ntt.edu.vn
[2] Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam
[3] Institute of Applied Mechanics and Informatics, Ho Chi Minh City, Vietnam

Abstract. Human always has the ability to view things and phenomena, but to understand about it that depends on the ability to perceive as well as combine with existing experiences and knowledge which is available in every human to understand it. This research studies the human vision organs, infographics capture characteristics of the human eye, human vision awareness progress, and human visual viewing - thinking mechanisms to design visual analytics system of multivariate data (mdVAS), model of viewing - thinking visually analyzing, and visual graphs for discovering knowledge from multivariate data (mD). This mdVAS will support to upgrade human vision awareness progress that includes stages such as viewing - perceiving - cognizing - analyzing - understanding - remembering ($vpcaurH_s$). This progress, especially the analysis stage plays a key that opens the door for human to understand the profound significance of things. Humans who want to understand the profound insights of things must bring out questions, at that time the mdVAS will enhance knowledge for humans to answer questions that humans themselves brought out. This approach will help humans, especially leaders, who have a profound insight into mD. This research illustrates experiments on educational data to manage the student intakes to help leadership on policy and decision making.

Keywords: Human vision awareness progress · Visual analytics system of multivariate data · mdVAS · Model of viewing - Thinking visually analyzing · Visual graph · Educational data

1 Introduction

Due to human demand wanting to discover the profound insights hidden within mD, human has used many methods in combination with different tools to collect data from various data sources in the real-world. Therefore, human is facing enormous challenges in extracting information and discovering knowledge from within these data sources. The

C. V. Phan and T. D. Nguyen (Eds.): ICTCC 2022, LNICST 473, pp. 88–120, 2023.
https://doi.org/10.1007/978-3-031-28790-9_7

data is analyzed and converted into many different formats to display on a 2D computer screen, but it depends on discovering purposes of every human. At that time, human will represent it in whatever forms to suit their purpose, such as visual graphs, images, symbols, etc. to help human perceive more and more significance as deep inside data as possible. This is the basis for answering the question, why do humans understand so quickly some forms of data representation and others don't? This question is extremely important for data scientists who are pursuing the design of mdVAS that exploits mD.

For having the formats as above, there are two approaches to transform data into information and knowledge in mdVAS that are approach according to the mathematics and visualization methods [1–3]. With the mathematics method, system discovered knowledge from data by providing algorithms to answer analytical questions but without the participation of users. With the visualization method, human is reputed the most important component in mdVAS to answer analytical questions through experiences and knowledge available in every human to perceive visually. Human looks at visual graphs representing data to perceive the deepest significance about data from which to support of policy making and makes decision.

Human always has the ability to view, deduce, remember based on deductive mechanisms, remind memory and with the great operation of the human vision organs (hVO_s) that the Creator bestows on human, it is also a means for human to think, create, accumulate knowledge and is able to contemplate the colorful beauty of nature. This research studies the hVO_s, human vision awareness progress ($hVAP_s$), infographics capture characteristics of the human eye ($hIGCE_s$) and human visual viewing - thinking mechanisms ($hVTM_s$) to design mdVAS that services users for extracting information and discovering knowledge from mD. Vision is the most important sensory organ, the channel that receives the most data. On those bases, this research proposes to make mdVAS that human is able to use and connect components as model of viewing - thinking visually analyzing (visVTM), computer science methodologies to support visual viewing – thinking, questions and answers, how to fit the visual graphs that support to human adapt to the $hVAP_s$ that includes stages such as $vpcaurH_s$ to extract information and discover knowledge from mD.

The remainder of this paper is organized as follows. Section 2 focuses on the discovery the hVO_s for the direction of the visual analysis of student intakes in the university in progresses by $hVAP_s$. Section 3 provides an overview of the research related to discover knowledge from mD and focuses on presenting concepts related to visual analysis of mD including visual analytics approach and question types applied on a mdVAS. Section 4 proposes to build a mdVAS, human observation abilities, visVTM, and visual analysis of the university data. Section 5 focuses on visual analysis of student intakes by using the types of questions that visually analyze mD. Section 6 presents the conclusions and development direction in the future.

2 Human Vision Organs

2.1 The Amazing Possibilities of Human Vision

2.1.1 The Main Components of the Human Eye

Look at the main components of the human eye (hE_s) in Fig. 1, we have a clear picture of the major components of the hE_s that the Creator bestows on humans and through a lot of time analyzed by scientists including components such as focusing optics and objects. With a focusing optic consisting of layers like layers of transparent lens, transparent corneas, iris, and pupil. The transparent lenses are able to pick up the light rays from the object to the convergent point focusing on the retina. With a focusing object having a retina layer which is behind focusing optic of the vision organs. In other words, light rays coming from objects through the cornea are focused on the retinal layer to create a visual image of the object. Thus, we see that each component of the hEs has distinct functions and tasks. To concretize the analysis of the main components of the hEs and the hVOs, we look at Fig. 1 and specifically in Fig. 2 to understand how the sublime functioning of the hVOs.

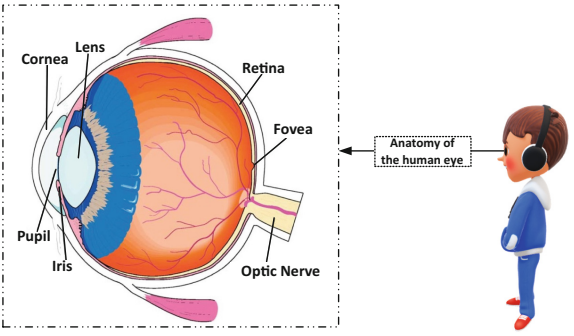

Fig. 1. Topology relational diagram of nine - time units

2.1.2 Human Vision Organs

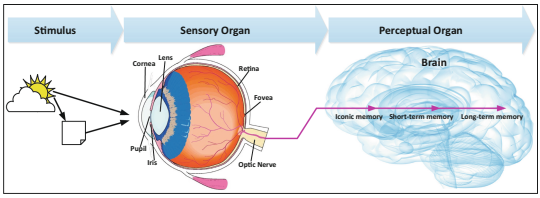

Fig. 2. The human visual perception organs [4, 5]

Look at the Fig. 2, we perform a survey of a light ray from a point on a certain object to the hE_s that light ray may be emitted by itself or reflected into that object from another

light source. According to electromagnetic wave, light rays are electromagnetic waves that carry data of object points. In theory, electromagnetic waves are called modulated data, while light rays are called carrier wares. When reaching the hE_s, light rays converge on the fovea layer in the retina layer to be demodulated by rod-shaped photoreceptors and cone-shaped photoreceptors to separate data from the carrier ware. One of both two types of photoreceptors has its own demodulation function. Rod-shaped cells are very sensitive to light, allowing them to see things in low light conditions and to help them see further. Cone-shaped cells need more light, but they tell us very small details of things when we look at them. The data is driven to the iconic memory (IM) of brain by optic nerve. At IM, the data is perceived to quickly select which data should be transferred to short-term memory (STM). Processing data of IM is understood as categorizing which data is of interest and which is not. Data of interest is transferred to STM. At STM, the data is identified cognition about object. This cognized data is exchanged with information and knowledge available in long-term memory (LTM), collated, filtered carefully and combined with the data stored in LTM to synthesize meaning, thus, the profound understanding of human data is enhanced by the combination of new data and previously available data stored in LTM. At this point, human has awaked to the profound significance hidden within the data, meaning awareness of things. After that, the data is uploaded to LTM in order to update awareness and long-term storage for later reuse or also known as the reminiscent of memories. This is a $hVAP_s$, also known as the progress of transferring data from objects to the human brain, which the hVO_s were progressed naturally by themselves.

2.1.3 Discovering the Efficiency of the hEs

With the ability to perceive the very rich light of the hE_s, the hE_s can distinguish light having different intensities from light to dark. If the light has too high intensity, the iris layer will adjust automatically the size of the pupil layer so that there is enough light reaching the retina. If the convergent point of light rays emitted by objects through the transparent cornea is not on the retina layer, at that time, the eye muscles will automatically adjust the lens that is attached behind the cornea to change the curvature and thickness of the cornea, and cause the convergent point to be shifted back into position on the retinal layer during the progress of receiving objects from the outside into the hVO_s [4–6]. Scientists still do not understand a lot of the mystery of human vision. One of the mysteries is the small number and size of PN_s, so how PN_s of the retina can be able to perform above magical and extremely complex abilities efficiently, accurately and quickly even before these signals are move to the central brain to form vision and control the organs of the hE_s or other parts of the human body. How the hE_s can adapt itself to the intensity and color of the ambient light in many different situations to always accurately and faithfully record the nature of the captured image is a mystery and are major challenges for scientists.

The hE_s can perceive images in a very detailed way thanks to the huge number of photoreceptor neurons (PN_s) that surround the fovea layer. The hE_s can also distinguish very precisely the depth, width and height of a scene thanks to the depth of the fovea layer on the retina. In order to analyze the received signals, each PN of the hE_s acts as a separate element, it performs an accurate sense of size, shape, position, depth, width,

height, amplification and color of the landscape outside the real-world and many other complex mysteries that scientists have not yet discovered. The PN_s network system of the hE_s works so sophisticated and complicated in the progress of image perception that scientists still don't understand its details today.

2.2 Human Vision Awareness Evolution

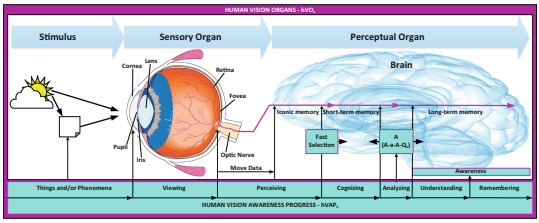

Fig. 3. Integrating human vision awareness progress in the hVO_s

Human perceives things with the five senses. Each sense is compared to a data acquisition system that supports very powerful and very sophisticated for human brain to perceive the real-world. Each sense has different amazing abilities. *Hearing* perceives sound in the real-world with ears. *Touch* perceives objects that exist in the real-world with touching the skin. *Smell* perceives objects and phenomena in the real-world with nose. *Taste* perceives the sweet and sour taste of dishes with tongue. *Vision* perceives things in the real-world with eyes. Among these senses, vision is considered to be the most important one and is carefully analyzed and studied in the section on the great abilities of human vision. Since then, we have found that in this hVOs having hidden superpowers such as the $hVAP_s$, the $hVTM_s$, and the $hIGCE_s$ to awake to things in the real-world, this progress includes stages as $vpcaurH_s$ (Fig. 3).

The analytical stage of the $hVAP_s$ is a core stage for upgrading human understanding, this stage is called the stage of data VA (A-a-A-Q_s) (Fig. 3). At this stage, we perform the combination and integration into the mdVAS with four components such as the component of questions and answers bank system, $hVTM_s$, and the component of $hIGCE_s$ in order to support human extracts information and discovers knowledge from data. With the component of question bank system is embedded in the mdVAS, it will help human answer the questions how to understand things in the real-world? To answer this question, humans must either bring out questions themselves or use the exist questions in system to understand things in the real-world that if you want to understand, you have to ask. With the component of answer system is embedded in the mdVAS, it will help human clearly understand about things in the progress of data VA based on visual graphs.

2.3 Human Visual Viewing-Thinking Mechanisms

To transform objects in the real-world into forms of information and knowledge representation that conform to the $hVTM_s$ as well as their goals are major challenges

for scientists. Humans have used many methods, specifically mathematical modeling method and data visualization method [1–3]. Through these methods, people choose the mode of representation and its presentation tools in accordance with the $hVTM_s$. With the existing system of great senses, people can sometimes use their five senses simultaneously to perceive things in the real-world. To aspect of viewing-thinking, vision is seen as the channel receiving the most information and knowledge compared with the rest of the human senses and previous studies have also shown that formats such as images, graphs, diagrams will be the easily forms conveying and representing the most information and knowledge [7]. Therefore, we choose the format graphs to represent data in this study in order to help human perceive the most information and knowledge from mD.

Through analysis of the hVO_s and the $hVAP_s$ that we performed above, hE_s perceives objects with the rays emitted by the object that reach the eyes, this light is led to the brain by the nervous system to processed into information and knowledge. Based on studies, we found that the brain analyzes information that leads to the eyes at two general and detailed levels. With the level of general, by parallel method the brain quickly analyzes the outstanding characteristics of objects. With the level of detail, the brain analyzes by serial method to detect specific details of an object [4, 8, 9]. The $hVTM_s$ of objects of the hE_s are carried out automatically. We discovered that the hE_s has a mechanism for seeing the whole and the mechanism for seeing the details of objects. With respect to the whole overall mechanism, the hE_s perceives the whole object, that is, it receives light rays coming from the object to provide the brain with processing, the brain draws information at the most general level as the main components of the object present on the representational forms. With respect to the detailed vision mechanism, the hE_s perceives details from the whole perception, the brain analyzes and processes it in more detail in the overall perception, based on the serial method, the hE_s identifies variables, characteristics of variables and detailed components on the representational forms. Through the overall perception and the detailed perception of representation forms of the objects, by viewing-thinking, each person, depending on their knowledge and experience, will perceive the relationship between the variables, between the components of the physical representation forms at different angles of vision with the expectation of discovering new relationships and knowledge available on the representational forms.

Visual analytics system discovers data applying methods of data analysis [2, 3, 10, 11] to render forms of data visual representation from that human can apply viewing - thinking mechanisms, in which contains a lot of information and knowledge. With ability of overall viewing-thinking as soon as detailed viewing-thinking, people will apply the ability to perceive the whole, perceive details and perceive the relationships among the available components on data representations to discover information and the knowledge hidden within it.

2.4 The Infographics Capture Characteristics of the hEs

The hE_s is a vision organs like an optical machine [5, 12, 13], having a very complex structure, having a task for capturing visual information of an object and then converting it to the brain, the brain is in charge of perceiving, cognizing, analyzing, understanding,

remembering and synthesizes images acquired and stored in LTM sent from millions of photoreceptors located on the retina to create a profound visual perception of object images. Environmental factors in which the presence of objects will greatly determine humans' cognitive abilities. Therefore, the selection of data representation formats to suit the human visual perception ability and the $hIGCE_s$ of the hE_s plays an extremely important role in extracting information and discovering knowledge from data. Through the aforementioned analysis and the $hIGCE_s$ of the hE_s are clearly shown through the following five characteristics [4, 6, 12].

- **The hE_s is sensitive to light.** The hE_s having automatic sensitivity to light is sophisticated and impeccable which is shown through aspects of capturing images and image quality are lifelike and almost instantaneous.
- **The hE_s automatically adjusts light.** The hE_s is able to automatically adjust the light of an object from light to dark and vice versa when this light enters the eye because the iris layer automatically performs enlargement or contraction.
- **The hE_s automatically focuses the image.** The hE_s is capable of autofocusing the image of an object very flexibly using a light-sensitive retina surface. The lens layer combines with the muscles in orbit to constantly change the shape of the lens layer to create flexibility in focusing images of objects.
- **The hE_s automatically balances the colors.** The hE_s from the moment it receives light, it automatically balances the colors according to the context in which the light enters the eye. The hE_s always creates harmony and combination closely with the brain to create visual images that help the brain maximize the perception of object images.
- **The hE_s automatically adjusts the focus of the image.** The hE_s automatically adjusts the focal of object images by bending the light through the lens so that it enters focus on the retina, and then received by the vision nerve system and transferred to the brain to analyze and process cognition of object images.

With the $hIGCE_s$ that we analyzed above, the hE_s is really regarded as a powerful and very sophisticated vision organs. Thus, the acquisition of image information of the object is almost perfect, so that this approach is complete and more perfect. Our mission is to make human view-to-think the image of the object on the representations in a more meaningful and profound way, based on the philosophy of Albert von Szent-Gyorgyi points out that "Discovery consists of seeing what everybody has seen and thinking what nobody has thought." [14] is very essential and highly practical science in the field of data visual analytics. To do this, we focus on building a visual analytics system that explores mD by combining and integrating into this system with five components such as the component of banking question system, answer system, the $hVTM_s$, the $hIGCE_s$, and visVTM to assist human in extracting information and discovering knowledge from data. Especially, today when human is more inclined to wake up the mind of each person, extremely important to support the leaders in policy and timely decision making.

3 Related Words

3.1 Discovering Knowledge from Multivariate Data

Extracting information and discovering knowledge from mD is a major challenge in analyzing data visual for researchers on data science and is a problem which is concerned now. In respects of data, data always exists in many different forms as writing, image, sound, light, etc. that its usefulness lies implicit within it which is endless. It challenges people to explore it that means information and knowledge hidden in the data is endless, it challenges people to explore it. According to Andrienko [15], the data represents the results of observations or measurements of phenomena. By means of data analysis, people can study these phenomena. Data analysis is to find answers to various questions related to things and phenomena. For data mining methods, from the past to the present, there are many methods of mining information and knowledge from the data, in there includes algorithmic methods, modeling methods, visual presentation method, visualization method, visual analysis method [1–3, 10]. Each method has different special abilities and will be presented in the next paragraphs and then we will propose to choose one of these methods as the basis for the research on the subject about extracting information and discovering knowledge from data.

Algorithm is a method that can extract information and discover knowledge from data by algorithms. To develop the algorithms, we must rely on the knowledge and experience of computer science specialists to understand and develop it. However, when the data arises quickly, specially, the data is in many different forms and accompanied by the generation of complex data at different times, leading to the method of algorithm development not in time and not timely analysis to respond the need to explore knowledge from data. On the other hand, algorithmic methods cannot mobilize the intelligence of all experts in many areas, also known as interdisciplinary as enrollment, epidemiology, weather, climate change, etc. to analyze and develop algorithms. Modeling is a method that can extract information and discover knowledge from data by modeling. Modeling one or more real-world contexts into a model is a huge challenge for researchers. In there, a data model is a model that provides forms of symbols to describe datasets and sets of behaviors on these datasets. The data model summarizes the data from outside reality and provides an overview of the data to represent a specific range and a limited range of real-world [16]. From the modeling method, researchers have come up with a variety of data models to extract information and discover knowledge from data, such as TLODs model [17], expanded CityGML model [18], ELUDM model for 2.5-3D objects [19], a new LoD and multi-representational concept for the CityGML model [20], etc.

Based on the philosophy of von Szent-Gyorgyi [14], from this philosophy, it paved the way for data visual representation, visualization and visual analytics method born. The visual representation method is data representation in a display form [21, 22], which allows users to view data visually, but limiting is the users haven't understood a lot of the datasets represented here. Data visual analytics and visualization method makes it possible for people to see what everyone saw and understands what people understand. The special point of data visual analytics and visualization method is thanks to the knowledge and experience of analysts in many areas, also known as interdisciplinary

(enrollment, epidemiology, weather, etc.). Data visual analytics is analytical theoretical science major supported by interactive visual interface [22]. There are now more specific definitions of data visual analytics [10] as follows data visual analytics is a combination of automated analytical techniques with visual interaction images to help people understand and reason more easily and make more effective decisions about large and complex datasets. The goal of visual analytics (VA) is to create tools and computer techniques that allows people to synthesize information and have wide and deep insight from large datasets, datasets that change over time, unclear dataset, and conflict datasets, detect what is already there hidden in the data and discover unexpected things from the datasets and provides timely and easy-to-understand assessments from datasets, effectively communicates the evaluation of human activities.

According to the team of Sun.G.D and his colleagues express that VA is the use of visual image of interaction to integrate human's knowledge and reasoning into analytical processes data. This is a dynamic and innovative field of research that is currently applied in a different variety of areas and has many applications in the area, such as security, finance, business, etc. [23]. According to the team of G.Andrienko, VA is aimed at combining human strong points and processing electronic data [24]. Visualization is a method for people and computers to be able to collaborate with each other through a computer graphic interface or being called a means whereby this collaboration can be achieved require. Researchers should find approached methods to solve the complexity of current data and find ways to create data visual analysis tools that are accessible and usable for the potential user community and contribute to resolve some of the big problems.

In today's applications, data volume is created at an unprecedented speed [1]. The ability to collect and store data in applications is increasing rapidly, while the ability of human analyzing this large amount of data to discover knowledge increases at a much lower level. This gap leads to enormous challenges for data scientists, data analysis experts, policy and decision making or in-depth research teams in data analysis that face the process of data analysis, extract information and discover the knowledge hidden in data. Today, with the emergence of a area of VA focused on data processing at different levels such as huge data volumes, heterogeneous data volumes and especially these data volumes change continuously over time. With the advantage of that area of VA is to have the integration of human judgment, existing knowledge as well as human experience from which humans apply data visual representation methods, visual interaction techniques for data visual analysis process is an innovative and scientific approach. The VA area is a combination of many other related research areas including visualization, information analysis, scientific analysis, knowledge discovery, visual interaction, knowledge representation, data management, cognitive and perceived sciences, spatial analysis and statistical analysis [1] that can upgrade VA into a area of full prospecting research and trends in the future.

3.2 Taxonomy of the Models Discovering Knowledge from Multivariate Data

Human's daily activities always take place in many different contexts, this is the basis for human to receive information and knowledge from many different data sources, leading to decision-making on problems that are great challenges for human. Therefore, the study

of methodologies, data visual analytics system, data analysis processes, data analysis models, or information extraction and knowledge discovery tools from data are promising and challenging works for data scientists. Especially today, humans are increasingly aiming to awaken each person's mind to make decisions about financial policies and improve the quality of training based on useful information and discovered knowledge that can be strategic for their units. This research systematizes study works related to areas such as data visualization, analysis of exploratory data, knowledge discovery in the database, and data mining.

Information extraction and knowledge discovery from mD by visual methods and VA of data have been proposed by many groups of authors [2, 3, 13, 25–27], these two areas are currently having a very strong development trend. Humans take advantage of data visualization and their own pre-existing knowledge to be able to perform information extraction and knowledge discovery from data based on visual graphs representing mD. Visual graphs should be designed to best visualize the range of mD visual representations, without omitting any useful information and knowledge, and without misunderstanding non-existent meanings in the data, whereby helping the analyst to observe and empathize the visual graphs using the visual viewing - thinking method to extract information and discover knowledge. During the length of the development history of the areas of data analysis, many groups of authors have researched and proposed models and processes of data analysis by visualization and VA. VA of data is based on the existing knowledge base of each person combined with computer science methods in the data discovery process. With visualization, this approach allows the creation of visual images and graphs that represent information and illustrate data in a way that is easy to understand and must reflect the truthfulness of the datasets to the user. Data visualization uses tools that support data processing and analytics to answer, clarify defined goals, and show current potential through the easiest acquired presentation most absorbed by the model of viewing – thinking and the vision method. Hereafter, we conduct a survey of some of the researches related to discover knowledge from data with the presence of human in the data analysis process.

Discovering knowledge from data is a challenging task for data analysts. Analysts must devise strategies for discovering knowledge from data, where data is transformed into information and knowledge by mathematical model, visualization and VA. With the mathematical model approach to help the analyst discover information and knowledge using mathematical model, this further confirms that the information and knowledge discovered have great useful value related to the support and improvement the quality of education in order to attract a steady and increasing annual student intake entering, and as a basis for maintaining the confidence of students studying at faculties in the university to aim at limiting the remaining student intakes occur. With the visualization approach, this approach applies the available knowledge and experience of humans and cooperates with computers in the process of discovering knowledge from data by vision methods. More specifically, in this research, we will combine mathematical model, visualization, and VA in extracting useful information and discovering knowledge from mD of the student intake types entering the university will give analysts the most insights into what they discover. In the context of increasingly focusing on influencing the minds of each human to help them in decision making, development strategies and other policies in the

unit. Therefore, data analysis systems require the presence of humans to cooperate with computers in the process of discovering knowledge from data, this approach not only brings new prospects, but also presents huge challenges for data analysts who aspire to design and build data analysis systems that focus on taking human as the core of the system. In the following, we analyze a number of studies related to data analysis that have been proposed by many groups of authors and make some comments as a premise for new proposals.

Approaching to extract information and discover knowledge from data by models and processes, the visual method and mathematical model are the most suitable choice and considered the key to success in the knowledge discovery process from data. This approach is the basis for the authors to propose the roles of humans and computers to cooperate with each other in data analysis models and processes including the reference visualization model proposed by Stuart Card et al. [28] to perform from the data transformation stage to the interactive visual form stage by means of visual mappings controlled by humans; The visual analysis process proposed by Keim et al. [10] was built from the visualization model proposed by Wijk [29], this process represents stage one of receiving data sources, stage two applies visualization to transform data into interactive visualization, stage three helps the analyst to apply the viewing - thinking method to perceive the deepest meaning of newly discovered knowledge, and the analyst can rely on the newly discovered knowledge and existing assumptions to continue the analysis further with the technical characteristics applied to the visualization of stage two; The VA process proposed by Keim et al. [30] helps human combine modeling and visualization methods to discover knowledge from data where modeling methods are carried out automatically data analysis while the visualization methods are conducted visual data exploration by human vision; Quality-metrics-driven automation pipeline is extended from the reference visualization model with a quality-metrics-driven automation proposed by Bertini et al. [31] helps the analyst to interact and control the entire process, and finally the analyst obtains the results which are visual structures that are displayed in different ways; The model of arising knowledge in the field of VA proposed by Sacha et al. [32] proposes the role of human involved in the implementation of the mining process, the testing process, the process of arising knowledge and the rest is a visual analysis system automatically performed by a computer; The data analysis system proposed by ThiNguyen et al. [2] has the ability to convert data into information and knowledge by modeling method to build mathematical models combined with visualization that help humans and computers work together with the available knowledge and experience of each human involved in discovery knowledge from data; The visual analytics system that mines mD based on visual graphs proposed by VanPham et al. [3] has two main layers, layer one combines mathematical model and visualization method to render visual graphs, layer two helps the analysts to interact on the visual graphs by asking questions and self-answering with a step-by-step visual viewing - thinking process, step one with the appearance of analysts and visual graphs, step two analysts view the visual graphs, step three analysts understand visual graphs to crystallize information and knowledge, step four analysts again interact with the visual graphs by asking and answering data analysis questions while observing the visual graphs, and the process continues to repeat step one if the human wants to do it for the next times.

Through the systematization and analysis of the above research works helped us get an overview of the area of data analysis by taking models and processes as the focus of the process of extracting information and discovering knowledge from data. In the research works that we have analyzed above, there is a research work by ThiNguyen [2] and VanPham [3] that mentions the types of elementary questions, variation questions, and relation questions used for visual analysis of data based on visual graphs. However, the data analysis system by ThiNguyen's group [2] mentions the types of elementary, variation, and relation questions but does not specify where to use it on the data analysis system, while the research work of VanPham's group [3] uses type of elementary, variation and relation questions to coordinate with criteria in visual analysis discovering knowledge from mD. Based on the survey results of the above research works, we propose a mdVAS applied on the mD of student intakes in the university. To build a mdVAS, we study the hVO_s, $hVAP_s$, $hVTM_s$, $hIGCE_s$ of the hE_s, and build visVTM to integrate into the mdVAS to help the analyst to $vpcaurH_s$. The mdVAS is an approach that thanks to the analyst's available experience and knowledge is used to contribute to the process of uncovering the deep meanings hidden in the data and allows analysts to easily understand data visual forms by using their memory, thinking and imagination ability.

3.3 Related Concepts

3.3.1 Data Analysis

Analyzing data for the purpose of extracting information, exploring knowledge and finding new laws from diverse data sources is an extremely difficult and challenging job for data analysts. The analysis of data is carried out by analysts in a multi-step process [3, 15], each step demonstrates the efficiency and flexibility in data visual analysis. Flexible and effective data analysis not only helps strategy analysts and strategy administrators achieve their desired goals, but also demonstrates the reliability and accuracy of information extraction, explore knowledge, and discover new rules from multivariate datasets for policy making and decision making. Each of these real-world objects has a certain use meaning, it is used for different purposes that people want to explore it. The analyst perceives it through the vision organs, visual information-gathering properties of the eye, visual viewing - thinking mechanisms, and vision awareness progress to view and think visually the existing image of objects in the real-world. From there, analysts will ask visual analysis questions of mD based on the objects that they see because of the need to discover objects and find out new rules of objects.

In visual analytics, we find that visual analysis questions of mD are the only clue, and also the most practical scientific approach to visual analysis of mD. In which, the characteristics of the visual analysis questions of mD include two main parts, that is the assumption and the conclusion [3]. In data visual analysis, question building is the first step and basis for the data visual analysis process to proceed. It can be said that it is impossible to perform data analysis if no questions are asked. Based on the relevance of the question to the data variables, the values of the data variables, and especially the related variables. Therefore, analysts play a central role in the process of visualizing data discovering knowledge and exploiting new rules from mD by applying a data visual analysis process that includes the following steps [3, 15].

- **Step 1:** Analyst proceeds to collect data through a number of specific legitimate sources. After the data has been collected, the analyst performs the removal of the unrelated types of data that are necessary for research purposes, and retain only those types of data for analysis by their purpose.
- **Step 2:** Analyst must focus on building questionnaires and then proceed to use these questionnaires to analyze data to explore knowledge and find new rules hidden within mD in step 1.
- **Step 3:** Analysts have to select reasonable methods they can choose mathematical modeling methods to extract information and explore knowledge from data. The analysts can choose the visual analysis method, especially this method of which is the coordination between computers and humans, apply human knowledge and experience to extract information, discover knowledge, find new rules hidden within mD.
- **Step 4:** Analysts apply the method they have chosen to apply it to the datasets collected in step 1 into the data analysis process.
- **Step 5:** Analysts look at visual graphs to extract information, discover knowledge, find out new laws and then conduct an evaluation of the results achieved to make specific decisions and employment policy for its unit.

3.3.2 Data Visual Analytics

VA is considered as a science of analytical theory, proven by scientists as one of the scientific methods of analyzing data visual analysis. This method uses an intuitive interface to assist the analyst in interacting with the intuitive interface in representing information and knowledge from variable datasets [22]. Therefore, we believe that the tools support efficiently in analyzing visual data are interactive visual interfaces. Interactive visual interfaces are means to represent visual objects that are surveyed, analyzed, designed, evaluated, and implemented by scientists in visual analysis systems that help analysts in extracting information, exploring knowledge, and finding out new rules hidden in multivariate datasets [2, 3]. Today, visual analysis is viewed by data scientists as a promising, prospecting, and challenging field in the future, concretely for strategic analysts, senior leaders of corporations, companies, and universities have to face and use it in data visual analysis [1, 10, 17, 22, 33].

3.3.3 The Types of Analysis Questions of Multivariate Data

To build types of analysis questions of data for mdVAS, we systematize some types of analysis questions suggested by many groups of authors for the category destination extracting information, discovering knowledge, and uncovering new laws from data. This is also the basis for us to propose more general types of analysis questions and are widely applied in the mdVAS of a university that we propose at the next sections of this research. In 2006, Andrienko author and colleagues [15] carried out dividing the questions of data analysis into two levels, level 1 is an elementary level question which is a question related to value of a certain variable, level 2 is a synoptic level question which is a type of question related to a group of values of a certain variable or a question type related to all values of a certain variable. In 2011, Bertin author [8] carried out dividing the questions of data analysis into levels. With level 1 is an elementary level question

which is a question related to value of a certain variable, with level 2 is a synoptic level question of Andrienko author [15]. After that, Bertin divided the synoptic level question of Andrienko into two levels. With level 1 is an intermediate level question which is a question related to value of a certain variable. With level 2 is a global level question which is a question type related to all values of a certain variable. In 2019, ThiNguyen [2] and VanPham [3] author groups also proposed types of analysis questions divided into three levels. With level 1 is an elementary question which is a type of question related to value of a certain variable. With level 2 is a global question which is a type of question related to a group of values of a certain variable or a question type related to all values of a certain variable to understand the characteristics of variables or to understand the law of variation. With level 3 is a relation question which is a type of question related to multiple data variables, used to find correlations between variables to help us discover new relational rules among variables as well as between datasets of related variables.

4 Visual Analytics System

4.1 Visual Analytics System of Multivariate Data

Humans (analysts, data scientists, or senior leaders) inherently have his/her own wonderful abilities that the Creator has bestowed upon their including his/her vision organs, $hIGCE_s$ of their eyes, $hVAP_s$, $hVTM_s$, and their existing experiences and knowledge. Based on these excellent capabilities, the analyst can apply it for both ask and answer his/her own questions while also having good support of the questionnaire and answer set components. These two components have been integrated in the mdVAS (see Fig. 4) in the knowledge discovery progress and draw new rules from the mD represented on the visual graphs. Through the studies above, we discovered that the philosophy of Szent-Gyorgyi [14] the more clarifying the profound meaning of seeing, everyone has the ability to seeing, but thinking the real-world objects they are seeing is a big challenge for the data analysts explore knowledge and find new relationships and rules hidden within objects. Based on this philosophy, we propose building a mdVAS that extracts information, discovers knowledge and draws new rules from mD as an urgent research direction and high practical scientific.

Looking at mdVAS we see that, at the computer layer, there is an external environment that includes many important data sources that need to be collected for the purpose of extracting information and discovering knowledge of data analysts. The visualization method in coordination with mathematical model method [1–3, 10, 13, 23, 25–27, 34] to render the mD into visual graphs to suit the possibility human visual perception. In the VA layer, the analyst plays a central role in the data VA process. In this process includes components such as the visual graph component, the data analyst component by the viewing - thinking method, the component of the analyst who observes the visual graphs by three seeing (see Fig. 5) such as whole seeing, partial seeing and partial detailed seeing, the $hVAP_s$ component, the questions component, and the answers component that the analyst can be used for knowledge discovery purposes and find new rules from visual graphs. At this layer of VA of mdVAS, the $hVAP_s$ component is shown through an automatic progress consisting of the following stages as $vpcaurH_s$. In this evolution, especially the analyzing phase is very important and a big challenge for analysts, in data

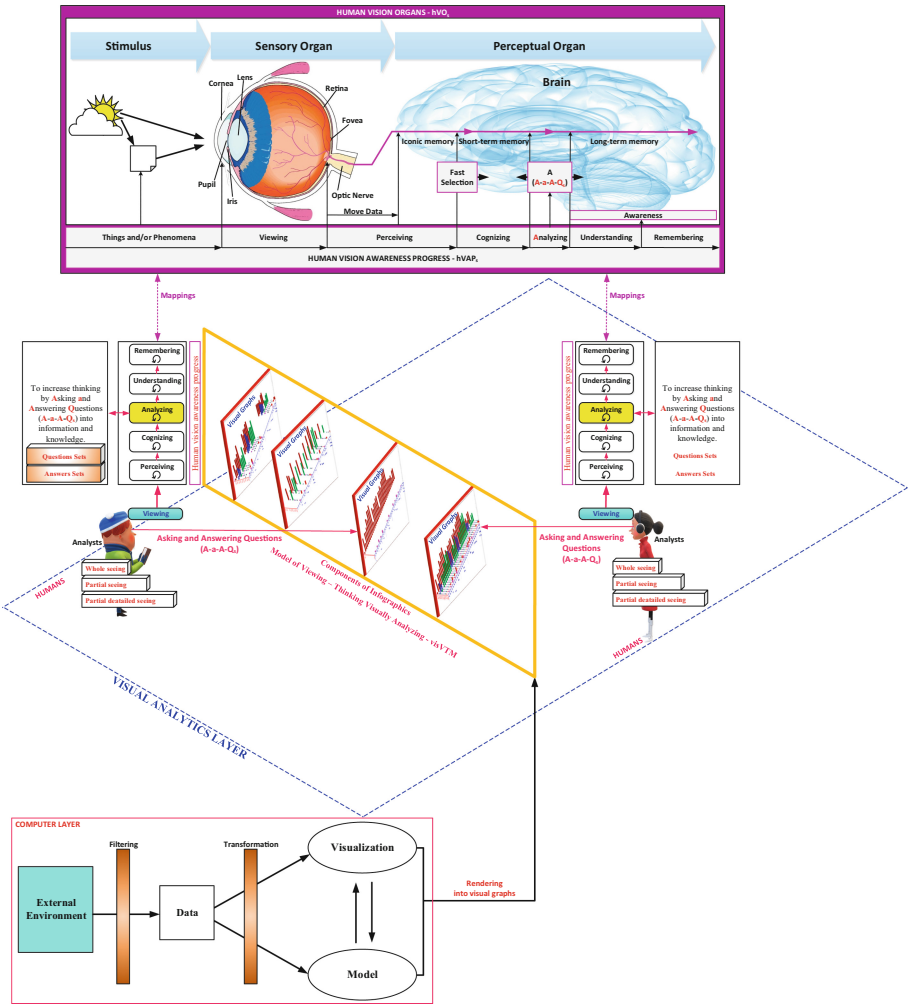

Fig. 4. The visual analytics system of multivariate data (mdVAS) extracts information, discovers knowledge and finds new rules from mD based on visual graphs, where the mdVAS indicates computer layer and VA layer (humans, visual graphs, detailed questionnaire component and the corresponding answers component, visVTM, and $hVAP_S$).

analysis consists of two important aspects, which are question and answer, if we want to understand, we must ask and answer ourselves. To solve this phase of analyzing, we integrated into the mdVAS as a detailed questionnaire component and the answers set correspondingly. Data analysts will use these two components to discover knowledge themselves and find new rules according to their wishes.

From mdVAS, we see that the biggest challenge in a mdVAS is the VA layer, especially the trend to solve the noticeable current problem is according to the approach that awakens the human mind in making decisions to satisfy a certain purpose. Therefore,

VA is understood as the dialogue between the analyst and the data visual representations that the data depends on the applications and according to the performer that the analyst must interpret and must be adapted to those data visual representations. In VA, analysts have a need to observe data visual representations, from which they think and interpret if they want to understand the profound meaning of what they see and next they will think about the next questions to ask. At this point, the types of analysis questions will arise in many different respects for the analyst, and new elements in the data visual representations will have to be considered further. The representations in this dialogue are visual interactions between the analyst and with the data visual representations, in particular, a visual graph representing mD.

To better understand the mdVAS clearly, especially in the VA layer, the types of analysis questions are conducted through many stages, each stage relies on the form of visual representation to questions, the viewing stage of analysts is to ask and answer questions themselves. The data analysis strategy using the VA method is by using the analyst as a key component of the mdVAS that searches for information, discovers knowledge, and discovers generate new rules from mD (Fig. 4). In the computer layer, the computer supports the analyst by displaying mD as visual graphs, the analyst views at the visual graph with his/her knowledge and experience and coordinate with the questions component and different information to think and understand the content hidden within the datasets by visual graphs. This combination is the questionnaire and corresponding answer set for that questionnaire, which means that the analyst understands the deep meaning hidden in the related datasets by a visual graph that the mdVAS provided, which is the questions and the answers component. The scientific and creative nature of this mdVAS is to take advantage of the hVO_s to analyze data visually, these advantages are expressed through the components such as the components of $hIGCE_s$, $hVAP_s$, and $hVTM_s$.

4.2 Human Observation Abilities

Following this, we propose a classification approach for seeing and types of analysis questions. With the above mdVAS, the analyst often performs the progress of observing the visual graph in the following ways as whole seeing (WS_q), partial seeing (PS_q), and partial detailed seeing (PDS_q) (see Fig. 5). The analyst can use the analytical questions available in the mdVAS to analyze the data, subsequent analysis questions will be generated by themselves during the analysis of the data, and the data analysis questions are classified into the following types of questions as elementary questions, variable questions, relationship questions, hierarchical questions, visually thinking questions, overall questions between variables and between datasets of one variable. We have the following conventions, human collectively known as the analyst will have ability whole seeing with forms $WS_q|q = 1, 2, 3, \ldots, n$; partial seeing with forms $PS_q|q = 1, 2, 3, \ldots, n$; partial detailed seeing with forms $PDS_q|q = 1, 2, 3, \ldots, n$; in which $\{q_1, q_2, q_3, \ldots, q_n\}$ are visual analysis questions of mD. Thus, humans can see real-world objects with $S = \{WS_q, PS_q, PDS_q\}$ in process of data visual analysis.

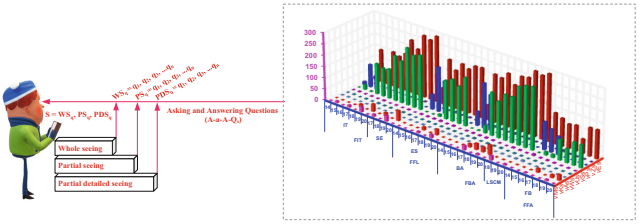

Fig. 5. Humans have the ability to see whole, partial, and partial detailed of real-world objects in progress of analyzing visually multivariate data.

4.3 Visual Analytics of a University Data

Einstein's theory of relativity says that time is relative and indefinite. The elapsed time depends on your frame of reference [35]. In managing any job, time management can be seen as a challenging job. Because the time it takes to complete something is an indefinite time. In structured datasets, time structure has semantics related to time distribution such as discrete time, continuous time, and absolute time [36]. On the time axis, discrete time means that there is a definite number of points between any two points on that axis; while continuous time means that between any two points on the time axis there is an indefinite number of points and it is suitable for representing constantly changing objects such as change of courses in a university; absolute time means to show a perfectly definite object as majors or faculties in a university. Time data types include time point, time segment, and time period. Time point means describing a point on the time axis, giving an example of a meeting that took place at 14:00 on January 12, 21 at the office of faculty of information technology (FIT), Nguyen Tat Thanh University (NTTU). Time segment means describing a time segment on the time axis and is limited by two points of time, for an example of a course year starting from 2020 ending until 2024; for example, the time for submitting online homework of students is from January 9, 21 to January 20, 21. Thus, the data of courses belongs to time segment is represented in the following formats: the discrete time format is 14, 15, 16, etc.; the absolute time format is 14–18, 15–19, 16–20, etc.; 14–15, 15–16, 16–17, etc. A period of time to describe an indeterminate segment of time, giving the example of a student who has been in college for 5 years and has not graduated from university.

Data of a university belongs to many different fields, including the part of data on the intake of students entering and leaving at a university according to each academic year which occurs regularly and continuously, a piece of data of particular interest to school leaders. In this research, which focuses on analyzing this data piece and student intakes are represented by the form of ratio data. The main purpose is to help school leaders know the student intakes by each course that the faculties are managing so that they have a basis to compare and make improvement decisions to improve the quality of training to meet all labor resource needs of society. Student intake entering means that every year after graduating from high school, students will perform a special task in their future that is to apply for admission into universities by many different university admission methods. The intake of students making reservation means that in the learning process, for some reason, students cannot continue to study anymore, at that

time, students can make a record to reserve their academic results. In the future, when possible, students can use this reservation to ask the school to go back to school. The intake of graduated students means students who have received a university diploma. The intake of current students means students who are studying a particular major in a department of a university. The intake of suspended students means students who are in violation of school rules and regulations. The intake of students withdrawing fees means students may encounter some personal problems, they cannot continue to study, according to the school's rules and regulations, at that time, students may be entitled to withdraw tuition. The intake of students dropping out means students making an announcement to the university that the student has stopped studying. This is voluntary of students and students can carry out the prescribed formalities of giving up university. The intake of students studying in excess of training time means that each student has a prescribed study period, which is a general regulation of the ministry of education and training. If during a period of study that a student has not completed the prescribed university program that student is called a student exceeding the training period. The intake of students transferring majors means students apply to transfer to another major. The school allows students to have one time transfer but can only transfer major within the internal university.

In this research, the data source is explored from the training management system (TMS) of NTTU, Hochiminh City, Vietnam where we are working. To secure this data source we have encrypted it. The research focuses on exploring the data sources of 6 majors including Information Technology (IT), Software Engineering (SE), English Language (ES), Business Administration (BA), Logistics and Supply Chain Management (LSCM), Finance and Accounting (FB), in which each major has opened 7 training courses including 14, 15, 16, 17, 18, 19 and 20. The above six majors are managed by 4 respective faculties after the FIT, the Faculty of Foreign Languages (FFL), the Faculty of Business Administration (FBA), the Faculty of Finance and Accounting (FFA). In the 7 courses of each major, there are 9 student intakes, including student intake in admission, student intake with retention of academic results, student intake of graduates, student intake studying, student intake of suspended, student intake of withdrawals, student intake dropping out, student intake studying beyond the prescribed training time, and student intake changing majors. Through exploiting this educational data sources (EDS), we have discovered that this data source needs to be divided into 9 types of student intakes for 7 courses, 6 majors and 4 faculties. From that, we calculate 7 courses $\times$ 6 majors $=$ 42 data tuples and also from here we calculate 7 courses $\times$ 6 majors $\times$ 19 data domain variables $=$ 798 data cells in the EDS.

4.4 Structured Hierarchical Tree by Multivariate Groups

Based on the above data structures, we have discovered 19 important data variables used to represent the student intakes participating in the university courses from here creating relations. The relations between the variables make a hierarchical tree (hTree) by multivariate groups (MG_s) (Fig. 6). The relations between variables on a hTree by MG_s are structured into a visVTM of mD, the correlation between variables on this model helps to upgrade the $hVAP_s$ of the analyst includes stages such as $vpcaurH_s$ in extracting information and discovering new knowledge from mD.

The hTree by MG_s helps the analyst to understand the degree of association between the new knowledge latent variables and especially where the correlation between variables of different student intake is hidden for which the analyst wants to discover knowledge. The faculty variable containing the dataset of faculties in the university, is called vF_s. There are two relational variables linked together by faculties, called vF_p and vF_c. The course variable contains the dataset of courses of each faculty in the university and is collectively referred to by the research as the temporal variable (vT_s). The major variable (vD_s) contains the data domain that is the majors in the university. The student intake variable (vI_s) contains a set of nine sub variables containing the data domains, these sub variables are variables that directly affect the student intakes participating in the majors and courses that help the analyst cover and enhance visual analysis of data. The data subdomain variables include vE_s which is the domain variable containing the dataset of enrolled students, vZ_s is the domain variable containing the dataset about students with retention of academic results, vP_s is the domain variable containing the dataset about students who have received diplomas, vS_s is the domain variable containing the dataset of students studying, vU_s is the domain variable containing the dataset of suspended students, vW_s is the domain variable containing the dataset about students applying for tuition withdrawal, vA_s is the domain variable containing the dataset of students withdrawing from school, vK_s is the domain variable containing the dataset of students studying beyond the training period, and vM_s is domain variable containing the dataset of students applying to change majors. All subdomain variables of this data correspond to each number of students enrolled in the university described by the variable vI_{Num}. The following are specific descriptions and classifications of data domain variables, data tuple variables, data relational variables, and relations related.

vF_s is a data domain variable belonging to the relation vF, but in the relation vD it is named vF_{ds} used to describe university faculties. At this point, vF_{ds} becomes the relational domain variable of data between the two relations vF and vD. In which, vF_s and vF_{ds} contain the datasets {'FIT', 'FBA', 'FFA', 'FFL', etc.} but may not be equal in number of corresponding relations in the two relations vF and vD. vF_p is a data domain variable belonging to the relation vF_{pc}, also this variable in the relation vF_{pc} we duplicate it into a data relation domain variable named vF_c, these two variables work together to create a hierarchical grouping relationship of faculties in the university. vT_s is a data domain variable belonging to the relation vT, but in the relation vI it is named vT_{is} used to describe university courses. At this point, vT_{is} becomes the data relation domain variable between the two relations vT and vI. In which, vT_s and vT_{is} contains the same datasets {14, 15, 16, 17, 18, 19, 20, etc.} but may not be equal in the number of corresponding relations in the two relations vT and vI. vD_s is a data domain variable belonging to the relation vD, but in the vI relation it is named vD_{is} used to describe the fields of study in the university. At this point, vD_{is} becomes the data relation domain variable between the two relations vD and vI. In which, vD_s and vD_{is} contains the same datasets {'IT', 'SE', 'BA', 'ES', 'FB', 'LSCM', etc.} but may not be equal in the number of corresponding relations in the two relations vD and vI. vI_s is a data domain variable belonging to the relation vI, this variable has data subdomain variables vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, and vM_s used to describe the student intakes of courses and majors in the university. vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s are data domain variables belonging to the

relation vI used to describe the student intakes corresponding to each number of students described by the vI_{Num} variable of majors according to university courses. In which, the corresponding student intakes variables have tasks such as containing the dataset of enrolled students (vE_s), containing the dataset about students with retention of academic results (vZ_s), containing the dataset about students who have received diplomas (vP_s), containing the dataset of students studying (vS_s), containing the dataset of suspended students (vU_s) containing the set data about students applying for tuition withdrawal (vW_s), containing the dataset of students withdrawing from school (vA_s), containing the dataset of students studying beyond the training period (vK_s), and containing the dataset of students applying to change majors (vM_s). With t, p, q, b, c are the contextual variables of the data visual analysis that these variables take a role as free variables or bound variables used in data tuples and domains relational calculus expressions.

The hTree by MG_s (Fig. 6) is structured according to the characteristics of hierarchical groups of multivariate. We conduct visual analysis of MG_s on the hTree, the MG_s on the left side of the hTree include variables vF_p, vF_c, vF_s, vFd_s, vD_s representing the roles of related faculties and majors associated with the respective faculties. The MG_s on the right side of the hTree include variables vM_s, vK_s, vA_s, vW_s, vU_s, vZ_s, vP_s, vS_s, vE_s, vI_{Num}, and vI_s representing the correlation with each other to compare the student intakes participating in the courses of each major and representing the dependent constraint between the student intakes studying and the student intakes enrolled by using the quartet variables vE_s, vS_s, vI_{Num}, and vI_s. This MG_s on the right side have features such as the higher the enrollment of students, the higher the characteristics of students studying. The MG_s in the middle of the hTree of variables vD_s, vD_{is}, vT_{is}, vT_s, and vI_s representing a major association will correspond to multiple courses represented by the student intakes participating in the university. From here, this study integrates the three MG_s on the left side, MG_s on the right side and MG_s in the middle of the hTree to create a visVTM of mD (Fig. 7), this approach is the scientific basis for improving cognitive progresses for analysts. The visVTM exploring information and knowledge from EDS is proposed, based on this model and depending on the nature of the MG_s, we perform the structure to visual graph (Fig. 8) in information extraction and knowledge discovery from EDS of the TMS of NTTU. The relationships between MG_s, the hierarchies of MG_s, the in-out of data sources, and the correlations between MG_s are generated on the hTree is the basis for structuring into a visVTM that will help analysts know many ways to approach information extraction and discover hidden knowledge on this model.

After analyzing the MG_s on the hTree, naming the variables and briefly describing the capabilities of the MG_s, we found that the variables vF_s, vF_p, vT_s, vD_s, vI_s, vI_{Num}, vF_{ds}, vF_c, vD_{is}, vT_{is}, vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s are called data domain variables. We choose a new approach by using the hierarchical group method of the above variables to create a hTree for managing of student intakes of the university. The hTree clearly shows the grouping of variables to display data domains to create related data domains. With data domain variables, we have the following variables vF_s, vF_p, vT_s, vD_s, vI_s, vI_{Num}, vF_{ds}, vF_c, vD_{is}, vT_{is}, vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s. With data tuple variables, we have the following variables t, p, q, b, and c. These variables have two types, which are free and bound variables depending on the form of visual analysis of mD, sometimes it is a free variable and sometimes it is a bound variable.

The variables vF_p, vF_c, vF_{ds}, vT_{is}, vD_{is} make up the correlation relations between the relations corresponding following vF_{pc}, vD, vI, and vT. From here, analysts can visually analyze multivariate data based on relational calculus expressions on the data tuples and domains. Each group of multivariate put together has the ability to visually represent multivariate datasets, closely related MG_s, and the combination of variables helps the analyst to answering his/her own questions of mD.

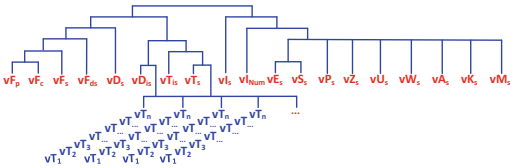

Fig. 6. Hierarchical tree (hTree) by multivariate groups in VA of mD of student intakes participating in each course in the major of each faculty in the university.

4.5 Viewing – Thinking Model Visually Analyzing Multivariate Data

From mdVAS (Fig. 4), data analysts can conduct knowledge discovery according to data tuples and domains relational calculus [37]. To visually analyze mD in extract information and discover new knowledge in the form of tuples, the analyst uses a data tuple relational calculus expression with form $\{t_1.A_i, t_2.A_j, ..., t_n.A_k | P(t_1, t_2, t_3, ..., t_n)\}$, in which $\{t_1, t_2, t_3, ..., t_n\}$ are tuple variables, $\{A_i, A_j, ..., A_k\}$ are the attributes in the t tuple variables correspondingly, and P is the formula formed from the prime formulas. To visually analyze mD in information extraction and discovery of new knowledge in the form of data domains, the analyst uses a relational calculation expression on data domains of the form $\{x_1, x_2, x_3, ..., x_n | P(x_1, x_2, x_3, ..., x_n)\}$, in which $\{x_1, x_2, x_3, ..., x_n\}$ are domain variables that receive a value as a domain of attribute values, P is a formula formed from prime formulas, finally the return result set of values $\{x_1, x_2, x_3, ..., x_n\}$ such that when the values are substituted for $\{x_i\}$ then P is true. The approach to using relational calculus expressions on the data tuples and domains in the data visual analytics of various student intakes of a university is the scientific basis in the VA of mD to help school leaders in implementing financial policies and making timely decisions.

This visVTM of mD of the university was integrated by us into the mdVAS (Fig. 4). With its strengths, in the analysis progress, the analyst applies available knowledge, the hVO_s, the $hIGCE_s$, the $hVAP_s$, and $hVTM_s$ for extracting information and discovering knowledge from mD. The visual analysis questions of mD will be formed based on the visVTM multivariate data presented in two aspects, the first aspect is the visual analysis questions of mD built on this system, the second aspect is the visual analysis questions of mD that will be generated in the progress of vision awareness of the analyst, depending on the topic of data they are analyzing, this progress will upgrade the analyst's knowledge and automatically generate questions and answers corresponding to the data topic they are discovering.

The visVTM multivariate data about student intakes of courses in each faculty of a university used to represent information extraction and knowledge discovery through

formal queries of relational calculus expressed on data tuples and domains. These formal queries representation helps the analyst to extract information and discover new knowledge. Through this visual analysis, leaders see useful and unhelpful pieces of information and knowledge to make the right decisions for each faculty in the university about managing student intakes in different courses. From that, faculties have taken measures to upgrade their approaches to student data management. The visual analysis of data by relational calculus expressions on data tuples and domains is shown in the next sections.

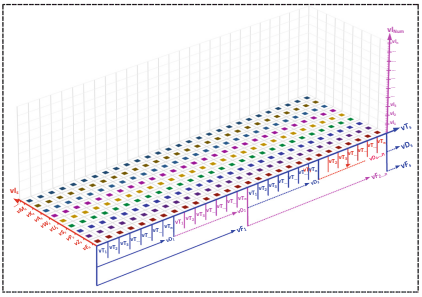

Fig. 7. Model of viewing - thinking visually analyzing (visVTM) multivariate data integrated by MG_S of hierarchical tree by MG_S related to student intake types of courses and majors.

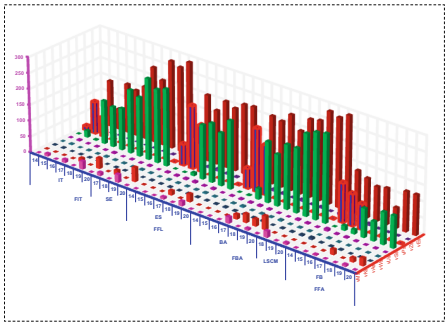

Fig. 8. A visual graphical representation of multivariate data of different the student intakes participating in the university by a viewing-thinking model analyzing in mdVAS.

4.6 Questions with Relational Calculus Expressions on Tuples and Domains

- **Question 1:** Indicate the faculties having majors and the student intakes opened for training in courses 18, 19 and 20. This question is conducted visual analysis including the following parts.

$$\{t.vD_s, b.vE_s, b.vZ_s, b.vP_s, b.vS_s, b.U_s, b.W_s, b.vA_s, b.vK_s, b.vM_s \mid t \in vD \wedge b \in vI\ (b.vD_{is} = t.vD_s) \wedge (\exists p \in vI\ (t.vD_s = p.vD_{is} \wedge p.vT_{is} = 18) \vee \exists q \in vI\ (t.vD_s = q.vD_{is} \wedge q.vT_{is} = 19) \vee \exists c \in vI\ (t.vD_s = c.vD_{is} \wedge c.vT_{is} = 20))\} \quad (1)$$

- **Question 2:** Indicate the majors of the faculties with the student intakes entering or the student intakes changing majors. This question is analyzed visually through the following sections.

$$\{t.vF_s, b.vD_s, c.vE_s, c.vM_s \mid t \in vF \wedge b \in vD \wedge c \in vI\ (t.vF_s = b.vF_{ds} \wedge b.vD_s = c.vD_{is}) \wedge (\exists p \in vI\ (b.vD_s = p.vD_{is} \wedge p.vE_s > 0) \vee \exists q \in vI\ (b.vD_s = q.vD_{is} \wedge q.vM_s > 0))\} \quad (2)$$

- **Question 3:** Indicate the faculties and majors that have both the student intakes suspended and the student intakes exceeding the training period.

$$\{t.vF_s, b.vD_s \mid t \in vF \wedge b \in vD\ (t.vF_s = b.vF_{ds}) \wedge (\exists p \in vI\ (b.vD_s = p.vD_{is} \wedge p.vU_s > 0) \wedge \exists q \in vI\ (b.vD_s = q.vD_{is} \wedge q.vK_s > 0)\} \quad (3)$$

- **Question 4:** For course 20, please indicate the majors that have the student intakes studying (vS_s) and the student intakes exceed the training time but do not have the student intakes suspended.

$$\{t.vD_s, p.vS_s, p.vK_s, q.vU_s \mid t \in vD \wedge (\exists p \in vI\ (t.vD_s = p.vD_{is} \wedge p.vS_s > 0) \wedge \neg\exists q \in vI\ (t.vD_s = q.vD_{is} \wedge q.vU_s \leq 0) \wedge p.vI_{is} = 20)\} \tag{4}$$

- **Question 5:** For each major in each faculty of courses 17, 19 and 20, please indicate the major, the faculty of charge and whole of student intakes involved in each of those majors.

$$\{t.vD_s, t.F_{ds}, p.vE_s, p.vZ_s, p.vP_s, p.vS_s, p.vU_s, p.vW_s, p.vA_s, p.vK_s, p.vM_s \mid t \in vD \wedge p \in vI \wedge \exists q \in vT\ (q.vT_s = p.vT_{is} \wedge p.vD_{is} = t.vD_s \wedge (q.vT_s = 17 \vee q.vT_s = 19 \vee q.vT_s = 20))\} \tag{5}$$

- **Question 6:** Find the majors of the faculties and the student intakes that are open for enrollment in all courses.

$$\{t.vD_s, t.vF_{ds}, b.vE_s, b.vZ_s, b.vP_s, b.vS_s, b.vU_s, b.vW_s, b.vA_s, b.vK_s, b.vM_s \mid t \in vD \wedge b \in vI\ (b.vD_s = b.vD_{is}) \wedge \forall p\ vT\ (\exists q \in vI\ (q.vT_{is} = p.vT_s \wedge q.vD_{is} = t.vD_s))\} \tag{6}$$

- **Question 7:** Indicate the majors of the faculties and the student intakes that are open for enrollment in all courses offered by the FIT, by the FFL, or by the FFA management.

$$\{t.vD_s, t.vF_{ds}, b.vE_s, b.vZ_s, b.vP_s, b.vS_s, b.vU_s, b.vW_s, b.vA_s, b.vK_s, b.vM_s \mid t \in vD \wedge b \in vI\ (b.vD_s = b.vD_{is}) \wedge \forall p\ vT(t.vF_{ds} = \text{`FIT'} \vee t.vF_{ds} = \text{`FFL'} \vee t.vF_{ds} = \text{`FFA'} \Rightarrow (\exists q \in vI\ (q.vT_{is} = p.vT_s \wedge q.vD_{is} = t.vD_s)))\} \tag{7}$$

- **Question 8:** Indicate faculties, majors and student intakes in which have the student intakes entering greater than 200 students participating in the university.

$$\{vF_{ds}, vD_s, vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s \mid \exists vD_s\ (<vF_{ds}, vD_s> \in vD \wedge \exists vD_{is}, vE_s\ (<vD_{is}, vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s> \in vI \wedge vE_s > 200 \wedge vD_s = vD_{is}))\} \tag{8}$$

- **Question 9:** Indicate the majors of the faculties having the student intake studying, the student intake being suspended, the student intake applying for tuition withdrawal, the student intakes withdrawing from school, and the student intakes changing majors all of these types of intakes are greater than or equal to 1 participating in the majors of the courses 15, 16, 17, 18, 19, or 20.

$$\{vF_{ds}, vD_s, vS_s, vU_s, vW_s, vA_s, vM_s \mid \exists vD_s\ (<vF_{ds}, vD_s, vS_s, vU_s, vW_s, vA_s, vM_s> \in vD \wedge \in vI \wedge vD_s = vD_{is} \wedge \exists vS_s, vU_s, vW_s, vA_s, vM_s, vD_{is}, vT_{is}\ (<vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s, vD_{is}, vT_{is}> \in vI \wedge vD_s = vD_{is} \wedge vS_s > 0 \wedge vU_s > 0 \wedge vW_s > 0 \wedge vA_s > 0 \wedge vM_s > 0 \wedge (vT_{is} = 15 \vee vT_{is} = 16 \vee vT_{is} = 17 \vee vT_{is} = 18 \vee vT_{is} = 19 \vee vT_{is} = 20))\} \tag{9}$$

- **Question 10:** Indicate the faculties, the majors and the courses having the student intake changing majors, the student intake withdrawing from school, the student intake being suspended, and the student intake reserving academic results.

$$\{vF_{ds}, vD_s, vT_{is}, vM_s \mid \exists vF_{ds}, vD_s\ (<vD_s, vF_{ds}> \in vD \wedge \exists vD_{is}, vM_s\ vA_s, vU_s, vZ_s\ (<vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s, vT_{is}, vD_{is} > \in vI \wedge vD_s = vD_{is} \wedge vM_s > 0 \vee (vA_s > 0 \vee vU_s > 0 \vee vZ_s > 0))\} \quad (10)$$

- **Question 11**: Please indicate the majors of each faculty not having the student intake exceeds the training time and indicate extra the student intake entering, the student intake studying, the student intake applying for tuition withdrawal and the student intake changing majors in courses belonging to faculties in the university.

$$\{vD_s, vF_{ds}, vK_s \mid \exists vD_s, vF_{ds}\ (<D_s, vF_{ds}> \in vD \wedge \neg\exists vD_{is}, vK_{is}\ (<vE_s, vZ_s, vP_s, vS_s, vA_s, vK_s, vM_s, vD_{is}> \in vI \wedge vD_s = vD_{is} \wedge vK_s > 0))\} \quad (11)$$

- **Question 12:** For each major that is open for training according to each course having the student intake more than 180 students, please indicate the majors, courses, the student intake studying and the student intake entering and indicate extra the types of student intakes remaining belonging to faculties in the university.

$$\{vD_{is}, vT_{is}, vS_s, vE_s, vF_{ds}, vZ_s, vP_s, vU_s, vW_s, vA_s, vK_s, vM_s \mid \exists vS_s\ (<vD_{is}, vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s, vT_{is}> \in vI \wedge \exists vD_s, vF_{ds}\ (<D_s, vF_{ds}> \in vD \wedge vD_s = vD_{is} \wedge vS_s > 180))\} \quad (12)$$

- **Question 13:** Indicate the faculties which has the majors having the student intake entering between 1 and 300 students and having the student intake graduated between 1 and 200 students.

$$\{vF_{ds}, vD_s, vE_s, vP_s \mid \exists vD_s\ (<vD_s, vF_{ds}> \in vD \wedge \exists vD_{is}, vE_s\ (<vD_{is}, vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s, vT_{is}> \in vI \wedge vD_s = vD_{is} \wedge vE_s \geq 1 \wedge vE_s \leq 300) \wedge \exists vD_{is}, vP_s\ (<vD_{is}, vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s, vT_{is}> \in vI \wedge vD_s = vD_{is} \wedge vP_s \geq 1 \wedge vP_s \leq 200))\} \quad (13)$$

- **Question 14:** With each major belonging to FIT and FBA faculties managing, please indicate the student intake entering, the student intake receiving diplomas, the courses and enclosed the majors.

$$\{vE_s, vP_s, vT_{is}, vD_{is}, vF_{ds} \mid \exists vD_{is}, vT_{is}\ (<vD_{is}, vE_s, vZ_s, vP_s, vS_s, vW_s, vA_s, vT_{is}> \in vI \wedge \exists vD_s, vF_{ds}\ (<vD_s, vF_{ds}> \in vD \wedge vD_{is} = vD_s \wedge (vF_{ds} = \text{`FIT'} \vee vF_{ds} = \text{`FBA'})))\} \quad (14)$$

- **Question 15:** For each course 14, 16, 18 and 20 having the student intake studying greater than or equal to 100 students of each faculty, please indicate the faculties, the majors, the student intake entering and the student intake studying in the majors of faculties belonging to each course.

$$\{vF_{ds}, vD_s, vT_{is}, vE_s, vS_s \mid \exists vD_s, vF_{ds}, vE_s\ (<vF_{ds}, vD_s, vS_s, vE_s> \in vD \wedge vI \wedge vD_s = vD_{is} \wedge \exists vS_s, vD_{is}, vT_{is}\ (<vD_{is}, vE_s, vZ_s, vP_s, vS_s, vU_s, vW_s, vA_s, vK_s, vM_s, vT_{is}> \in vI \wedge vD_s = vD_{is} \wedge vS_s > 100 \wedge vT_{is} = 14 \vee vT_{is} = 16 \vee vT_{is} = 18 \vee vT_{is} = 20))\} \quad (15)$$

5 Visually Analyzing of University Multivariate Data

5.1 Visually Analyzing Multivariate Data with Questions and Answers

The $hVAP_s$ has many different stages in which the analysis stage is one of the most important stages of the progress of perceiving objects that exist in the real-world of humans. Humans perceive the real-world objects with the vision organs with the desire to discover the objects, but also to understand the objects, humans have to analyze it. Therefore, the analytical stage is the stage in which humans ask and answer questions themselves in the process of discovering new knowledge, the philosophy of Szent-Gyorgyi [14] pointed out that humans want to understand objects, they must ask questions, because questions are the first step and the basis for motivating humans to carry out the process of analyzing and discovering objects.

In the analytical stage of the humans' visual perception progress, if humans want to understand objects, they must ask and answer questions to extract information and discover new knowledge. Therefore, developing a questionnaire is an important initial step and is the basis for accelerating the data analysis process to be conducted. Research shows that data scientists cannot perform data analysis without questions which were asked with the desire to discover objects. Types of analysis questions have been proposed by many authors in the past to discover knowledge from data based on the relationships between variables and the values of data variables. From here, we propose to create a visVTM as a foundation for integrating visual graphs representing mD. The group of authors Andrienko has divided the data analysis questions into two groups as one is elementary questions and the other is synoptic questions [15]. The author Bertin has divided data analysis questions into three levels as elementary level, intermediate level, and overall level [8]. The group of authors ThiNguyen has divided the questions into 3 levels, level 1 is elementary questions, level 2 is variation questions, and level 3 is relationship questions [2].

Through surveying related research works, we draw the most important point in data analysis that if humans want to extract information and discover knowledge from data, they must perform data visual analysis by using question types and accompanying answers. We survey the data variables to create a hTree by MG_s from which to combine the variables and the data of these variables to answer questions for the purpose of extracting information and discovering new knowledge from data. Based on the types of questions proposed by the group of authors above, we find the elementary questions, the variable questions, and the relational questions are suitable for this study, so we choose these three types of questions to apply to the mdVAS. In addition, we propose the types of elementary questions (Q_{es}), variation questions (Q_{vs}), relationship questions (Q_{rs}), hierarchical questions (Q_{hs}), visual thinking questions (Q_{vts}), and overall questions (Q_{os}) to applying for mdVAS. To become a component capable of classifying standard visual analysis questions of mD, we provide specific definitions of question types used in mdVAS (Fig. 4) as follows.

- **Definition 1:** Q_{es} include questions regarding a value of a variable or regarding a value of many variables combined together to help the analyst understand the value of a variable or mD.

- **Definition 2:** Q_{vs} include questions regarding a value group of a variable or variables that combine together to help the analyst understand the variational rules of variables.
- **Definition 3:** Q_{rs} include questions regarding many data variables that help the analyst find correlations and relationships between variables to discover correlation and new relational rules between data variables.
- **Definition 4:** Q_{hs} include questions regarding many data variables that help the analyst find hierarchies between variables.
- **Definition 5:** Q_{vts} include questions regarding abstraction between data variables that help the analyst visualize new knowledge in data abstraction.
- **Definition 6:** Q_{os} include questions regarding overall datasets of relational variables that help the analyst see the overall work of an object in the work process.

5.2 Visually Analyzing Multivariate Data of the Student Intakes

For visually analyzing of mD of student intakes in a university, we have conducted research to make a visVTM to integrate into the mdVAS. From visVTM, we apply relationship expressions on the data tuples and domains to help the analyst extract information and explore new knowledge. The visual graphs visually representing mD in Fig. 8 about student intakes participating in the university is rendered from visVTM in mdVAS to help the analyst easily explore new knowledge. Based on this basis, we have applied relational calculation expressions on the data domains and tuples to assist the analyst in extracting information and discovering knowledge about student intakes participating in the university shown in visually analyzing data.

For mdVAS, humans play the most important role in the process of analyzing extracting information and discovering knowledge from data. To evaluate the student intakes participating in the university to find out which types of intakes directly affect the financial policy and making decision of school leaders, we survey data variables, subgroup hierarchy by MG_s to create hTree to make visVTM, from that visually representing the student intakes on the visual graphs, through which we pose some analytical questions to verify the effectiveness of the visual graphs representing mD. The analyst looks at visual graphs and with the powerful support of components in mdVAS can ask questions themselves and also can answer these questions themselves when interacting with visual graphs to extract information and discover new knowledge. Let's observe together to investigate the visual graphs that visually represent the student intakes of a university in Fig. 8 above it which is the result rendering from visVTM in the mdVAS.

For visual analysis questions of mD are questions asked by the analysts with the desire to extract information and discover new knowledge while observing visual graphs. Depending on the data characteristics the analysts can ask appropriate analytical questions when observing the visual graphs. The visual analysis questions of mD are divided into six categories of questions that we have defined standard above, and for each visual analysis question of mD including an assumptive and conclusive part. The assumptive part is the data that the analyst will have to provide when the question begins to be asked, the conclusive part is the results obtained after the question is analyzed visually. In the following, we focus on visual analysis of different student intakes by coordinating with all components on the mdVAS (see Fig. 4) to extract information and discover new knowledge.

5.3 Strategy for Answering Visual Analysis Questions

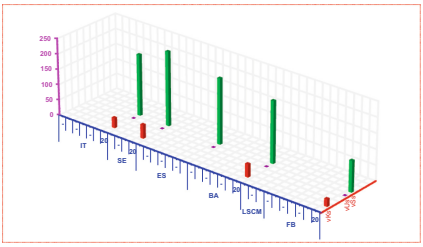

Fig. 9. Visual representation of the student intakes studying, the student intakes suspended, and the student intakes exceed the training time in course 20 of majors in the university by visual graph representing of mD.

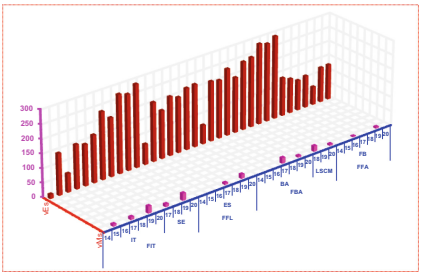

Fig. 10. Visual representation of the student intakes entering and student intakes changing majors of courses belonging to the majors of faculties in the university by visual graph representing of mD.

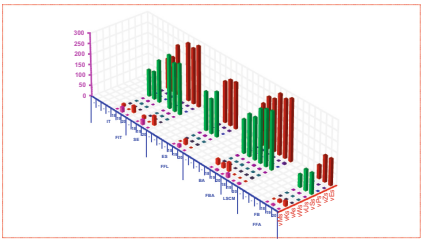

Fig. 11. Visual representation of the student intakes in courses 18, 19 and 20 of IT, SE, ES, BA, LSCM and FB majors belonging to FIT, FFL, FBA and FFA faculties respectively managed by visual graph representing of mD.

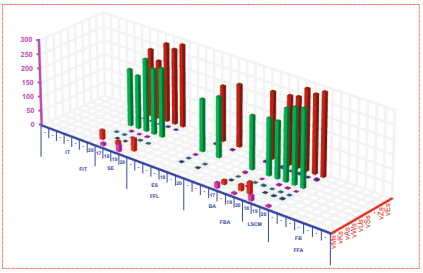

Fig. 12. Visual representation of each major is open to training belonging to each course having the student intake studying more than 180 students and the student intake entering enclosed the student intakes remaining in courses of majors of faculties in the university.

5.3.1 Elementary Questions

Approaching to extract information and discover knowledge from mD of student intakes using strategy of combining Q_{es} with relational calculus expressions on data tuples on the visual analytics system to apply for question $4^{(4)}$ having given assumption is course 20, the student intakes studying, the student intakes exceed the training time and the student intakes suspended. The conclusion of the analysis process is the majors, the student intakes studying, the student intakes exceed the training time and the student intakes suspended. The analyst views and thinks the visual graph in Fig. 8 and then conducts visual analysis of mD on this visual graph. The results of visual analysis of mD in Fig. 9 help the analyst to answer this question, they discover that course 20 has the student intakes studying and the student intakes exceed the training time, but there are no the student intakes suspended. Course 20 of any majors also has the student intakes

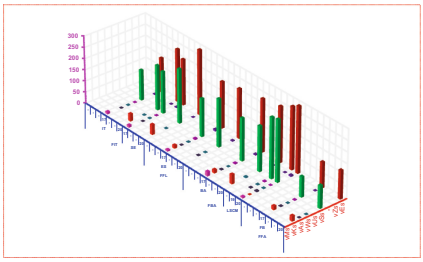

Fig. 13. Visual representation whole of the student intakes in course 17, 19 and 20 of majors belonging to the faculties in the university by visual graph representing of mD.

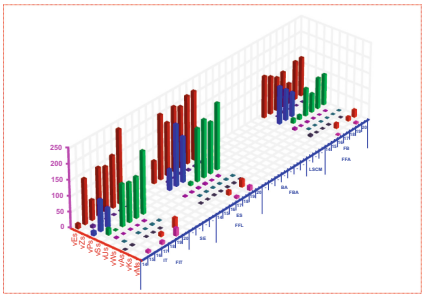

Fig. 14. Visual representation of the majors that are open for enrollment in all courses offered by the managed by the FIT, FFL and FFA faculties enclosed with all student intakes in the university.

exceed the training time, except for the course 20 of ES major, which there is no the student intakes exceed the training time. The analyst can easily see that the student intake exceeding the training time is much lower than the student intake studying in course 20 of majors.

5.3.2 Variable Questions

Using relational calculus expressions on data tuples combine with Q_{vs} to extract information and discover knowledge from mD of student intakes on the visual analysis system applying for question $2^{(2)}$ having given assumption is the student intakes entering and the student intakes changing majors. The conclusion of the analysis process is faculties, majors, the student intakes entering and the student intakes changing majors. The analyst views and thinks the visual graph in Fig. 8 and then conducts visual analysis of mD on this visual graph. To answer this question, they rely on the assumptions given to explore the student intakes entering and the student intakes changing majors in the university. The results of visual analysis of mD in Fig. 10 help the analyst to see the variation of the increasing the student intakes entering for each course 14, 15, 16, 17, 18, 19 and 20 in contrast to the student intakes changing majors, there is no variation by course and shows that the student intakes changing majors is not significant. From the results of this analysis, the university's leaders understand the period variation of the student intakes in courses which are most important to promptly invest in developing all resources for the units directly under the university.

5.3.3 Relationship Questions

To extract information and discover knowledge from mD of student intakes, this research combines Q_{rs} with relational calculus expressions on tuples of data on the visual analysis system to apply for question $1^{(1)}$ having given assumption is the courses 18, 19 and 20. The conclusion of the analysis process is faculties, majors and student intakes for each of 18, 19 and 20 courses respectively. The analyst views and thinks the visual graph in Fig. 8

and then conducts mD visual analysis on this visual graph. To answer this question, the analyst relies on the assumptions given for courses 18, 19 and 20 to explore the student intakes belonging to the majors of the faculties in the university. The analysis results in Fig. 11 help them to see that the student intakes entering and studying FB major of FFA faculty is much lower than that of BA, LSCM, ES, SE and IT majors of FBA, FFL and FIT respectively. In addition, it is easy for the analyst to recognize the largest student intakes in the FBA faculty, the student intakes of FIT faculty are the second largest, the student intakes of FFL faculty are the third largest, while the student intakes of FFA faculty are the lowest. From the results of this analysis, the university's leaders will see the correlation between the student intakes at the faculties to help them make appropriate financial investments policing for each faculty in the university.

5.3.4 Hierarchical Questions

Strategies for visual analysis of mD of student intakes using Q_{hs} combine with relational calculus expressions on domains of data on the visual analysis system to apply for question $12^{(12)}$ having given assumption is the student intake studying more than 180 students. The conclusion of the analysis process is the majors, courses, the student intake studying, the student intake entering and the student intakes remaining belonging to faculties. From Fig. 8 and the assumptions that this question has allowed the analyst to discover the student intake studying, the student intake entering and the student intakes participating visually performed in course 20 of IT major, in courses from 17 to 20 of SE major, in courses 18 and 20 of ES major, in courses 17, 19, 20 of BA major, and in courses from 18 to 20 of LSCM major (see result in Fig. 12). In addition to these two student intakes studying and entering, the student intake showed insignificant student intake remaining in courses of majors belonging to faculties in the university.

5.3.5 Visually Thinking Questions

Approaching to combine Q_{vts} with relational calculus expressions on tuples of data on the visual analysis system to apply for question $5^{(5)}$ having given assumption is the courses 17, 19 and 20. The conclusion of the analysis process is the majors, the faculty of charge, whole of student intakes involved in each of the majors. The analyst views and thinks the visual graph in Fig. 8 and then conducts visual analysis of mD on this visual graph. To answer this question, the analysts based on given assumptions help them easily to discover some important information such as no the student intakes graduated in courses 17, 19 and 20 also in these courses showing them that there are equal correlations the student intakes studying compared to the student intakes entering, see result in Fig. 13. Especially, the student intakes changing majors of IT, SE, BA and FB majors in course 20 did not occur, which proves that these majors are showing a very high resource demand of the labor market and a trend strong development direction in the future, on the contrary, these majors have the student intakes exceed the training time, which proves that these majors are very difficult to study, requiring students to have a high sense of self-study and to spend more time studying. In addition, the student intakes in the remaining courses 17, 19 and 20 occurred but has negligible intake.

5.3.6 Overall Questions

From on the visual analysis system combining Q_{os} with relational calculus expressions on data tuples to extract information and discover knowledge from mD of student intakes applying for question $7^{(7)}$ having given assumption is the FIT, FFL, FFA, and courses. The conclusion of the analysis process is the majors are open for training in all courses offered by the FIT, FFL, and/or FFA. The analyst views and thinks the visual graph in Fig. 8 combining with all components in the mdVAS, then they conduct VA of mD on this visual graph. From the assumptions that this question has allowed the analyst to discover all three faculties of FIT, FFL, or/and FFA having in turn IT, ES and FB majors are all open for training in all courses 14, 15, 16, 17, 18, 19 and 20. By observing the results of extracting and exploring IT, ES and FB majors, we see that the student intakes including vE_s, vP_s and vS_s of two majors IT and ES are much larger than those of the FB major, the remaining the student intakes including vK_s and vM_s of these three majors are quite similar. The student intakes vZ_s, vU_s, vW_s and vA_s remaining of the IT, ES and FB faculties all have intakes happening but are insignificant, so they do not affect the development of faculties and university, see result in Fig. 14.

6 Conclusions

This paper has systematized research works related to data mining, data visualization, data analysis processes, and visual analysis from different data sources. This research conducts analysis of educational data on student intakes, hierarchizes multivariate groups to create a hierarchical tree (hTree), and conducts restructuring the hierarchical tree to build a viewing - thinking model analyzing visually of mD of student intakes. Since then, this paper has focused on exploring vision organs, infographics capture characteristics, vision awareness progress, and visual viewing-thinking mechanisms of human to create a visual analytics system of multivariate data (mdVAS) of student intakes to help upgrade the human visual perception progress such as $vpcaurH_s$ in information and knowledge discovery, and find out new rules.

Visual graphs represent student intakes where there is a lot of latent information, containing a lot of new knowledge and rules that humans wish to discover. The analyst has the ability to observe visual graphs and then conducts data visual analysis by asking and answering corresponding questions based on relational calculations on tuples and domains of data, at the same time, flexible support of all components on a mdVAS, thereby not only helping analysts visually analyze data and compare the correlation between data variables to discover new patterns in student intakes according to courses of majors of faculties in the university so that there are solutions to promote the strengths as a premise to overcome the outstanding points in the university, but also especially support the leaders in the university to better understand the discovered information and knowledge to make financial policies and make timely decisions for each unit in the school. In addition, the mdVAS can be developed and extended to always adapt to multidimensional datasets, multiple data sources, and especially in exploiting data sources on school finances and comparing the learning outcomes of many students in the same subject, class, or the same course to help schools and parents have orientations to develop knowledge chains as well as career orientations for students in the future.

References

1. Keim, D.A., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H.: Visual analytics: scope and challenges. In: Simoff, S.J., Böhlen, M.H., Mazeika, A. (eds.) Visual Data Mining. LNCS, vol. 4404, pp. 76–90. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-71080-6_6
2. Thi Nguyen, H., Thi Pham, T.M., Thi Nguyen, T.A., Thi Tran, A.V., Vinh Tran, P., Van Pham, D.: Two-stage approach to classifying multidimensional cubes for visualization of multivariate data. In: Cong Vinh, P., Alagar, V. (eds.) ICCASA/ICTCC 2018. LNICSSITE, vol. 266, pp. 70–80. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-06152-4_7
3. Van Pham, D., Vinh Tran, P.: A system and model of visual data analytics related to junior high school students. In: Vinh, P.C., Rakib, A. (eds.) ICCASA/ICTCC 2019. LNICSSITE, vol. 298, pp. 105–126. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34365-1_9
4. Fairchild, M.D.: Color Appearance Models, 2nd edn. Association with the Society for Imaging Science and Technology (USA Wiley–IS&T Series in Imaging Science and Technology) (2005)
5. Few, S.: Tapping the Power of Visual Perception. Perceptual Edge, pp. 1–8 (2004)
6. Kuehni, R.G.: Color An Introduction to Practice and Principles, 3rd edn. Wiley, Hoboken (2013)
7. Few, S.: Data visualization for human perception. In: Soegaard, M., Dam, R.F. (eds.) The Encyclopedia of Human-Computer Interaction, Aarhus, Denmark, 2nd edn. (2014)
8. Bertin, J.: General theory, from semiology of graphics. In: Perkins, M.D.R.K.C. (ed.) The Map Reader: Theories of Mapping Practice and Cartographic Representation, pp. 8–16. Wiley (2011)
9. Green, M.: Toward a perceptual science of multidimensional data visualization: bertin and beyond. In: ERGO/GERO Human Factors Science. Citeseer (1998)
10. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 154–175. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70956-5_7
11. Feldkamp, N., Bergmann, S., Strassburger, S.: Visual analytics of manufacturing simulation data. In: The 2015 Winter Simulation Conference (WSC), Huntington Beach, CA, USA, 6–9 December 2015, pp. 779–790. IEEE (2015). https://doi.org/10.1109/WSC.2015.7408215
12. Francois, G., Gautron, P., Breton, G., Bouatouch, K.: Image-based modeling of the human eye. IEEE Trans. Visual. Comput. Graph. **15**(5), 815–827 (2009). https://doi.org/10.1109/TVCG.2009.24
13. Thi Nguyen, H., et al.: Integrating retinal variables into graph visualizing multivariate data to increase visual features. In: Vinh, P.C., Rakib, A. (eds.) ICCASA/ICTCC 2019. LNICSSITE, vol. 298, pp. 74–89. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34365-1_7
14. von Szent-Gyorgyi, A.: American (Hungarian-born) Biochemist (Nobel Laureate1937), "who said: Discovery consists of seeing what everybody has seen and thinking what nobody has thought," Edited by Irving Good. The Scientist Speculates (1962)
15. Andrienko, N., Andrienko, G.: Exploratory Analysis of Spatial and Temporal Data - A Systematic Approach. Springer, Heidelberg (2006)
16. Schneider, M.: Spatial and Spatio-Temporal Data Models and Languages. In: Liu, L., Özsu, M.T. (eds.) Encyclopedia of Database Systems, pp. 2681–2685. Springer, Boston (2009). https://doi.org/10.1007/978-0-387-39940-9_360
17. Van Pham, D., Vinh Tran, P.: Visually analyzing evolution of geographic objects at different levels of details over time. In: Cong Vinh, P., Alagar, V. (eds.) ICCASA/ICTCC 2018. LNICSSITE, vol. 266, pp. 98–115. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-06152-4_9

18. Kolbe, T.H.: Representing and exchanging 3D city models with CityGML. In: Lee, J., Zlatanova, S. (eds.) 3D Geo-Information Sciences. LNGC, pp. 15–31. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-87395-2_2
19. Tuan Anh, N.G., Tran Vinh, P., Vu, T.P., Van Pham, D., Sy, A.T.: Representing multiple levels for objects in three-dimensional GIS model. Presented at the the 13thInternational Conference on Information Integration and Web-based Applications & Service (iiWAS2011) (2011)
20. Löwner, M.O., Gröger, G., Benner, J., Biljecki, F., Nagel, C.: Proposal for a new LOD and multi-representation concept for CITYGML. ISPRS Ann. Photogrammetry Remote Sens. Spat. Inf. Sci. **IV-2/W1**, 3–12 (2016). https://doi.org/10.5194/isprs-annals-IV-2-W1-3-2016
21. Dang, P.V., Phuoc, T.V., Phuoc Tuyen, H.N.: Visual representation of geographic objects in 3D space at levels of different details. Presented at the FAIR - Fundamental and Applied Information Technology (2017)
22. Thomas, J.J., Cook, K.A.: Illuminating the Path the Research and Development Agenda for Visual Analytics. IEEE Computer Society Press, Los Alamitos (2005)
23. Sun, G.-D., Wu, Y.-C., Liang, R.-H., Liu, S.-X.: A survey of visual analytics techniques and applications: state-of-the-art research and future challenges. J. Comput. Sci. Technol. **28**(5), 852–867 (2013). https://doi.org/10.1007/s11390-013-1383-8
24. Andrienko, G., et al.: Space, time and visual analytics. Int. J. Geograph. Inf. Sci. **24**(10), 1577–1600 (2010). https://doi.org/10.1080/13658816.2010.508043
25. Vinh Tran, P., Xuan Le, T.: Approaching human vision perception to designing visual graph in data visualization. Concurr. Comput. Pract. Experience, no. Special Issue Paper (2020). https://doi.org/10.1002/cpe.5722
26. Thi Nguyen, H., Xuan Le, T., Vinh Tran, P., Van Pham, D.: An approach of taxonomy of multidimensional cubes representing visually multivariable data. In: Vinh, P.C., Rakib, A. (eds.) ICCASA 2019, pp. 90–104. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-34365-1_8
27. Nguyen, H.T., Tran, A.V.T., Nguyen, T.A.T., Vo, L.T., Tran, P.V.: Multivariate cube integrated retinal variable to visually represent multivariable data. EAI Endorsed Trans. Context-Aware Syst. Appl. **4**(12) (2017). https://doi.org/10.4108/eai.6-7-2017.152757
28. Card, S.T., Mackinlay, J.D., Scheiderman, B.: Readings in Information Visualization: Using Vision to Think. Academic Press (1999)
29. van Wijk, J.J.: The value of visualization. In: IEEE Visualization, VIS 2005, 23–28 October 2005, pp. 79–86 (2005). https://doi.org/10.1109/VISUAL.2005.1532781
30. Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F.: Mastering the Information Age: Solving Problems with Visual Analytics. Eurographics Association, Goslar (2010). https://diglib.eg.org/handle/10.2312/14803. ISBN 978-3-905673-77-7
31. Bertini, E., Tatu, A., Keim, D.: Quality metrics in high-dimensional data visualization: an overview and systematization. IEEE Trans. Vis. Comput. Graph. **17**(12), 2203–2212 (2011). https://doi.org/10.1109/TVCG.2011.229
32. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G., Keim, D.A.: Knowledge generation model for visual analytics. IEEE Trans. Vis. Comput. Graph. **20**(12), 1604–1613 (2014). https://doi.org/10.1109/TVCG.2014.2346481
33. Andrienko, G., Andrienko, N., Burch, M., Weiskopf, D.: Visual analytics methodology for eye movement studies. IEEE Trans. Vis. Comput. Graph. **18**(12), 2889–2898 (2012). https://doi.org/10.1109/TVCG.2012.276
34. Von Landesberger, T., et al.: Visual Analysis of large graphs: state-of-the-art and future research challenges. Comput. Graph. Forum **30**(6), 1719–1749 (2011)
35. Part of the Einstein exhibition. "A Matter of Time." The Museum. https://www.amnh.org/exhibitions/einstein/time/a-matter-of-time#:~:text=In%20the%20Special%20Theory%20of,on%20your%20frame%20of%20reference.&text=The%20faster%20a%20clock%20moves,a%20different%20frame%20of%20reference

36. Ning, Z.: Spatio-temporal cadastral data model: geo-Information management perspective in China. Master thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands (2006)
37. Codd, E.F.: Relational compleness of database sublanguages. In: Rustin, R. (ed.) Database Systems, pp. 65–98. Prentice Hall and IBM Research Report RJ 987, San Jose, California (1972)