

# **Journal of Digital Science**



**ISSN 2686-8296**

**Volume 2 Issue 2**

**December 2020**

**© Institute of Certified Specialists**

## CONTENTS

<b>Design and Evaluation of Visible Light Wireless Data Communication Models ..</b>	<b>3</b>
Simona Riurean	
<b>Predicting the Signs of the Links in a Network .....</b>	<b>14</b>
Quang-Vinh Dang	
<b>The origins of severe software defects method .....</b>	<b>23</b>
Nico Hillah	
<b>The development of a classification model of accounting in the digital economy .....</b>	<b>31</b>
Aleksei Kovalev	
<b>Development principles for preparing financial reporting in the context of digitalization .....</b>	<b>44</b>
Elena Dombrovskaya	

# Design and Evaluation of Visible Light Wireless Data Communication Models

Simona Riurean [0000-0002-5283-6374]

University of Petroșani, Petroșani 332006, Romania

[https://doi.org/10.33847/2686-8296.2.2\\_1](https://doi.org/10.33847/2686-8296.2.2_1)

Received 30.09.2020/Revised 01.12.2020/Accepted 21.12.2020/Published 29.12.2020

**Abstract.** The technologies based on the radio frequency used for wireless transmission indoor are more congested than ever these days, therefore several wireless communication alternatives are intensively searched. Some most promising technologies are based on the optical part of the electromagnetic spectrum. The benefits and drawbacks in optical wireless data communication models design are presented in this work. The key characteristics of the VLC systems are briefly discussed, as well. Some models developed for wireless data transfer in visible light communication and infrared are presented, and their operation related to the data rate and the length of the optical link, are briefly compared.

**Keywords:** Visible Light Communication, LED and PD key characteristics, Optical link, prototypes.

## 1. OWC Technologies and Applications

This article is an extension of the paper [1] presented at the ICCS 2020 conference.

The wireless transmission has gone through many paradigm changes after the advent of electromagnetic (EM) waves, remote communication such as telegraphy and the invention of radio [2, 3]. EM waves consists of radio, infrared (IR), visible, ultraviolet, etc. [2]. Today, the RF part, among these all, is the most used for communication purposes.

The RF communication does not require a Line of Sight (LoS) architecture between the transmitter (Tx) and receiver (Rx). The RF signal can travel long distances from Tx to Rx and also reach high data rates [3], however, faces some challenges such as interference, security issues as they can easily penetrate through solid objects, bandwidth and power limitations resulting in power inefficiency, lots of vulnerabilities in the respect of data security, and also affects the health and human wellbeing. Most important, the RF space has become increasingly crowded lately because of the many smart devices (desktops, laptops, tablets, smart phones and so on) connected simultaneously, especially indoor where the most data traffic occurs [4]. Therefore, developing alternative solutions to wireless RF transmissions especially indoor but also outdoor is imperious necessary to solve these congestions. The various optical wireless communication (OWC) technologies and applications developed so far (Infrared communication - IR, Visible Light Communication - VLC, Optical Camera Communication - OCC, Light Fidelity LiFi and Free Space Optics - FSO), emerge as a potential partner to the RF technologies (Wi-Fi, Bluetooth Low Energy, Cellular Data, ZigBee, 6LoWPAN, WiMax and so on – fig. 1.) and, in some specific scenarios, the major player. The IR Communication, VLC, OCC and LiFi, dedicated to indoor implementations and FSO aiming to transfer remote data outdoor, have been intensively researched lately. Today, these technologies and applications, early used as indoor positioning and tracking systems already tested and installed in

museums, airplanes, schools, underground, retail or medical facilities are about to become ubiquitous [5-8]. The instruments dedicated to human wellbeing and health, as smart watches, IR temperature measurement devices or finger pulse oximeters are these days more necessary than ever. IR wireless communication has been already searched for decades and some of its most used applications worldwide are the TV remote control, robots' control and positioning. The discovery of the white LED (in 1990 by Shuji Nakamura from USA and in 1992 by Isamu Akasaki and Hiroshi Amano from Ja-pan that received in 2014 the Nobel prize for physics [9]) triggered intensive research on VLC technology.

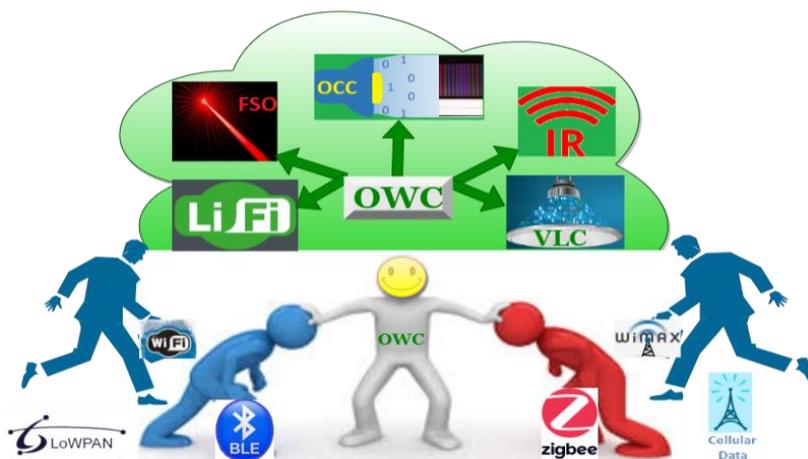


Fig. 1. Optical Wireless Communication versus RF technologies

OCC considered in IEEE 802.15.7m standardization along with VLC, covers the transmission between light sources and cameras for low rate communication and localization [10]. LiFi (coined by Harald Haas during a Ted conference in 2011) is a networked, MIMO, full duplex (VLC download and IR upload) wireless communication technology using access points embedded into the lighting fixtures [11].

The automotive applications such as Vehicle to Vehicle (V2V) communication, Vehicle to Infrastructure (V2I), Infrastructure to Vehicle (I2V) or Infrastructure to Infrastructure (I2I) and many other applications are more investigated than ever and aim to improve transport security, the human comfort, and become energy efficient [12]. V2V wireless communication is one of the most intense researched and advanced findings in FSO area [13].

### 1.1. Standardization of VLC

A significant step towards the widespread commercialization of VLC networks was the advent of the IEEE 802.15.7 standard in 2011. It defines three physical layers (PHY) (from PHY I to PHY III) and the Media Access Control (MAC) layer. PHY I layer is designed for outdoor scenarios, while PHY II and PHY III layers aims the indoor applications. A typical physical layer model of VLC is presented in figure 2.

There is a total of thirty combinations of modulation and coding schemes defined for VLC. PHY I are specified uses Reed-Solomon (RS) and Convolutional Codes (CC) for Forward Error Correction (FEC), while PHY II and III rely mostly on RS codes for FEC only.

Both PHY I and PHY II use On-Off Keying (OOK) and Variable Pulse Position Modulation (VPPM), while PHY III uses only color shift keying (CSK), thus PHY III is compatible only with Red Green Blue (RGB) LEDs.

Depending on the modulation used, the Run-Length Limited code (RLL), the optical clock rate, the FEC code, the three PHY modes, there are different data rates: 11.67 to 266.6 Kbps, 1.25 to 96 Mbps and 12 to 96 Mbps. The physical network topologies supported by the MAC layer are star, peer-to-peer, and broadcast. Three classes of devices are considered: vehicles, mobile and infrastructure.

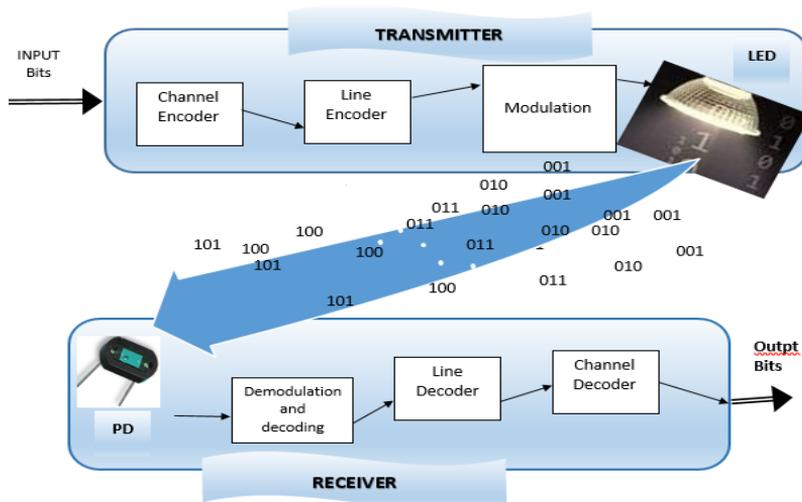


Fig. 2. Typical physical layer model of VLC

The MAC layer addresses characteristics as: support for mobility, dimming, color function and visibility, avoid flicker, support for pairing and disassociation in the VLC Personal Area Network (VPAN), generation of network beacons if the device is a coordinator, and reliability of the connection between MAC entities. Security considerations are applied at the application level. The MAC 802.15.7 layer is very close to the IEEE 802.15.4 MAC Personal Area Network (WPAN) [14].

However, many situations of using VLC technology are not considered in these standards, therefore, the IEEE Task Group 7m, first known as the IEEE 802.17.5r1, was initiated in late 2014 to revise the IEEE 802.15.7-2011 standard [15].

## 2. Key characteristics of oTx and oRx in a VLC setup

### 2.1. A general setup of the VLC system

A VLC system (regardless the topology setup) comprises the optical transmitter (oTx) and optical receiver (oRx), two separate modules responsible for sending/receiving data piggyback by the light.

Possible topologies in optical wireless transmissions (fig. 3) used for communication between transmitter (oTx) and receiver (oRx) can be a directed line of sight (LoS), non-directed LOS (NLoS), diffuse or quasi-diffuse links.

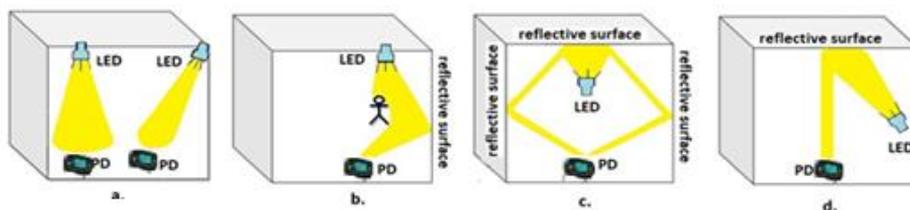


Fig. 3. a. Direct Line of Side (LoS) link, b. NLoS link, c. Diffuse link d. Quasi-diffuse link

The general diagram of a generic VLC system (fig. 4) consists of an oTx, the environment where the optical signal is affected by different noises, and the oRx.

Channel coding and modulation in oTx aim to adapt the signal to the channel and the part responsible for the conversion of the electrical signal into an optical one (E/O circuit in figure 4) is the main part of the oTx, the light source.

Noises into the environment (where the optical beam of light travels) affect the signal due to the ambient light. The ambient light can be the natural light and/or other sources of artificial light (fluorescent, incandescent bulbs or LEDs).

The photodetector (PD), part of the oRx module, is responsible to convert the optical signal into an electrical one (O/E circuit in Fig. 4).

The signal already converted is amplified and then demodulated and decoded to be recovered as accurately as possible.

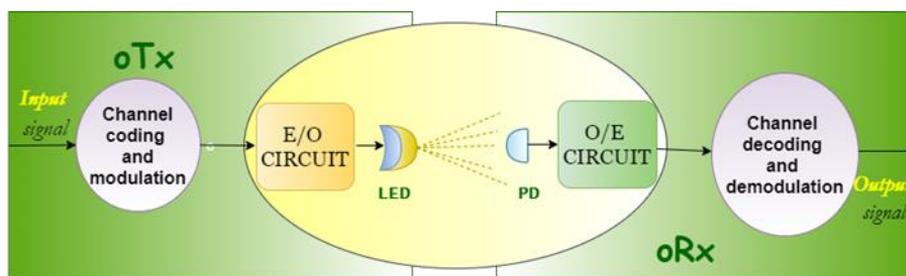


Fig. 4. A general diagram of a VLC system

## 2.2. LED's key characteristics for oTx setups

VLC systems can use different types of light sources such as white LEDs (WLEDs), array or stripe of SMD LEDs,  $\mu$ LEDs, OLEDs, Laser Diodes (LDs), according to the requirements of the application and specific environment. The most challenging characteristics of the LED for stable, mature VLC systems, are their nonlinearity and the limited modulation bandwidth. Numerous factors, such as quantum confined stark effect (QCSE), carrier lifetime, carrier recombination time, crystal orientation, etc. affect the modulation bandwidth of LEDs [16,17].

The modulation bandwidth of the VLC (generally limited to MHz) is directly influenced by the frequency response of the LED, therefore, negatively affects the data rates. It depends on the lifetime ( $\tau$ ) of the carrier recombination on the active region and on the p-n junction capacity. The frequency response of the off-the-shelf white LEDs is fundamental constraint by the slow temporal response of the yellow phosphor layer.

There are a number of proposals in literature for addressing the LEDs' non-linearity drawback in VLC systems. The simplest method is to use binary transmission such as the On-Off Keying (OOK) or Pulse Position Modulation (PPM), when the LED is simply switched on and off at specific current values. To mitigate this effect, higher spectral efficient modulation schemes such as multi-band Carrier-less Amplitude and Phase modulation (m-CAP), Orthogonal Frequency Division Multiplexing (OFDM) or color shift keying (CSK) have been used [18]. On the other hand, using Wiener model, although demonstrated only by simulation, a valid adaptive pre-distortion technique was proposed by Qian et. al. for the VLC systems. It consists of a linear time-invariant and memoryless nonlinear model, but with significant complexity VLC setup [19]. Since then, many other pre-distortion methods are being proposed [20,21].

A different solution is the use of a blue filter in front of the PD, to remove the slow response of the yellow phosphorus. Both the frequency response of the LEDs and the modulation bandwidth are related to the input current level and the junction and parasitic capacity. The LED's cut off frequency is also important. It refers to the maximum frequency at which the LED's light emission decreases to half the initial light intensity. Whether the transmission system is simplex or duplex, it is important to switch the LED between ON and OFF at high speed. The requirements for the signal that determines the output color are usually lower in terms of voltage compared to their power supply

When selecting a LED, for a robust VLC setup, the light intensity, the wavelength of light, the light scattering, the modulation speed and the power requirements must be considered.

For example, in case of using a stripe of Surface Mounted Diodes (SMD) LEDs (such as 3528s, 5050s, 2835s or 5630s) each LED can be addressed independently. Although the signal degradation occurs, the problem can be solved by connecting the LEDs to a series of synchronized controllers.

If the wavelength of LED is matched with that of the PD from the optical receiver (oRx), the noise level produced by the ambient light (natural and artificial) is decreased, therefore the correct reading of the optical signal is ensured. A high optical power requirement may also involve a cooling solution when the prototype is designed for a VLC network working in special scenarios with high ambient light noise.

### ***2.3. PD' key characteristics for oRx setups***

Light Dependent Resistors (LDRs), imaging sensors or a photodetectors (PDs) can be used in a VLC systems use as oRx a. PDs can support higher bandwidths for higher downlink speeds than camera-based VLC.

The Positive Intrinsic Negative Photodetectors (PIN PDs), and the avalanche photodiode (APD) are the highest sensitive semiconductor PD used in VLC systems that exploits the photoelectric effect to convert light into electricity. Silicon APDs are de-signed to offer high conversion gain, increased sensitivity and lower noise compared to standard PIN detectors but are more sensitive to extreme temperatures than PIN PDs [22]. On the other hand, APD produce a stronger signal but add more Poisson noise.

As Si PIN PDs have low voltage operation, high tolerance to a wide range of temperature fluctuations and excessive lighting, linear response characteristics, and low costs, are most often used in VLC systems.

PDs can be configured in two modes, photovoltaic and photoconductive. In photovoltaic mode the diode is unpolarized and the relationship between voltage and the amount of light is nonlinear. Usually, the photovoltaic panels operate in this way, being useful in producing electricity but limiting the sensitivity of the diode. Photovoltaic mode is preferable if we want to make accurate measurements since it is less susceptible to noise. The photoconductive mode applies a reverse bias voltage to the PD and the relationship is linear, up to 1mW optical power for a 1mm diameter diode.

The applied voltage produces a lower capacity and a higher speed of the diode to the detriment of heat production. Photoconductive circuits have applicability in bright light environments and where fast readings are needed.

### ***2.4. Boards used for VLC development***

The Arduino, Raspberry Pi, BeagleBone Black or STMicroelectronics Nucleo 32L0 are some of the development boards that are affordable for early prototyping and proved to have acceptable performance in different VLC topologies [23-27].

Arduino certainly has several options in the field of prototyping, their specialty being in robotics and small, localized systems. There are many options for compatible modules including sensors, prototyping boards, lighting, motors, and anything else that requires a short program to be controlled.

The reason the Raspberry Pi can be a proper choice for prototyping, is that it has the same behavior as a computer. It can connect to the internet, process text files, is compatible with peripheral devices (keyboard, mouse, screen), processes media files and is generally capable of anything as long as it needs enhanced processing or enhanced graphics capabilities. Last year, the newest Raspberry Pi 4B model was launched with significant improvements as 8Gb of RAM, 2 micro HDMI ports capable of 4k resolutions at 60 frames per second, Bluetooth 5.0, Gigabit Ethernet and a 1.5 GHz ARM v8 64bit processor.

The BeagleBone Black board has Sitara AM3358BZCZ100 processor, a Dual Data Rate RAM memory Micron 512MB DDR3L or Kingston 512mB DDR3 the physical interface to the network through SMSC Ethernet PHY and a high-definition multimedia interface (HDMI) Framer that provides control for an HDMI or DVI-D display with an adapter, as well as with various connectors, LEDs and switches (DC Power, power button, 10/100 Ethernet, serial debug port, four blue LEDs, USB, mini-USB, microSD, microHDMI BOOT switch and reset button).

STM Nucleo 32L0 board is designed around the STM32 microcontrollers in a 64-pin. It provides a high-power efficiency, with dynamic voltage scaling, an ultra-low-power clock oscillator and short wakeup time. The autonomous peripherals (including USART, I<sup>2</sup>C, and touch sense controller) reduce the load of the Arm® Cortex®-M0+ core leading to fewer CPU wakeups and contribute to decreased processing time and power consumption.

However, Raspberry Pi and Arduino are the main PCBs used in VLC prototyping to keep the features high and the cost low of the final product.

### **3. oTx and oRx modules for indoor data transfer**

The challenging journey from idea to testing a prototype has to follow few important steps [28].

The prototypes developed aim to test VLC system for indoor data transfer in a LoS topology under regular ambient light (sunlight and fluorescent light sources embedded into the lighting fixture).

In a more advanced scenario, an additive light source coming from a desk lamp with white LED placed over the VLC system aims to determine the influence of this additional noise on the system performance in terms of distance between oTx and oRx.

#### **3.1. VLC model with Arduino**

The prototype with the Arduino PCBs (Arduino Uno for the oTx and Arduino Mega for oRx) a text message was sent to a 400 mm distance (see fig. 5.). The oTx circuit was built with NPN transistors 2N3904 and 2SC5200, resistors and capacitors

The LED used is VLHW4100 with luminous intensity 90-110 LM, VF min 3.0 V max 3.6V, Max current 350 mA and CCT of 3200 - 3500K. The oRx circuit was built with the NE5532 Op Amp, resistors and capacitors. For the first scenario, a solar panel was used and then a PD type is VTB8440BH with IR filter (Fig. 5).

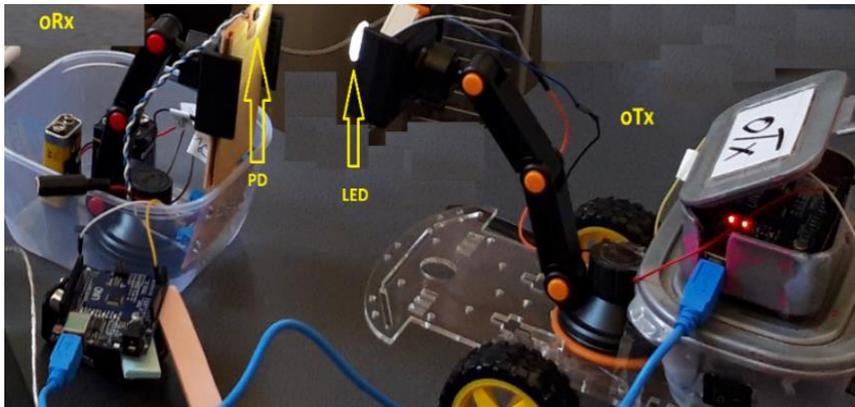


Fig. 5. VLC prototyped with Arduino PCB

### 3.2. VLC model with Raspberry Pi

The prototype with the Raspberry Pi PCB (Fig. 6. and 7.) has the LED type APA102C and a circuit with PNP transistors and resistors.

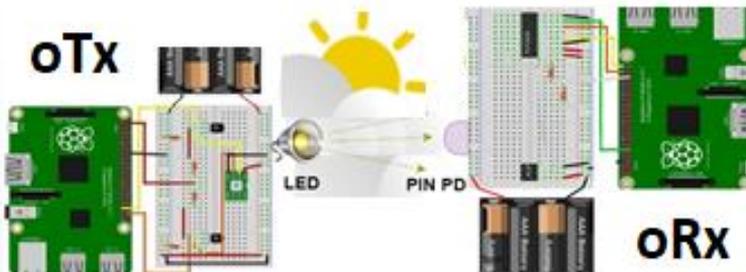


Fig. 6. oTx and oRx simulated with Raspberry Pi

The APA102C is an addressable LEDs that operates on +5V power input, as well as 0-5V logic levels for clock and data. It works on 2-wire communication protocol consisting of a clock line and a data line.

It requires one more wire than standard WS2812B LEDs because the communication with the LEDs becomes timing-independent. This feature allows it to be used directly with a Raspberry Pi, for example. The oRx circuit has the buffer/line driver 74HCT125 that aims to convert the optical signal into an electrical one and ALS-PT19-315C/L177/TR8 as optical detector.

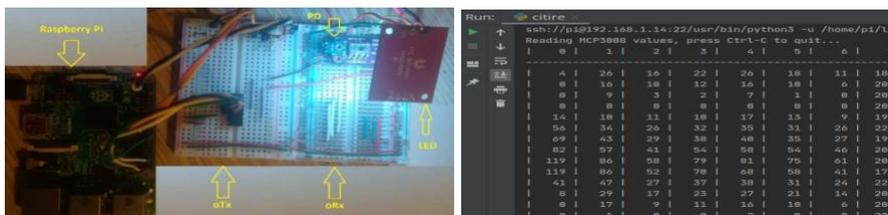


Fig. 7. VLC model with Raspberry Pi PCB and communication results

The ALS-PT19-315C/L177/TR8 is a low cost 2-SMD phototransistor. It has the sensitivity wavelength range ( $\lambda_p$ ) between 390 and 700. It has the peak sensitivity wavelength at 630 nm. The operating temperature is between -40 and +85 0C and the operating voltage 2.2 and 5 V. Its Rise time ( $t_r$ ) is 0.11 ms and the fall time ( $t_f$ ) 0.22 ms.

The output voltage ( $V_{out}$ ) is the product of photocurrent ( $I_{PH}$ ) and loading resistor ( $R_L$ ) (see eq. 1 ad eq. 2.).

$$V_{out(max.)} = I_{PH(max.)} \times R_L \leq V_{out(saturation)} = V_{CC} - 0.4V \tag{1}$$

$$R_L \times C_L \geq 0.5 \tag{2}$$

### 3.3. VLC model with BeagleBone Black

The BeagleBoard BeagleBone® Black (BBB) with ARM Cortex-A8 32bit RISC microprocessor, up to 1GHz. It has on board a HDMI to connect directly to monitors and/or TVs. The Linux Debian distribution can be flashed to the eMMC. It has a USB 2.0 client and host ports, a UART0 serial port access via 6 pin 3.3V TTL header and 10/100 RJ45 Ethernet connector [29].

OpenVLC 3.1 board is an open-source, low-cost and flexible VLC system platform. Research and development is coordinated by the Pervasive Wireless Systems group at IMDEA Networks Institute. It can achieve a throughput of 400 kbps at the Transport Layer at a distance of more than 4 meter, powered BBB. OpenVLC is code based and the use of VLC easy to adjust the behaviour of the platform by modifying the software code, such as introducing new media access control (MAC) protocols. Its interface is designed as a Linux kernel module [30].

OpenVLC's source codes are free: they can be redistributed and/or modified under the terms of the GNU General Public License.

The frame format built has the configuration as in Fig. 8.

Field	PRE	SFD	LT	Dst	Src	Protocol	Payload	CRC
Byte	4	1	2	2	2	2	0 ... 255	2

Fig. 8. The frame formats

The frame consists of 4B for preamble (PRE), 1B for the start of frame delimiter (SFD), 2B for the length (LT), 2B for the destination address, 2B for the source address, 2B protocol, between 0 and 255 B payload and 2B for the Cyclic Redundancy Check (CRC).

The error-correcting codes used are Reed–Solomon codes (RSc) that operate on a block of data treated as a set of finite-field elements called symbols, being able to detect and correct multiple symbol errors. By adding  $t = n - k$  check symbols to the data, a RSc can detect (but not correct) any combination of up to and including  $t$  erroneous symbols, or locate and correct up to and including  $\lfloor t/2 \rfloor$  erroneous symbols at unknown locations. The "reed-solomon.h" file with generic Reed Solomon encoder / decoder library is used for this purpose.

The VLC frame is prepared and then symbol flow is sent to the shared memory from where it is read by the firmware in the PRU. The PRU then controls the GPIOsto modulates the LED light for data transmission.

At the receiver, the light signals are detected by the PD and are sampled by firm-ware from PRUs. Once a valid PR and SFD are detected, the received data is sent to the shared memory and then received and processed by the OpenVLC driver. Finally, the received data is sent to the network layer, where it is processed using the Linux TPC / IP kernel.

The OpenVLC kernel module allows to create a network interface, meaning that any user is able to see the OpenVLC module as if it were just another network device, such as Wi-Fi or Ethernet, and any application to run would be connected via the VLC

network interface. The kernel runs in the BBB processor; its processing power is much higher than that of the PRU microcontroller. For this reason, the most computationally demanding tasks remain at the core level.

The designers used the Manchester line encoding. It encodes one bit into two sym-bols with On-Off-Keying (OOK) modulation (a symbol is either a HIGH or a LOW) ensuring that the average signal power remains constant. To avoid flickering, the Manchester line coding converts a "1" bit into a LOW-HIGH symbol pair and a "0" bit into a HIGH-LOW. Both RSc and Manchester encoding are also used in the 802.15.7-2011 standard.

The software solution is implemented as a Linux driver communicating directly with the OpenVLC cape and the Linux networking stack. Into the cape, the VLC interface is set up as a new communication interface, using a number of Linux tools.

The cape uses the duplex communication, having both the oTx and oRx functions. Once flashing the BBB, setting up the operating system, connecting the VLC cape and setting up the Tx and Rx, the system (OpenVLC cape with the BBB) behaves as a common network interface. With an open source script (online available on GitHub [31]), the interface's IP (IPv4) can be changed. Once the module driver is installed, the built-in microcontroller (PRU) can be boot up with the code necessary to transmit and received data over VLC and IR (Fig. 9).

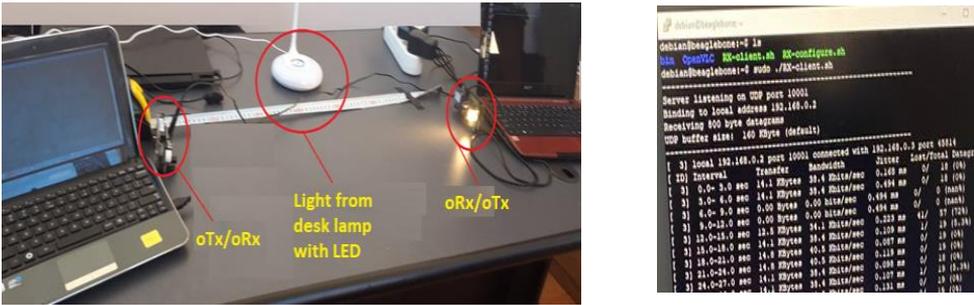


Fig. 9. Setup for data communication with BBB and OpenVLC Cap 1.3. and transmission results.

The most affordable model is the one with Arduino, but the most performant is the one with BBB. The model developed with Raspberry Pi, although more expensive than the model developed with Arduino, has a better throughput and a longer optical achievable link. Both models with Arduino and Raspberry Pi are simplex communication based solely on visible light and the OpenVLC cap with BBB has duplex transmission base both on IR and visible light.

A brief comparison between the three tested models is presented in Table 1.

Table 1. Comparison between the three tested models.

Nr Crt	oTx Board/ oRx Board	LED/PD	oTx PCB/ oRx PCB	Cost	Software/ Application	Optical Length [mm]
1	Arduino Uno/ Arduino Mega	VLHW4100/ VTB8440BH	2N3904; 2SC5200/NE5532	low	Arduino IDE	400
2	Raspberry Pi/ Raspberry Pi	APA102C/ ALS-PT19	74HCT125/MCP3008	moderate	Python	200
3	BeagleBone Black/ BeagleBone Black	XHP35/ SFH206K with TINA10645 lens	OpenVLC 1.3. ADS7883 LTC6269 FQPF30N06L	high	Debian GNU Linux/C PuTTY	1800

## 4. Conclusions

The key characteristics of the parts used for the models are important to be considered when the future manufactured prototype is simulated. There are many benefits, downsides and also challenges to overcome when VLC models are design with the off-the-shelf components especially when the prototype must have a low cost with planned throughputs. The models developed are tested in a LoS scenario under regular indoor ambient light (natural and fluorescent bulbs as artificial light from the lighting fixture), as well as a desk lamp with LED for additional noise, resulting acceptable performance. The link length is not the one targeted, but, as a first solution to keep low the entire budget of the project, and increase the VLC link is to use proper optics such as biconvex lenses in front of the LED and lenses with concentrators in front of PD to increase the optical gain. Although the prototypes tested have their particular benefits and drawbacks, with the current technological boom in this field, new, and improved spare parts are to be expected to become available on the market with low costs.

## References

1. Riurean S.: VLC Prototypes Developed with Off-the-Shelf Components for Wireless Indoor Data Transfer. In: ICCS 2020, LNNS, Vol.186. DOI: 10.1007/978-3-030-66093-2\_30.
2. Rehman, S. U.; Ullah, S.; Chong, P. H.; Yongchareon, S.; Komosny, D., Visible Light Communication: A System Perspective—Overview and Challenges. *Sensors*, 19 (5). (2019).
3. Leba M., Riurean S. and Ionica A., LiFi — The path to a new way of communication, 12th Iberian Conference on Information Systems and Technologies (CISTI), Lisbon, 2017, pp. 1-6, doi: 10.23919/CISTI.2017.7975997, (2017).
4. Marcu A., Dobre R. and Vlădescu M.: Investigation on available bandwidth in visible-light communications, 2016 IEEE 22nd International Symposium for Design and Technology in Electronic Packaging (SIITME), Oradea, 2016, pp. 244-247, doi: 10.1109/SIITME.2016.7777287.
5. Riurean, S., Antipova, T., Rocha, Á., Leba, M., Ionica, A. VLC, OCC, IR and LiFi Reliable Optical Wireless Technologies to be Embedded in Medical Facilities and Medical Devices. *J Med Syst* 43, 308 (2019).
6. Dimitrov S., Haas H.: Principles of LED Light Communications. Towards Networked Li-Fi, Cambridge University Press, (2015).
7. Riurean S., Antipova T., Rocha A., Leba M., Ionica A. (2019) Li-Fi Embedded Wireless Integrated Medical Assistance System. In: Rocha Á., Adeli H., Reis L., Costanzo S. (eds) *New Knowledge in Information Systems and Technologies. WorldCIST'19 2019. Advances in Intelligent Systems and Computing*, vol 931. Springer, Cham. [https://doi.org/10.1007/978-3-030-16184-2\\_34](https://doi.org/10.1007/978-3-030-16184-2_34).
8. Riurean S., Olar M., Leba M., Ionica, A.: Underground positioning system based on visible light communication and augmented reality, 17th International Technical-Scientific Conference on Modern Technologies for the 3rd Millennium, Oradea, ROMANIA, MAR 22-23, 2018, Pages: 345-350, <http://orcid.org/0000-0002-5283-6374>, (2018).
9. Nobelprize Homepage, <https://www.nobelprize.org/prizes/physics/2014/press-release/>, last accessed 2020/08/25.
10. Marcu A. E., Dobre R. A. and Vlădescu M., Flicker Free Optical Camera Communication for Cameras Capturing 30 Frames per Second, 43rd International Conference on Telecommunications and Signal Processing (TSP), Milan, Italy, 2020, pp. 166-169, (2020).
11. Ted Homepage [https://www.ted.com/talks/harald\\_haas\\_wireless\\_data\\_from\\_every\\_light\\_bulb](https://www.ted.com/talks/harald_haas_wireless_data_from_every_light_bulb), last accessed 2020/08/25
12. Avătămăniței, S.A.; Căilean, A.-M.; Beguni, C.; Dimian, M.; Popa, V. Analysis concerning the usage of Visible Light Communications in Automotive Applications: achievable communication distances vs optical noise. *International Conference on Development and Application Systems*, (2020).
13. Căilean, A.M., Dimian, M., Done, A., Enhanced design of visible light communication sensor for automotive 785 applications: Experimental demonstration of a 130meters link. *Global LIFI Congress (GLC), Paris, 2018, 1-4. 786 (2018).*

14. IEEE Computer Society. IEEE Standard for Local and metropolitan area networks - Part 15.7: Short-Range Wireless Optical Communication Using Visible Light. Number September (2011).
15. IEEE 802.15 WPAN 15.7 Amendment Study Group (2015).
16. Huang Y.M., Singh K.J., Liu AC, et.al. Advances in Quantum-Dot-Based Displays. *Nanomaterials* (Basel). Jul 6;10(7):1327. (2020).
17. Liao C. et al., Light-emitting diodes for visible light communication, International Wireless Communications and Mobile Computing Conference (IWCMC), Dubrovnik, 2015, pp. 665-667, (2015).
18. Burton, et. al., Investigation into Using Compensation for the Nonlinear Effects of the Output of LEDs in Visible Light Communication Systems, 2nd West Asian Colloquium on Optical Wireless Communications (WACOWC), Tehran, Iran, 2019, pp. 80-84 (2019).
19. Qian H., Yao S. J., Cai S. Z, and Zhou T., Adaptive Postdistortion for Nonlinear LEDs in Visible Light Communications, *IEEE Photonics Journal*, vol. 6, pp. 1- 8, (2014).
20. Aggarwal P., Ahmad R., Bohara V. A. and Srivastava A., Adaptive predistortion technique for nonlinear LED with dimming control in VLC system, 2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Bhubaneswar, (2017).
21. Ying Kai, Yu Zhenhua, Baxley Robert j., Qian H., Chang G.-K., and Zhou T. Nonlinear Distortion Mitigation in Visible Light Communications, *IEEE Wireless Communications* 22(2) April (2015).
22. Jie Lian, Zafer Vatanserver, Mohammad Noshad and Maité Brandt-Pearce, Indoor visible light communications, networking, and applications, *Journal of Physics: Photonics*, Volume 1, Number 1 (2019).
23. Riurean S.M., Nagy A.A, Leba M. and Ionic A. C.: A small step in VLC systems – a big step in Li-Fi implementation, IOP Conference Series: Materials Science and Engineering, Volume 294, International Conference on Applied Sciences (ICAS2017) 10–12 May 2017, Hunedoara, Romania (2017).
24. Swain K.P., Prasad M.V.S.V., Palai G., Sahoo J., Mohanty M.N., Exploiting VLC Technique for Smart Home Automation Using Arduino. In: Dash S., Vijayakumar K., Panigrahi B., Das S. (eds) *Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing*, vol 517. Springer, Singapore. (2017).
25. Wang T. and Zhao Z., An implementation of visible light communication based on Raspberry Pi, IEEE Integrated STEM Education Conference (ISEC), Princeton, NJ, 2018, pp. 218-219, (2018)
26. Riurean S., Leba, M. Ionica A. and Nassar Y. Technical Solution for Burnout, the Modern Age Health Issue, IEEE 20th Mediterranean Electrotechnical Conference (MELECON), Palermo, Italy, 2020, pp. 350-353, (2020).
27. Duque A., Bidirectional Visible Light Communications for the Internet of Things. *Networking and Internet Architecture [cs.NI]*. Université de Lyon - INSA Lyon, (2018).
28. Rosca S., Riurean S., Leba M., Ionica A., An Educational Model of Graduation Project for Students at Automation and Computer Engineering, *Journal of Digital Science*, Vol.1, Iss. 1, Dec (2019). DOI: 10.33847/2686-8296.1.1\_4.
29. Nayyar A., Puri V.: Comprehensive Review of BeagleBone Technology: Smart Board Powered by ARM, *International Journal of Smart Home* 10(4):95-108 (2016).
30. Galisteo A., Juara D., Giustiniano D.: Research in Visible Light Communication Systems with OpenVLC1.3, *Networking and Internet Architecture (cs.NI)* arXiv:1812.06788 (2019).
31. [Online] <https://github.com/openvlc/openvlc>, last accessed 10.11.2020.

# Predicting the Signs of the Links in a Network

Quang-Vinh Dang [0000-0002-3877-8024]

Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

[https://doi.org/10.33847/2686-8296.2.2\\_2](https://doi.org/10.33847/2686-8296.2.2_2)

Received 30.09.2020/Revised 01.12.2020/Accepted 21.12.2020/Published 29.12.2020

**Abstract.** It is hard to deny the importance of graph analysis techniques, particularly the problem of link and link-sign prediction, in many real-world applications. Predicting future sign of connections in a network is an important task for online systems such as social networks, e-commerce, scientific research, and others. Several research studies have been presented since the early days of this century to predict either the existence of a link in the future or the property of the link. In this study we present a novel approach that combine both families by using machine learning techniques. Instead of focusing on the established links, we follow a new research approach that focusing on no-link relationship. We aim to understand the move between two states of no-link and link. We evaluate our methods in popular real-world signed networks datasets. We believe that the new approach by understanding the no-link relation has a lot of potential improvement in the future.

**Keywords:** Signed Network, Machine learning, Link Prediction.

## 1. Introduction

Many real-world systems can be modelled by a signed-directed graph [14,39], such as connection graph between users on social networks like Facebook, citation graphs in scientific publication, recommendation graph in e-commerce, or the activities like view/like/dislike graphs on YouTube. The other application domains include sociology, biology, transportation and operational research [5, 9, 34,36,22].

Recommendation systems, in particular, play a crucial role in many social networks and e-commerce websites today [49]. For instance, it is estimated that 30% of product views on Amazon [47] and 80% of views on Netflix [15] are generated via recommendation. Definitely the recommendation engine should be able to understand the users to recommend the service that the user want to use or the item that the user want to buy. Besides the well-known collaborative filtering technique [49], a few research studies look at the problem is to consider the recommendation as the link-attribute prediction problem [14], i.e. we would like to predict the property of the link between two nodes. For instance, in ecommerce systems like Amazon, if we consider buyers and items are nodes and connections between them are the rating scores the buyers give to the items, link-attribute prediction shares a same goal with collaborative filtering.

In this study, we limit ourselves to the problem of link-sign prediction, i.e. we consider only two possible attributes of a link which are *positive or negative* [28]. We can think of positive links as a node (user on Facebook, buyer on Amazon, viewer on YouTube) like another node, and negative links as dislike relation. If we can predict accurately the signs of the links, we can improve significantly the quality of the recommender systems [13].

The problem of link-sign prediction has attracted a lot of studies in the last two decades. However, the problem is divided into link-prediction and sign prediction separately. In this study we consider unifying the two problems into one single framework.

## 2. Related Works

The problem of link-sign prediction is usually divided into the link-prediction problem [32] and the sign-prediction problem [28]. The link-predictor aims to predict the existence of the link in the future, given the non-existence of this link at present. The sign-predictor aims to predict the sign of a link in the future, given that the existence of the link is confirmed. The difference is visualized in Fig. 1.

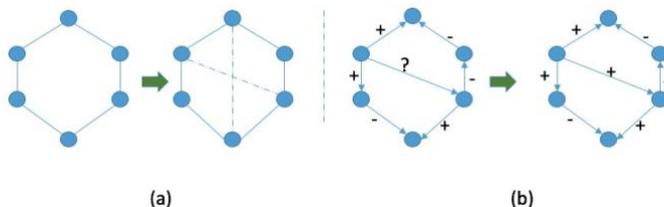


Fig.1. (a) Link-prediction; and (b) Sign-prediction

### 2.1. Link-Prediction Problem

Link-prediction can be roughly divided into proximity-based and learning-based methods [31]. Proximity-based methods mostly rely on homophily theory [37] which states that similar users tend to connect to each other's. The proximity metrics can be nodal proximity-based metrics [61] or structural proximity-based metrics [32].

The nodal-based proximity methods use the personal profile information of the nodes to design the proximity metric. For instance, the authors of [60] proposed to use demographics information such as age, education, and occupation. In [57] the authors also used geographic information in physical space to estimate the distances between two nodes in cyber-space. Several studies employed user activities on social networks such as keywords, hashtags or semantic of the content that the users share or view on social network to calculate their similarity [63]. Among different metrics designed in different studies, cosine similarity is mostly used as the proximity metric [31]. Other popular distance metrics have been used in literature are KL-divergence [48] or Jaccard's coefficient [46].

Structural-based methods do not rely on the individual information of the nodes but based on the topological information of the graph [32]. An early attempt is credited to [1] that defines the Adamic-Adar measure. The metric is inspired by the Erdos-Renyi model [3]. Both theoretical and empirical studies showed that the likelihood of a link depends on the size of direct neighborhood of a node [32]. Other well-known structural-based measures that rely on neighborhood information are SimRank [24], Sorensen Index or HP/HD Index [43].

Another kind of structural-based proximity is to rely not on the neighborhood information but on the connection between two nodes. The core idea is to calculate the distance between two nodes as the surrogate estimation of how close they are to each other, hence how likely that they will connect. One of very early approaches belongs to [25]. The core idea of Katz index [25] is to combine all connected paths between two nodes with associated weights as their corresponding length. The authors of [69] updated the Katz index to use local-path information. Other studies [55] modified the PageRank algorithm for the link prediction problem. Recently, [62] added the interaction among paths, i.e., the authors consider the interaction between nodes that belong to different paths into consideration, to achieve a better prediction.

[6] explored a very new and interesting aspect of the network, which is the uncertainty of the network topology, i.e., when we are not 100% sure if there is a link or no-link between two nodes.

On the other hand, learning-based link-prediction methods build machine learning models to predict the probability that a link will be established in the future. To build the machine learning models, features needed to be extracted. The features might be similar to the similarity metrics described above, i.e. neighbor-based features such as common friends [4] or path-based features such as the shortest path between two nodes [33]. The most common algorithms are logistic regression [40] or SVM [64].

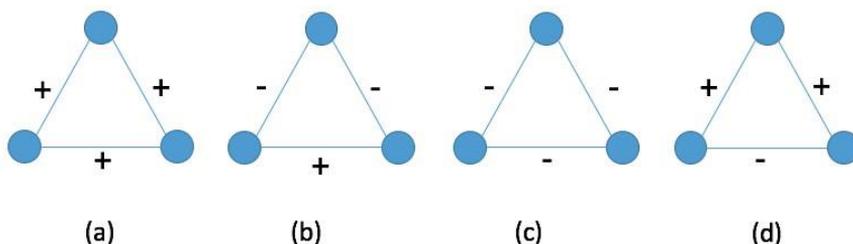
In recent years, many studies utilize the power of deep learning methods for representation learning [20]. Instead of manually defining the graph features, the authors train a model to learn the features [39]. Similar to the embedding techniques in natural language processing field that tries to keep the similar words close to each other's [38], graph embedding techniques try to keep the similar nodes close to each other in a low-dimension space [17]. Some prominent studies include the graph neural networks [45], graph convolutional neural networks [27], Graph Gated Neural Networks [52,67] and Graph Attention Networks [56,51, 68,59,58].

## 2.2. Sign-Prediction Problem

In contrast to link-prediction problem where the problem is formed as predicting whether a link will be established in the future or not, the sign-prediction problem tries to predict the attribute of the link, here positive or negative sign, given the existence of the link in the future [19].

Early attempts in sign-prediction used rule-based systems, such as "friends of friends are friends" and "friends of enemies are enemies" for prediction [19]. To speed up the calculation, the prediction is performed as matrix multiplication. The following studies turned the problem into a low-rank matrix factorization problem [35] as in the recommender systems.

Starting with the work of [28,65], structural balance theory and social status theory are used to interpret the structure of the social graphs hence to predict the sign of the connections. We visualized two theories in Figure 2 and 3 [14, 8]. Several studies have followed the work of [28] such as [7] by extending the cycle length or [23] by integrating the social theories into matrix factorization techniques. In the same direction, [26] focused on micro-structure of three users with bidirectional links and their similarities, for the sign prediction.

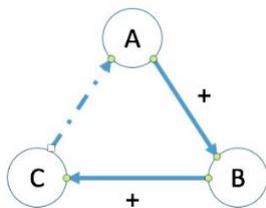


**Fig.2.** Visualization of structural balance theory [28].

*Structural balance theory* considers the triads (a) and (b) are balanced, while (c) and (d) are not. The *weak balance theory* considers (c) as balanced as well, regardless the direction.

The authors [50] argued that, (i) even structural balance theory and social status theory played an important role in existing research studies on link sign

prediction, they are not scalable up to the size of the modern networks, and (ii) a fully observed network is usually not available in practice. They develop Bayesian node features based on partially observed networks and used a logistic regression classifier for link sign prediction. However, the obtained performance is weak compared to other recent studies.



**Fig.3.** The sign of the dash line from C to A is inferred by their social status. Because  
 $A \rightarrow B$  and  $B \rightarrow C$ , therefore hence the *social status theory* states that  $B > A$  and  $C > B$ , so  $C > A$ , leads to the prediction that the sign of line from C to A is negative.

Graph embedding techniques attracted a lot of attention recently [18]. Many graph embedding research studies are inspired by word embedding techniques [42] as they share a similar goal. The authors of [14] combined word embedding technique, random walk, and Long-Short Term Memory (LSTM) networks to utilize the time information of the network for sign-prediction.

As pretrained models are more and more popular in natural language processing community and achieve a lot of success, transfer learning has been used in sign-prediction as well. This direction has a lot of potential usage as it does not require labelled data [66]. An interesting application of the sign-prediction is presented in the work of [2] that applied the sign prediction in basketball prediction.

Most existing works in sign-prediction assumes the previous task which is link-prediction has been done, so they assume the existence of the link and focus only on the sign of the link.

### 3. Understanding No-Link

As we discussed in Section 2, the work of link-sign prediction is usually divided into two steps: link-prediction and sign-prediction [29]. Given that the performance of state-of-the-art link-prediction algorithms are far from perfect [31], it is not realistic to expect a high-accuracy prediction in sign-prediction if we consider the task link-sign prediction as an entire task.

Given that most of a network should be no-link [14], it is important to understand the no-link relationship to achieve a higher-accuracy prediction. Hence, we focus on the *move* from no-link into link between two nodes. We collected features based on sociology theories and other studies [29] such as Balance Theory, Status Theory, Reciprocity or Frequent Subgraph [28]. We also derived features automatically through some graph embedding techniques such as DeepWalk [42], LINE [53] and node2vec [18]. We try to capture the semantic of the network just before a link is established. Our approach shares some similarities with the work of [30] as we combine both explicit and implicit features, but different from [30] we treated different links with different weights. Furthermore, in [30] the authors simply add two explicit and latent scores then solving an optimization problem, so they consider two sets of features as the same weight without any learning process.

Regarding the no-link relations, we consider the time when a node decides to make a link to a particular other node but no other nodes. At this very point of time,

we consider that it means the first node decides to not make a link to other nodes. We collected the same features but with label of no-link.

Instead of treating the problem as the traditional classification problem [28, 14] we convert the problem into the ranking problem. We use a state-of-the art ranking algorithm implemented in XGBRanker version 1.0.0 <sup>1</sup> which has just been released at the time of writing. In our opinion, this approach makes more sense as what we need to optimize in practice, such as in an e-commerce system, is the ranking metric, i.e., which item a user will most likely like, rather than a normal classification problem [29]. Indeed, we can use the output of any classification problem as the ranking score but actually they are optimized for different purposes.

## 4. Experiments

### 4.1. Datasets

We used well-known signed network datasets that have been used widely in the literature: Epinions, Slashdot and Wikipedia RfA [14,28]. They are real-world signed networks collected from e-commerce website (Epinions), sharing news website (Slashdot) or crowd-sourcing encyclopedia (Wikipedia). In these websites, users can form explicitly trust (positive) / distrust (negative) links to other users. The connections are provided with the established time that allow us to understand the history of the network. We display some basic statistics of the dataset in Table 1 [14].

Table 1. Basic statistics of datasets. WCC stands for *weakly connected component*.

	<b>Epinions</b>	<b>Slashdot</b>	<b>Wikipedia</b>
# of nodes	119 217	82 140	7 118
# of edges	841 200	549 202	103 747
fraction of edges	6e-5	8e-5	2e-3
+ edges (%)	85.0	77.4	78.8
- edges (%)	15.0	22.6	21.2
largest WCC (%)	99.1	100	100
average # of directed connection	590	327	418
# of triads	13 375 407	1 508 105	790 532
fraction of triads	1.35e-10	5.46e-11	4.25e-9

The authors of [30] studied the change from no-link to link between two nodes given they have common neighbors. The analysis is presented in Figure 4 for Epinions dataset. It is agreed between several recent studies [30,29] that the process of moving from no-link to link is important (Fig.4).

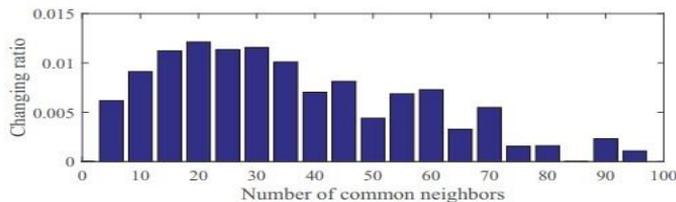


Fig.4. Changing ratio given the common neighbors between two nodes [30]. The X-axis represents the number of common neighbors in 15 months and the Y-axis represents the proportion of two nodes that actually formed a link.

<sup>1</sup> <https://xgboost.readthedocs.io>

For each dataset, we divide the train/validation/test set according to the ratio of 60 : 20 : 20 by temporal order, i.e. we use 60% of the links that are established first in each network as the training set, 20% of the links that are established next for the validation set, and 20% of the links that are established last for testing purpose.

To avoid over-fitting, we apply early stopping technique. Other hyperparameters are optimized by using Bayesian optimization.

## 4.2 Evaluation Metric

As the matter of fact, the network in general is extremely balanced: most of the potential links do not exist. As we present in Table 1, there is around only less than 0.1% of the potential links in the network, i.e., a connection that might be exist between any two nodes, is established in a network. It is understandable because usually a typical person can only make a few connections, and even if the degree of a node is thousands it is still a very small number compared to billions of nodes in modern networks like Facebook.

Some popular metrics which are sensitive to the imbalanced dataset such as accuracy score or even F-score are not usable to evaluate link prediction problem [12]. AUC is used more widely [14] but the AUC is designed only for a binary classification problem. In this paper we will use a more generalized AUC metric designed by [21] which has been adapted in multi-class classification problem [11,12]. The generalized AUC is chosen because it is based on the same principle of measuring the ranking order as in typical AUC. Like typical AUC, the generalized AUC ranges from 0.0 to 1.0, higher is better.

## 4.3 Experimental Results

We compare our method with several other baseline methods, such as multi-class classification [14]. We note that it is our method is not comparable to existing works because it focuses on both link-prediction and sign-prediction at the same time.

Here, we consider the problem of predicting at a given point of time, given that there is no link between node A and node B, which action will be performed by node A: (i) establish a positive link, or (ii) establish a negative link, or (iii) do not establish a link.

**Table 2.** Generalized AUC of our method in different datasets.

Methods	Epinions	Slashdot	Wikipedia
Classification	0.72	0.74	0.77
Ranking	0.74	0.75	0.79

The result is presented in Table 2. We notice that in general the ranking approach achieve a better performance as the AUC by its nature is optimized by a ranking method.

Like other link-prediction [28] or sign-prediction work [14], both methods perform best in Wikipedia dataset. The reason might be the topological structure of Wikipedia that is much more dense than other networks, but we let it be an open question for further research works.

## 5. Conclusion

In this study we presented our work on link-sign prediction in signed networks. Together with the preliminary work of [29], they are very first working in considering link-prediction and sign-prediction into one problem. By combining the problems, we eliminate the accumulated error over multiple prediction steps. We note that the approach presented in this paper is a topological-based approach [54], i.e., we only rely on the features that can be derived from the topology of the network and we do not use any personal information of node such as gender or activity history.

In the future research we plan to incorporate more temporal features as described in [14,39,41,16,44]. We believe that by understanding the movement from no-link to a link in a network we will gain more insights of the link-sign prediction problem. In other words, we would like to know what bring a node to a very precise point of time before he or she decide to form a link or not, not only just before a connection is made. Another interesting research direction is to predict the time when a link is made rather than considering this time as a given input. On the other hand, we would like to incorporate trust calculation to the network as the links as the interaction between nodes [10].

## References

1. Adamic, L.A., Adar, E.: Friends and neighbors on the web. *Social Networks* 25(3), 211–230 (2003)
2. Ahmadalinezhad, M., Makrehchi, M., Seward, N.: Basketball lineup performance prediction using network analysis. In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. pp. 519–524 (2019)
3. Barabási, A.L., et al.: *Network science*. Cambridge university press (2016)
4. Benchettara, N., Kanawati, R., Rouveirol, C.: A supervised machine learning linkprediction approach for academic collaboration recommendation. In: *RecSys*. pp. 253–256. ACM (2010)
5. Boffey, T.: *Graph theory in operations research*. Macmillan International Higher Education (1982).
6. Chen, X., Guo, J., Pan, X., Zhang, C.: Link prediction in signed networks based on connection degree. *J. AIHC* (2019)
7. Chiang, K., Natarajan, N., Tewari, A., Dhillon, I.S.: Exploiting longer cycles for link prediction in signed networks. In: *CIKM*. pp. 1157–1162. ACM (2011)
8. Dang, Q.V.: *Trust assessment in large-scale collaborative systems*. Ph.D. thesis, University of Lorraine, France (2018)
9. Dang, Q.V.: Link-sign prediction in signed directed networks from no link perspective. In: *International Conference on Integrated Science*. pp. 291–300. Springer (2020). DOI: 10.1007/978-3-030-49264-9\_26.
10. Dang, Q.V., Ignat, C.L.: Computational trust model for repeated trust games. In: *Trustcom/BigDataSE/ISPA*. pp. 34–41. IEEE (2016)
11. Dang, Q.V., Ignat, C.L.: Measuring quality of collaboratively edited documents: The case of wikipedia. In: *CIC*. pp. 266–275. IEEE Computer Society (2016)
12. Dang, Q.V., Ignat, C.L.: Quality assessment of wikipedia articles without feature engineering. In: *JCDL*. pp. 27–30. ACM (2016)
13. Dang, Q.V., Ignat, C.L.: dTrust: A simple deep learning approach for social recommendation. In: *CIC*. pp. 209–218. IEEE (2017)
14. Dang, Q., Ignat, C.: Link-sign prediction in dynamic signed directed networks. In: *CIC* (2018).
15. Gomez-Uribe, C.A., Hunt, N.: The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)* 6(4), 1–19 (2015)
16. Goyal, P., Chhetri, S.R., Canedo, A.: dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems* 187, 104816 (2020)
17. Goyal, P., Ferrara, E.: Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems* 151, 78–94 (2018)

18. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: KDD, pp. 855–864. ACM (2016).
19. Guha, R.V., Kumar, R., Raghavan, P., Tomkins, A.: Propagation of trust and distrust. In: WWW (2004)
20. Hamilton, W.L., Ying, R., Leskovec, J.: Representation learning on graphs: Methods and applications. arXiv preprint arXiv:1709.05584 (2017)
21. Hand, D.J., Till, R.J.: A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning* 45(2), 171–186 (2001)
22. Harary, F., Norman, R.Z.: Graph theory as a mathematical model in social science. No. 2, University of Michigan, Institute for Social Research Ann Arbor (1953)
23. Hsieh, C., Chiang, K., Dhillon, I.S.: Low rank modeling of signed networks. In: KDD. pp. 507–515. ACM (2012)
24. Jeh, G., Widom, J.: Simrank: a measure of structural-context similarity. In: KDD. pp. 538–543. ACM (2002)
25. Katz, L.: A new status index derived from sociometric analysis. *Psychometrika*(1953)
26. Khodadadi, A., Jalili, M.: Sign prediction in social networks based on tendency rate of equivalent micro-structures. *Neurocomputing* p. 10 (2017)
27. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: ICLR. p. 14 (2017)
28. Leskovec, J., Huttenlocher, D.P., Kleinberg, J.M.: Signed networks in social media. In: CHI. pp. 1361–1370. ACM (2010)
29. Li, X.: Towards practical link prediction approaches in signed social networks. In: UMAP (2018)
30. Li, X., Fang, H., Zhang, J.: FILE: A novel framework for predicting social status in signed networks. In: AAAI. pp. 330–337. AAAI Press (2018)
31. Li, Z.L., Fang, X., Sheng, O.R.L.: A survey of link recommendation for social networks: Methods, theoretical foundations, and future research directions. *ACM Trans. Management Inf. Syst.* (2018)
32. Liben-Nowell, D., Kleinberg, J.M.: The link-prediction problem for social networks. *JASIST* (2007)
33. Lichtenwalter, R.N., Lussier, J.T., Chawla, N.V.: New perspectives and methods in link prediction. In: KDD. pp. 243–252. ACM (2010)
34. Likaj, R., Shala, A., Mehmetaj, M., Hyseni, P., Bajrami, X.: Application of graph theory to find optimal paths for the transportation problem. *IFAC Proceedings Volumes* 46(8), 235–240 (2013)
35. Ma, H., Yang, H., Lyu, M.R., King, I.: Sorec: social recommendation using probabilistic matrix factorization. In: CIKM. pp. 931–940. ACM (2008)
36. Mason, O., Verwoerd, M.: Graph theory and networks in biology. *IET systems biology* 1(2), 89–119 (2007)
37. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. *Annual review of sociology* 27(1), 415–444 (2001)
38. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
39. Mutlu, E.C., Oghaz, T.A.: Review on graph feature learning and feature extraction techniques for link prediction. arXiv preprint arXiv:1901.03425 (2019)
40. O'Madadhain, J., Hutchins, J., Smyth, P.: Prediction and ranking algorithms for event based network data. *ACM SIGKDD explorations newsletter* (2005)
41. Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T.B., Leiserson, C.E.: Evolvegn: Evolving graph convolutional networks for dynamic graphs. In: AAAI. pp. 5363–5370 (2020)
42. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: online learning of social representations. In: KDD. pp. 701–710. ACM (2014)
43. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.L.: Hierarchical organization of modularity in metabolic networks. *Science* (2002)
44. Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., Bronstein, M.: Temporal graph networks for deep learning on dynamic graphs. arXiv preprint arXiv:2006.10637 (2020)
45. Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G.: The graph neural network model. *IEEE Transactions on Neural Networks* 20(1), 61–80 (2008)
46. Scellato, S., Noulas, A., Mascolo, C.: Exploiting place features in link prediction on location-based social networks. In: KDD. pp. 1046–1054. ACM (2011)

47. Sharma, A., Hofman, J.M., Watts, D.J.: Estimating the causal impact of recommendation systems from observational data. In: Proceedings of the Sixteenth ACM Conference on Economics and Computation. pp. 453–470 (2015)
48. Shen, D., Sun, J., Yang, Q., Chen, Z.: Latent friend mining from blog data. In: ICDM. pp. 552–561. IEEE Computer Society (2006)
49. Smith, B., Linden, G.: Two decades of recommender systems at amazon. com. *Ieee internet computing* 21(3), 12–18 (2017)
50. Song, D., Meyer, D.A.: Link sign prediction and ranking in signed directed social networks. *Social Netw. Analys. Mining* 5(1), 52:1–52:14 (2015)
51. Song, W., Xiao, Z., Wang, Y., Charlin, L., Zhang, M., Tang, J.: Session-based social recommendation via dynamic graph attention networks. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. pp. 555–563 (2019)
52. Tai, K.S., Socher, R., Manning, C.D.: Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint arXiv:1503.00075 (2015)
53. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: LINE: large-scale information network embedding. In: WWW. pp. 1067–1077. ACM (2015)
54. Tang, J., Chang, Y., Aggarwal, C., Liu, H.: A survey of signed network mining in social media. *ACM Comput. Surv.* (2016)
55. Tong, H., Faloutsos, C., Pan, J.Y.: Fast random walk with restart and its applications. In: ICDM. pp. 613–622. IEEE (2006)
56. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
57. Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabási, A.: Human mobility, social ties, and link prediction. In: KDD. pp. 1100–1108. ACM (2011)
58. Wang, X., He, X., Cao, Y., Liu, M., Chua, T.S.: Kgat: Knowledge graph attention network for recommendation. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 950–958 (2019)
59. Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., Yu, P.S.: Heterogeneous graph attention network. In: The World Wide Web Conference. pp. 2022–2032 (2019)
60. Xu, Y., Rockmore, D.N.: Feature selection for link prediction. In: PIKM. pp. 25–32. ACM (2012)
61. Yang, Y., Chawla, N.V., Sun, Y., Han, J.: Predicting links in multi-relational and heterogeneous networks. In: ICDM. pp. 755–764. IEEE Computer Society (2012)
62. Yao, Y., Zhang, R., Yang, F., Tang, J., Yuan, Y., Hu, R.: Link prediction in complex networks based on the interactions among paths. *Physica A: Statistical Mechanics and its Applications* 510, 52–67 (2018)
63. Yuan, G., Murukannaiah, P.K., Zhang, Z., Singh, M.P.: Exploiting sentiment homophily for link prediction. In: RecSys. pp. 17–24. ACM (2014)
64. Yuan, W., He, K., Guan, D., Zhou, L., Li, C.: Graph kernel based link prediction for signed social networks. *Information Fusion* 46, 1–10 (2019)
65. Yuan, W., Li, C., Han, G., Guan, D., Zhou, L., He, K.: Negative sign prediction for signed social networks. *Future Generation Computer Systems* 93, 962–970 (2019)
66. Yuan, W., Pang, J., Guan, D., Tian, Y., Al-Dhelaan, A., Al-Dhelaan, M.: Sign prediction on unlabeled social networks using branch and bound optimized transfer learning. *Complexity* 2019, 4906903:1–4906903:11 (2019)
67. Zayats, V., Ostendorf, M.: Conversation modeling on reddit using a graph structured lstm. *Transactions of the Association for Computational Linguistics* 6, 121–132 (2018)
68. Zhang, J., Shi, X., Xie, J., Ma, H., King, I., Yeung, D.Y.: Gaan: Gated attention networks for learning on large and spatiotemporal graphs. arXiv preprint arXiv:1803.07294 (2018)
69. Zhou, T., Lu, L., Zhang, Y.C.: Predicting missing links via local information. *The European Physical Journal B* 71(4), 623–630 (2009).

# The origins of severe software defects method

Nico Hillah

Institute of Certified Specialists, Perm, Russia

[https://doi.org/10.33847/2686-8296.2.2\\_3](https://doi.org/10.33847/2686-8296.2.2_3)

Received 01.10.2020/Revised 10.11.2020/Accepted 21.12.2020/Published 29.12.2020

**Abstract.** Identifying the causes which may potentially generate high financial damage was the goal of our research. To reach this goal, we conducted a case study on a system in the field of education. We studied the software defects of this system over several months and classified them based on two classification concepts: the EVOLIS and their severity. These defects prevent any essential operation or activity to be conducted through the concerned system or other systems connected to it. In fact, the occurrence of these failures causes a double financial cost to organizations: one in fixing them and the other one because of the unavailability of the system or systems. We targeted three types of software defects as sources of these failures. We conducted this study by classifying 665 software defects of a school management system and we found that the top two trigger groups are the technology and the IS architecture groups. This result led us to propose a method to identify the origins of severe software defects.

**Keywords:** Severe software defect, Software defect triggers, Software defect classification.

## 1. Introduction

Hackers take advantage of existing information system deficiencies, in particular using failures in software applications to cause damage not only to systems; but also, to institutions and users. With the event of Covid-19 crisis, every organization uses software systems to perform all types of activities in different domains. Unfortunately, most of the time, they are subject to failures [1], [2]. Software failures are defined as *“Termination of the ability of a product to perform a required function or its inability to perform within previously specified limits”* [p5, 3].

In fact, depending on their severity, these failures induce not only financial loss to organizations but also time and resource loss in correcting them. Software defects (SDs) are the sources of these failures. IEEE standard 1044-2009 [p5, 3] defines a defect as: *“An imperfection or deficiency in a work product where that work product does not meet its requirements or specifications and needs to be either repaired or replaced”*. Different studies have investigated the sources and factors triggering SDs [4], [5]. Nevertheless, there is no existing literature on the types of triggers associated with the level of SDs severity that they generate. Knowing which types of triggers generate which level of SDs severity will help systems administrators in particular and organizations, in general, to better allocate their resources in order to address software failures. In this regard, the question we address in this paper is *which types of trigger factors generate the most severe SDs and how to identify them?*

To answer this question, we conducted a case study on a software system. In fact, we performed two main classifications of its SDs. The first classification was to identify the severity of SDs; then we classified the same SDs based on the trigger factors using EVOLIS framework [4].

The paper is structured as follows: first, we will introduce existing techniques software defects classification; secondly, we will present the methodology that we

used in conducting this study; then, we will show the results that we obtained; and finally, we will present our main contribution which is the origin of severe SDs method.

## 2. Related Works

In the software life cycle, the classification of defects presents many advantages [6]. The classification of defects helps the software development teams to reduce the cost of correcting them, to detect defective modules, and to have efficient resource planning. Various studies have proposed and evaluated different approaches to collect and to analyze these SDs. The main approaches are (1) taxonomies [7], [8], root cause analysis [9], schemes [3], and the classification of these SDs [10].

There are different existing schemes in classifying SDs [10]. (1) The Orthogonal Defect Classification (ODC) of IBM [11] was developed in 1992 by R. Chillarege et al. [11] and it classifies defects across "*the dimensions (1) defect type, (2) source, (3) impact, (4) trigger, (5) phase found, and (6) severity*" [12]. (2) The HP Defect Origins, Types and Modes, the approach of Hewlett Packard, was developed by the HP software metrics in 1986 [13], and this scheme classifies the defects according to their types, their origins and their mode [12]. (3) The IEEE standard 1044-2009 is the scheme we retain for our first classification project. We selected this approach because it proposes the most complete definition of the SDs severity types among the three schemes. Moreover, this severity attribute is one of the most used attributes in SDs classification in practice [12]. The main advantage of choosing the severity attribute is the possibility for managers to identify which defect to correct first [12]. We retain the severity attribute for our first classification. Other defect classification studies have performed SDs classification using classical data mining techniques and algorithms such as association rules [14], Naïve Bayes Model [15], and clustering analysis [16].

## 3. Methodology

To be able to understand the relation existing between triggering factors for SDs and the severity of SDs, we conducted a case study of a system that we will name system A. This system is developed by an educational organization and it is a school management system. Its purpose is to help schools in managing the grades of their students. It is used for managing more than 90000 students' grades. The first version of system A had been released in the middle of 2012. We classified 665 SDs of system A. The collection of SDs covers a period of one year and four months from January 2015 to April 2016. System A has nine released versions over this period. The defects repository tool used by this organization is Jira [17].

Our objective is to classify SDs according to their severity and then classify these same SDs according to the factors that trigger them. In fact, we analyzed the SDs of this system A by classifying them according to the defect severity attribute of IEEE 1044-2009 standards [3] and then by classifying them with EVOLIS framework [4]. The software team in charge of maintenance of system A and a member of our research team had conducted both classifications.

### 3.1. The classification of SDs based on the severity

The IEEE standard defines this attribute as "*The highest failure impact that the defect could (or did) cause, as determined by (from the perspective of) the organization responsible for software engineering.*" [3]. The five values of severity are classified from the most significant to the least significant ones (see Table 1).

Table 1. Severity values [3]

Attribute	Value	Definition
Severity	Blocking (B)	Testing is inhibited, or suspended pending, correction or identification of suitable workaround.
	Critical (C)	Essential operations are unavoidably disrupted, safety is jeopardized, and security is compromised.
	Major (Maj)	Essential operations are affected but can proceed.
	Minor (Min)	Nonessential operations are disrupted.
	Inconsequential (Inc)	No significant impact on operations.

For the purpose of this study, we define any software defect (SD) as severe as long as it belongs to one of these severity levels: (1) Blocking (B), (2) Critical (C), and (3) Major (Maj). The Minor (Min) and the Inconsequential (Inc) are not considered as severe SDs. Our first classification is done based on the severity attribute of the IEEE standard (Table 2).

Table 2. Classification of System A's SDs based on their severity

	B	C	Maj	Min	Inc	Total
Jan	10	12	46	22	0	90
Feb	5	8	24	17	0	54
March	12	17	49	31	0	109
April	9	8	50	16	2	85
May	3	1	20	5	0	29
June	0	2	25	14	0	41
July	0	1	5	4	0	10
Aug	2	7	8	7	3	27
Sept	5	11	18	15	4	53
Oct	7	4	22	6	0	39
Nov	15	8	37	20	3	83
Dec	7	5	18	12	3	45
<b>Total</b>	75	84	322	169	15	665

### 3.2. The classification of SDs based on the EVOLIS Framework

For our second classification project, we chose the EVOLIS framework [4] (see Table 3.). This framework proposes a technique to classify SDs according to the factors that trigger them. "EVOLIS can be caused by a large variety of factors: bugs that need to be fixed, users that wish to have new functionalities, new market opportunities that require new software features, performance standards that the system must reach, technical changes in the environment with which the system must interact, obsolescence of applications and so on" [3]. EVOLIS identifies four main groups of factors that trigger SDs: (1) IS/users fit (U.F) triggers that are defined as any failure related to the user interface, the user documentation and aptitude to use the system. (2) The technology (TCH) triggers are related to defects that concern the software as well as the hardware platforms as information system components. (3) According to the authors, the IS architecture (ACH) triggers concern "different types of integration evolution, namely an evolution of integration among components of the system, among business functionalities, or an integration with systems outside of the company." [3]; and finally (4) the Business-IS (Bs-IS) alignment triggers that "address the co-alignment between business and information systems" [3].

Table 3. Classification of System A's SDs based on their trigger factors

	ACH	Bs-IS	TCH	U.F	Total
Jan	13	23	21	33	90
Feb	13	9	18	14	54
March	15	26	25	43	109
April	27	26	13	19	85
May	4	8	10	7	29
June	9	2	19	11	41
July	1	1	2	6	10
Aug	1	4	9	13	27
Sept	15	8	8	22	53
Oct	16	4	9	10	39
Nov	23	19	18	23	83
Dec	5	10	12	18	45
<b>Total</b>	142	140	164	219	665

Subsequently, we grouped these results into a two-dimensional table (see Table 4.). Each dimension represents the results obtained for each classification project.

Table 4. Two-dimensional classification of SDs of system A

	B	C	Maj	Min	Inc	Total
ACH	17	12	99	14	0	142
Bs-IS	24	21	66	28	1	140
TCH	19	30	87	27	1	164
U.F	15	21	70	100	13	219
<b>Total</b>	75	84	322	169	15	665

## 4. Discussion and Contribution

### 4.1. Discussion

We analyzed the results threefold: the results of severity classification, followed by the results of EVOLIS and then we combined and analyzed both results together. First, the severity classification showed us that the top three high types of SDs are respectively the major type of SDs with 322 SDs, followed by the minor type with 169 and the critical type with 84 SDs (see Table 2.). Second, the EVOLIS classification showed us that the top three groups of factors that trigger SDs are respectively the IS/users factors with 219, followed by the technology factors with 164 and then the factors related to the IS architecture and business-IS alignment (see Table. 3). These last trigger groups have almost the same number of SDs: 142 SDs for the IS architecture and 140 SDs for the business-IS alignment SDs. Further, the analysis of both combined results showed us that technology triggers represent respectively 12%, 18% and 53% for blocking SDs, critical SDs and major SDs (see Fig. 1). In total, the technology triggers are responsible for 83% of the severe SDs. Similarly, the architecture triggers represent respectively 12%, 8% and 70% for blocking SDs, critical SDs, and major SDs (see Fig. 1). The business-IS alignment represents respectively 17%, 15%, and 47% for blocking SDs, critical SDs and major SDs. Finally, the IS/users triggers represent 20% of the total of severe SDs (see Fig. 1).

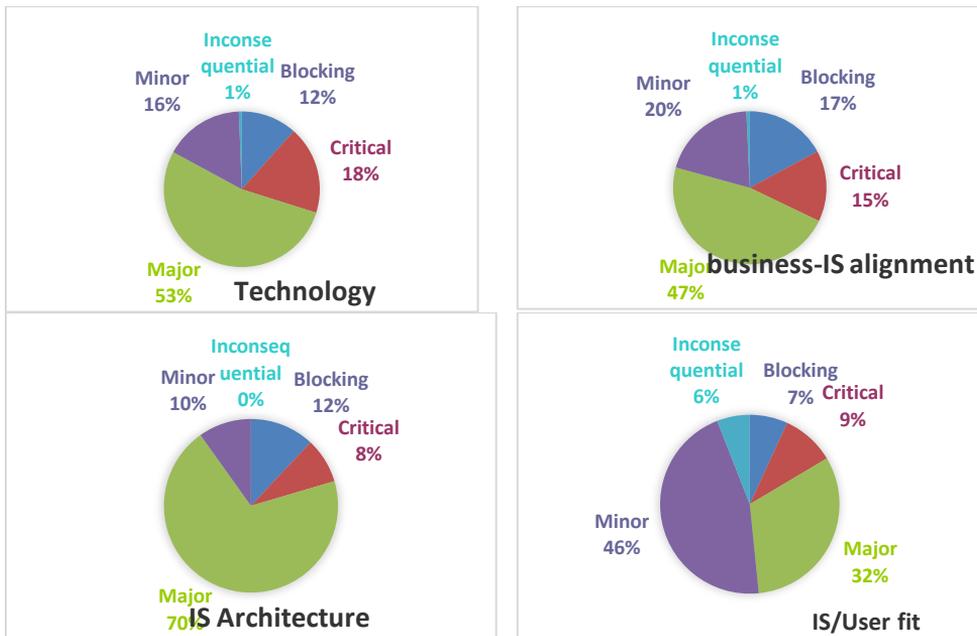


Fig. 1. Trigger factors and severity of SDs of System A

Analyzing these results separately did not give so much information to organize the SDs management. However, when we put them together, we found that some type of SDs trigger factors is the source of some specific severe groups of SDs. In fact, we observed that the majority of the inconsequential SDs are triggered by the IS/users fit factors.

This implies that the probability of an inconsequential SDs to be triggered by either the IS architecture, the business-IS alignment or the technology factors are very low or barely existent. Furthermore, we assigned weighting factors to each severity level according to their importance as followed (Table 5.).

Table 5. The weighting factors for the severity level.

Severity level	Weighting factor
Blocking	40%
Critical	30%
Major	20%
Minor	8%
Inconsequential	2%

We then apply this weighted scoring model to our two-dimension table to calculate the weighted scores (W) for severe SDs per trigger factor (see Table 6.).

Table 6. Severe SDs weighted score.

Severe SD	Weight (W)	Bs-IS	W- Bs-IS	ACH	W- ARC	TCH	W- TCH	U.F	W- U.F
Blocking	0.4	24	9.6	17	6.8	19	7.6	15	6
Critical	0.3	21	6.3	12	3.6	30	9	21	6.3
Major	0.2	66	13.2	99	19.8	87	17.4	70	14
Total	0.9	111	<b>29.1</b>	128	<b>30.2</b>	136	<b>34</b>	106	<b>26.3</b>

Looking at these results, we can conclude that the technology trigger factors, with the highest weighted score of 34, are responsible for most of the severe SDs followed by the IS architecture factors, with a 30 weighted score. Then the business-IS alignment, with 29.1, and finally the IS/users fit triggers, with 26.3 (see Fig. 2.). We can also notice that there is a considerable gap between the number of SDs of the first two groups of triggers (IS architecture and technology and the last two of them (business-IS alignment and IS/User fit).

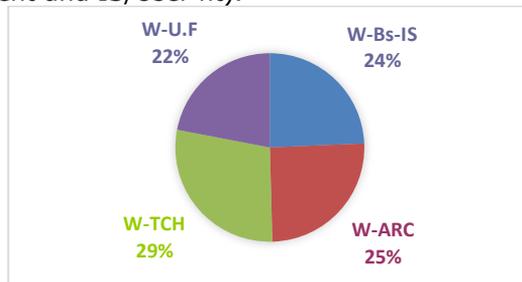


Fig. 2. Trigger factors by weighted severity of system A

#### 4.2. Contribution

Based on the results of these classifications, we propose a method in order to identify the origins of severe SDs. This method is addressed to the people responsible for managing and controlling software evolution such as software portfolio managers, software maintenance teams, software development teams, and to researchers studying software evolution, and conducting studies in the SDs and change requests mining field. This method empowers these stakeholders to identify the origins or sources of severe SDs on applications. We named this method “The origins of severe software defects method”.

This method consists of three stages: SDs collection or acquisition stage, SDs analysis stage, and the SDs classification stage. We define each of these stages as follows:

- 1<sup>st</sup> stage: The collection stage consists of collecting SDs into a software repository.
- 2<sup>nd</sup> stage: The analysis stage consists of identifying the trigger factors (origins) of the SDs and evaluating their severity impact on the system. We named a SD having its origin and severity impact identified “Analyzed SD” (ASD).
- 3<sup>rd</sup> stage: The classification stage consists of grouping the analyzed SDs (ASDs) at the preceding level into two groups: the group of severe SDs and the group of nonsevere SDs. Any ASD that causes a partial or total disruption of the system’s essential operations falls under the second group. The severity seriousness of the SDs is defined by the person responsible for software engineering in this context. Any other ASD having no impact on the system’s essential operations must be categorized into the group of nonsevere SDs. Performing this classification will lead to identify the origins of severe SDs of the system being studied.

We represent this method in a pyramid form. The first stage is the initial phase, we place it at the base of the pyramid. It is followed by the second stage in the middle of the pyramid and then the third stage at the top of the pyramid (see Fig. 3). In summary, the activities at the three stages are presented as follows:

At the first stage (the collection stage), the activity to perform is:

- The collection of SDs.

At the second stage (the analysis stage), the activities to perform are:

- For each SD, identifying its trigger factor or origin, and identifying the severity of its impact on the system.

At the third stage (the classification stage) the activity to perform is:

- Classification of the ASDs based on their severity impact into severe and no severe groups.

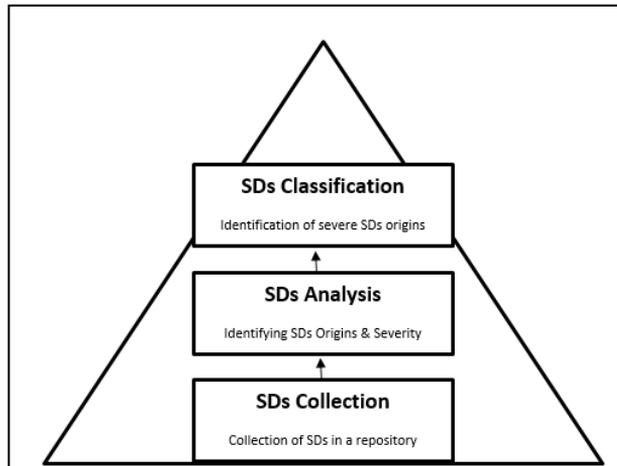


Fig. 3. The origins of severe software defects method

## 5. Conclusion

Not only do SDs cause huge financial loss to system owners, but their management also causes up to 80% of the total cost of a system during its life cycle [6]. In this context, identifying the ones which may potentially generate high financial damage was the goal of our research. To reach this goal, we conducted a case study on a system in the field of education. We studied the SDs of this system over several months. We classified them based on two SD classification concepts: the EVOLIS [4] and their severity [3]. To the question of which groups of SDs triggers generate the most severe SDs, we answered that the technology triggers are at the head position with a total of 34 weighted score. In the second position is the IS architecture triggers which come with a 30.2 weighted score, and then followed by the business-IS alignment triggers with 29.1. The last position is occupied by the IS/users fit triggers with a 26.3 weighted score of the total severe SDs analyzed. We also found that the majority of the defects triggered by IS/user factors are either minor or inconsequential types of SDs.

The results obtained from this study will help software managers to improve the management of SDs by allocating the SDs correction resources more accurately thus reduce the cost of managing SDs. In our future work, we will analyze other software systems and compare their results to the ones we obtained in this study. From a practical perspective, a possible future work will be to build an automatized SD management system (SDM) based on our proposed method. Just like the method, this SDM system will inform in real-time on the real impact of SDs and their sources. This system will help software maintenance teams and software portfolio managers in managing their system's SDs by deploying the right resource at the right moment on the right instance. It also allows them to have an overview of the actual state of their systems.

## References

1. R. N. Charette, "Why software fails [software failure]," *Ieee Spectr.*, vol. 42.9, pp. 42-49, 2005.
2. R. Kaur and D. J. Sengupta, "Software Process Models and Analysis on Failure of Software Development Projects," *Int. J. Sci. Eng. Res.*, vol. 2, no. 2, p. 4.

3. 1044-2009 IEEE Standard Classification for Software Anomalies. 2009.
4. A. Métrailler and T. Estier, "EVOLIS Framework: A Method to Study Information Systems Evolution Records," in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, 2014, pp. 3798–3807.
5. A. A. Alshazly, A. M. Elfatraty, and M. S. Abougabal, "Detecting defects in software requirements specification," *Alex. Eng. J.*, vol. 53, no. 3, pp. 513–527, Sep. 2014.
6. G. K. Rajbahadur, S. Wang, Y. Kamei, and A. E. Hassan, "The impact of using regression models to build defect classifiers," in *Proceedings of the 14th International Conference on Mining Software Repositories*, 2017, pp. 135–145.
7. R. Binder, *Testing object-oriented systems: models, patterns, and tools*. Reading, Mass: Addison-Wesley, 2000.
8. D. Vallespir, F. Grazioli, and J. Herbert, "A framework to evaluate defect taxonomies," in *XV Congreso Argentino de Ciencias de La Computación*, 2009.
9. M. Leszak, P. Dewayne E., and D. Stoll, "Classification and evaluation of defects in a project retrospective," *Elsevier*, no. 61, pp. 173–187, 2002.
10. N. Mellegård, *Improving Defect Management in Automotive Software Development, LiDeC—A Light-weight Defect Classification Scheme*. Chalmers University of Technology, 2013.
11. R. Chillarege et al., "Orthogonal defect classification—a concept for in-process measurements," *IEEE Trans. Softw. Eng.*, vol. 18, no. 11, pp. 943–956, 1992.
12. S. Wagner, "Defect classification and defect types revisited," in *Proceedings of the 2008 workshop on Defects in large software systems*, 2008, pp. 39–40.
13. J. T. Huber, "A Comparison of IBM's Orthogonal Defect Classification to Hewlett Packard's Defect Origins, Types, and Modes." Hewlett Packard Company, 1999.
14. L. Yu and S. R. Schach, "Applying association mining to change propagation," *International Journal of Software Engineering and Knowledge Engineering*, vol. 18, no. 08, pp. 1043–1061, 2008.
15. G. Murphy and D. Cubranic, "Automatic bug triage using text categorization," in *Proceedings of the Sixteenth International Conference on Software Engineering & Knowledge Engineering*, 2004.
16. W. Dickinson, D. Leon, and A. Fodgurski, "Finding failures by cluster analysis of execution profiles," in *Proceedings of the 23rd International Conference on Software Engineering. ICSE 2001*, Toronto, Ont., Canada, 2001, pp. 339–348.
17. Atlassian, "Jira | Logiciel de suivi des tickets et des projets," *Atlassian*. [Online]. Available: <https://fr.atlassian.com/software/jira>. [Accessed: 06-Apr-2020].

# The development of a classification model of accounting in the digital economy

Aleksei Kovalev<sup>[0000-0002-3277-5586]</sup>

<sup>1</sup> Novosibirsk State University of Economics and Management, 630099 Novosibirsk, Russia

[https://doi.org/10.33847/2686-8296.2.2\\_4](https://doi.org/10.33847/2686-8296.2.2_4)

Received 18.05.2020/Revised 07.11.2020/Accepted 21.12.2020/Published 29.12.2020

**Abstract.** A multidimensional classification scheme and a semantic multidimensional accounting data model are defined in this article. Instead of accounts, multidimensional accounting uses categories of economic activity. The proposed multidimensional data model is more flexible than the traditional account model and allows you to expand the capabilities of accounting, taking into account the different needs of users of accounting information. The multidimensional data model allows you to expand the capabilities of accounting, taking into account the different needs of users of accounting information. To create a multidimensional accounting system, the categories of economic activity registered in accounting have been determined, the concept of double entry and balance in a multidimensional representation (probalance) has been formulated. The features of planning in a multidimensional accounting system have been described and the implementation of the financial results plan has been considered.

**Keywords:** Bookkeeping, Accounting, Semantic data model of accounting, Multidimensional accounting, Multidimensional data model, Hierarchical classification scheme, Categories of economic activity, Commercial activity, multidimensional classification scheme, OLAP, Internal area of accounting, External area of accounting, Pro-balance, Digital economy.

## 1. Introduction

In recent decades, corporate financial information, conveyed by complex quarterly and annual reports, has lost most of its usefulness to many users, also to investors, and is urgently in need of revitalization and restructuring. Given the crucial role of financial (accounting) information in fostering prosperity and growth of business enterprises and the economy at large, the serious deficiencies of this information, should be of great concern not only to investors, but also to managers and accountants [1]. AQ2

Humankind uses accounting since 12–13 century. The basis of accounting was the work with information technology based on the use of paper. Accounting as an information system has not changed significantly since that time. Modern information space formed by complex technologies, hardware, and software opens up new possibilities for the development of accounting and presents it to the new requirements.

We endure an epoch of formation of a digital economy. Developed not only new equipment and new principles of information processing.

There is a growing value of information. The information itself is transformed into a new quality, more suitable for decision-making. This process is depicted as DIKW Pyramid. Standard DIKW Pyramid and consists of four levels: (D) data, (I) information, (K) knowledge, and (W) wisdom [2]. An extended version of DIKW

Pyramid consists of 6 levels (see Fig. 1) measurement; facts; data, information; knowledge; and wisdom [3].

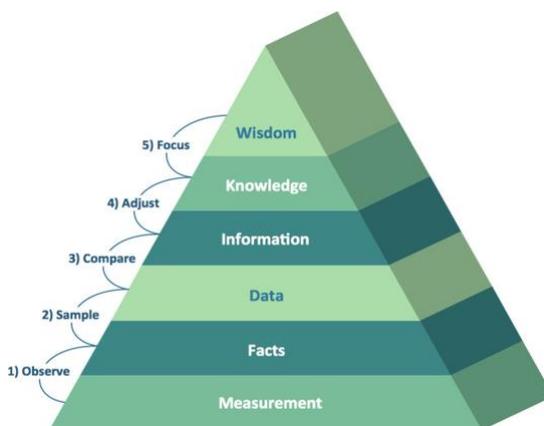


Fig. 1. Extended version of Data Information Knowledge Wisdom pyramid.

Accounting as information system cannot stand away from this process. But developing of accounting lags behind the growing demands of users of accounting information. The accounting system stay the level of “Data” and does not increase above. Accounting is the source of data for the level of “Information”. At the level of “Information” are data warehouse. At the level of «Knowledge» of the class Business Intelligence (BI) or Business Analysis (BA) for example, online analytical processing (OLAP) [4].

In accordance with the definition given by the American Association of Accountants’ Accounting - the process of identifying information, calculation and evaluation of performance and reporting of information to users to create, study and decision-making “[5, c. 10]. The subject of accounting are the facts of economic activity, which in the accounting process are the identification, classification and registration [6]. There are three basic types of classifications: hierarchical, faceted and multidimensional.

It should be noted that the terminology of classification is developing. Some terms have definitions that differ from one author to another. The facet classification was developed by S. Ranganathan and was first named by him “Colon Classification”. Facet classification generates a code that consists of a predefined set of signs arranged in a specific order. Ranganathan shared the facets in this code, separated by a symbol ":" (colon). Over time, it became known as the “facet classification.” The facet classification was further developed, which led to the development of a multidimensional classification. Some authors continue to use the term “facet classification” in relation to a multidimensional data structure.

Hierarchical classification is very useful in terms of navigation and paper technology for information processing. But at the same time, a hierarchical classification has a rigid structure and changes in the structure entail incomparability of data and require time-consuming. At the same division level, but in different branches of hierarchical tree different division bases are allowed. Hierarchical charts of accounts are often found.

Facet classification is more perfect. In this classification, the classification signs are arranged strictly in order and a specific sign is located in a certain place. This structure allows you to process each attribute separately or in any order when generating reports. Facet classification retains the disadvantages of hierarchical

classification: a strict order of signs, a certain number of signs that cannot be changed.

Multidimensional classification is most universal of three kinds of classifications. Multidimensional classification supposes addition of new signs of classification, change of interrelation of elements. But thus, work effective work with it is possible only with computer facilities application. It is possible to construct any hierarchical classifications of multidimensional classification within considered signs. In this sense multidimensional classification is base in relation to hierarchical classification and gives to the user the maximum freedom in work with the information.

In the multidimensional classification of signs does not stand out major and minor and the order of signs is not defined in advance.

It provides possibility of development of system without reorganization of all tree of signs (in case of hierarchical classification), and also to build the reports containing any possible variants of hierarchical classifications within the limits of considered signs.

## **2. Previous Study**

The important role of classification is mentioned in a large number of publications. References to scientific sources are given as an example of such works. In these works, the classification is considered in a variety of applications: classification of balance sheet items [6, 7], classification of articles of the statement of financial results [8], classification of accounts [9], Classification of assets and liabilities [7, 10, 11] accounting system classification [12, 13] Classification of the cash adjustments [11] Accounts referred to any element of the classification or assigns a classification function [14]. Schmalenbach developed a structured chart of accounts based on facet classification [15]. The categorical structure of accounting information was developed by W. McCarthy. He used three categories: Resources events agents (REA) [16]. The General role of classification in accounting is described in book by Van Der Poll. [6]

The closest to the topic of classification as the basis for the formation of accounting information came Professor Shaposhnikov [17], who considered the use of classification in various tasks of accounting but his research has not received further development. Some key features of multidimensional classification are described in the article "Multidimensional Model of Accounting" [18]

The study of classification as the basis for the creation of accounting information is considered for the first time.

## **3. The Main Body**

### **3.1 Methods**

As research methods were used multidimensional classification and OLAP. Now the business accounting is based on use of accounts and sub-accounts. This system corresponds to the hierarchical model, in which the first level of division in which are the business accounting accounts, the second a sub-account. Structure accounts and subaccounts it is generated on the basis of a feature set reflecting interests of dominating group of users. Interests of groups of users in relation to the organization considerably differ. Occurrence of new requirements of users leads to necessity of labour-intensive process of reorganization of hierarchy or creation and simultaneous use of alternative charts of accounts. This fact is one of the reasons of process of forming of various kinds of accounting observed now: financial accounting, tax accounting, management accounting, etc. Multidimensional classification allows you

to take into account in a single system different points of view on information. As a result of the lack of system limitations, not one of these points of view should be recognized as the main one. Therefore, the most suitable to the business accounting organization is multidimensional classification which in computer science concerns multidimensional model of the data. Multidimensional model of the data used in technology OLAP (OnLine Analytical Processing - operative analytical handling). Basic purpose OLAP - handling of the information for analysis and decision making carrying out. OLAP refers to the level «Knowledge» in the system DIKW Pyramid.

In relation to other conceptual models of the data, OLAP has a number of advantages:

1) It is easily perceived by end users (not specialists in the field of information technology), occurrence of new aspects of use of the data and necessity of entering of new communications does not lead to re-structuring of all model of the data and a database as a whole [19].

2) In OLAP tools of the analysis of the data are brightly expressed. She allows to receive an intellectual estimation, i.e. to make generalization, grouping, removal of the excessive data and to raise reliability at the expense of an exception of errors. Advantage OLAP consists that inquiries here are formed not on the basis of rigidly set forms, and by means of the flexible independent approaches which are not demanding for modification attraction of the programmer [20]. Groupings can be built, freely choosing a detailed elaboration order in a cut of signs. OLAP provides revealing of associations, laws, trends, carrying out of multiple classifications, generalization or detailed elaborations, drawing up of forecasts, carrying out of the intellectual analysis («Data mining») i.e. are given by the tool for enterprise management in real time. Thus, execution of the widest scope of inquiries of users to information system of accounting is provided.

3) The Multidimensional model of the registration information supposes with the minimum costs expansion and increase in detailed elaboration of the data according to inquiries of users of the registration information. In multidimensional model it is possible to add new analytical attributes (to increase detail of the data) and numerical indicators without any changes in existing forms of input and reports [21; 129].

The classifier of a multidimensional classification consists of a of signs and values of these signs.

The multidimensional model allows to register signs in any combination and to receive any slices of the information within the limits of the considered signs, as reflects graphical representation of a 3-dimensional hypercube.

Now there are highly effective software of handling OLAP of cubes: Targit, Tableau, Yellowfin, WebFOCUS etc. One of the most accessible and intuitively clear means OLAP, is Pivot table in MS Excel.

### **3.2. Conceptual Bases of Multidimensional Accounting**

We call accounting, which is built on a multidimensional data model, multidimensional accounting. Multidimensional structure and the corresponding hierarchical structure relate as a general to specific. Thus, and multidimensional accounting concerns the business accounting as a general to specific.

At a design stage of multidimensional model, it is necessary to comprehend domain knowledge of accounting. Features of the structure of the resulting model should not violate the rules of formal logic. Logically consistent structure of the model provides the benefits that arise when working with data. Disadvantages of the conceptual domain data model cause of errors and contradictions at work with accounting data. In this aspect, the multidimensional model is means of check of correctness of understanding of base concepts and the technician of the descriptions of the considered essence.

Therefore, development of a technique of accounting of the facts of economic activities in multidimensional model has demanded the critical analysis of bases of an accounting method.

Such bases consist of: a subject of accounting, the accounting account double record and balance, a ratio of analytical and synthetic accounting.

1. For a basis the point of view is taken that an accounting subject are the facts of economic activities affecting, directly or indirectly to the financial position the organization (capital). It is the point of view has been stated by Schar [22] and it is supported by modern writers, such as Sokolov [23], etc.

2. Multidimensional accounting uses some the basic, most general concept base primitive things (or categories in philosophical understanding of the term) instead of business Accounts. Used categories are the most general concepts inherent in all economic operations. The semantic model of accounting consists of the three main categories: the object of economic operation, the entity of economic operation, the type of relation between object and economic entity. As usual, the relation is the relation of property between object and economic entity.

<b>The object</b>	<b>The relation</b>	<b>The economic entity</b>
-----------------------	-------------------------	--------------------------------

For the basic accounting model, it is categories: the object of economic operation; the economic entity; the type of relation between object and economic entity; circumstances of the event, time of event and a measure. All main properties of economic operation are characteristics of the specified categories.

In every economic relation of property, there are two sides: the subject (owner) and the object (property). In the work, the following content is devoted to them: the Subject is an active motivated beginning that expresses its will and manifests energy in achieving its goals. The object is a passive beginning, the object of relations (goods that have value for the subject).

The relation details the type of economic and legal relations between an object and a subject. By default, this relationship is ownership. However, certain elements of ownership can also be specified: the right to use, own or dispose of, etc. It is also intended to indicate the "control" relationship provided for in the conceptual framework of IFRS.

There are examples of the similar approach at the western accounting school. It is the system named REA (from the name of primitive things - resources, events, agents). For the first time concept REA has been entered by McCarthy in 1982 [16]. REA has a number of differences from the multidimensional accounting, most essential of which is absence of double-entry and balance sheet model in an explicit form. Even earlier, in 1975 role Michael A. Crew underlined a role of agents and resources in the description of the facts of economic activities. In his book "Theory of the firm," he wrote, that the facts of economic activity are combined into a few basic types of:

- Transfer a resource or money from one agent to another;
- Hand over the information from one agent to another;
- Transform a number of resources to other kind of resources.

Generally active business accounts concern the characteristic of objects of operations, and passive and is active-passive accounts concern the characteristic of entity of operations.

3. For multidimensional accounting the scheme of warehouse of the data "Star" most approaches. It provides storage of essential properties of economic operations concerning accounting categories (see Fig. 2):

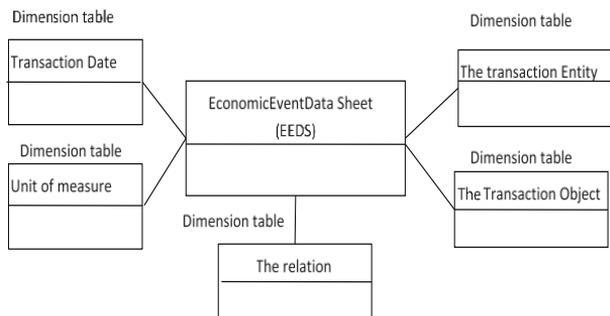


Fig. 2. The base scheme of a data structure in multidimensional accounting.

In solving practical problems of accounting schema “snowflake” replaces the schema “star”, which admits the existence of hierarchical classifications in the dimension tables. Given the variability of analytical signs for analytical dimensions must be applied temporal databases. Base categories allow you to identify a specific event. Each such event will be described by a large number of analytical signs. These analytical signs characterize the values of each of the categories involved in the event: a specific object, a specific subject, and the event as a whole. Each category can be described in detail by a large number of analytical signs. The number of attributes and their values may change.

This system allows very detailed and flexible characterization of each event in the accounting system. Changing the analytical attributes will not compromise the integrity of the credential data, since the structure of registered events will be preserved.

4. In accounting double entry uses a different combination of accounts. Application of categories in multidimensional accounting has provided a constancy of a design of double-entry. In multidimensional accounting double-entry contains constant quantity of elements for registration of all facts of economic activities. In a multidimensional accounting the double entry records changes in each category or sign. Changes between categories (or sign) are impossible. Equality of the sums on the number module in double-entry of multidimensional accounting provides relationship of cause and effect and possibility of reconstruction of events on the basis of the registration information. Double-entry defines the end of the previous state (quality) and the beginning of a new state (quality) of the object and the subject of accounting. Thus, the continuity is preserved signs and recording of change in their values (see Fig. 3). Double-entry is expressed in the completion of one situation and the appearance of new situation. This scheme is directly implemented in the double-entry of multidimensional accounting.

Instead of the debit and the credit of accounts in multidimensional accounting are used positive and negative numbers. Has offered and researched use possibilities in accounting of rational numbers of Tsygankov [24].

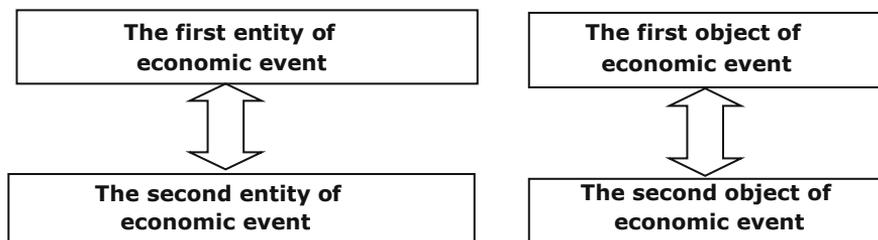


Fig. 3. Conceptual structure of double-entry in multidimensional accounting.

Positive and negative numbers with success are use in cameral accountings and single-entry accounting. At use of rational numbers, the basic scheme of registration of economic operations looks as follows (see Fig. 4):

Where: X - the sum of money and (or) in natural indicators. Values for the first and second line are equal modulo.

The first example concerns the events connected with change of the proprietor or moving of objects between divisions of one proprietor.

<b>1. Change of the economic entity.</b>					
<b>№</b>	<b>Moment in time</b>	<b>Subject of the event</b>	<b>Relation</b>	<b>Object of the event</b>	<b>Value</b>
1	YY.MM.DD	The economic entity <u>1</u>	Property	The Object 1	+ X1
1	YY.MM.DD	The economic entity <u>2</u>	Property	The Object 1	- X1

<b>2. Change of Object.</b>					
<b>№</b>	<b>Moment in time</b>	<b>Subject of the event</b>	<b>Relation</b>	<b>Object of the event</b>	<b>Value</b>
1	YY.MM.DD	The economic entity 1	Property	The Object <u>1</u>	+ X1
1	YY.MM.DD	The economic entity 1	Property	The Object <u>2</u>	- X1

<b>3. Mixed type of event.</b>					
<b>№</b>	<b>Moment in time</b>	<b>Subject of the event</b>	<b>Relation</b>	<b>Object of the event</b>	<b>Value</b>
1	YY.MM.DD	The economic entity <u>1</u>	Property	The Object <u>1</u>	+ X1
1	YY.MM.DD	The economic entity <u>2</u>	Property	The Object <u>2</u>	- X1

Fig. 4. The main scheme of registration of economic events in multidimensional accounting.

The second example describes the scheme of events when one object of operation replaces another, such as: transfer of materials to production, output finished product etc.

The third example reflects the mixed type of events when there is a change of the economic entity and object as a result of the economic events.

The completed state indicated by a negative value is called "outflow". The state that started is indicated by a positive value, called "inflow".

The table containing double-entry of registration of the facts of economic events is called «Economic event data sheet».

5. The Following task consisted in development of a basis of balance sheet in multidimensional system. The first problem was practice of allocation accounts receivable to an asset balance.

In the traditional balance sheet, a positive balance of settlements with counterparties is reflected in the asset balance and negative balance in the passive. This principle at the beginning does not allow to determine what part of the balance sheet will reflect the result of the relationship with the counterparty at the balance sheet date. I.e. it is situational, not systematic and does not allow to allocate a steady sign on which the balance sheet parties are formed.

In multidimensional accounting the balance sheet model is organized in another way. In our opinion, the balance sheet model is reflection of isolation of property of the organization from set of the property rights existing in economy.

Principle of isolation follows of the direct reading of the definition of a legal entity in the Civil Code Russian Federation, paragraph 48 [25], in which isolated property described as the main criterion for the existence of a legal entity. Separate

property is the economic basis of Organization on which behalf the accounting is. The other entities may become contractors of this Organization.

In multidimensional accounting the balance sheet is formed on the basis of values of a special sign: «accounting area». This sign is given for entities of operations. Generally, it has only two values: «Internal area» and «External area». Sign of “internal area” corresponds to the organization in terms of which is accounting. «The external area» is used for counterparts of this Organization.

We name the received design of the balance sheet “Pro-balance sheet”.

The isolated property relates to the asset of the balance. We call this asset of the balance sheet of “pro-activ”. The other side of the balance sheet is detailed balance of any mutual settlements with all counterparts and owners. We call this other side of the balance of “Pro-passive”. Feature of the pro-passive is that it reflects accounts receivable with the opposite sign. Pro-activ and pro-passive results are equal in absolute numbers. The pro-balance sheet circuit diagram is shown in a Fig. 5:

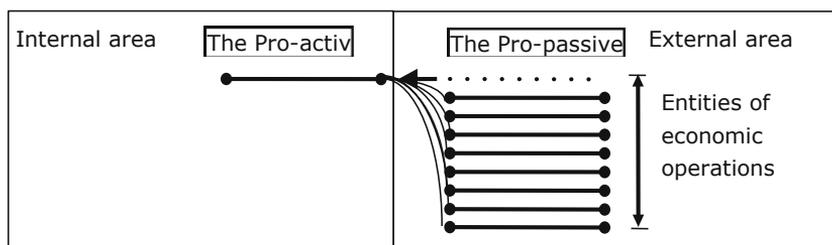


Fig. 5. Dichotomous division by values of accounting area.

The division of entities into two classes corresponds to the logical operations: the dichotomy. Using only a dichotomy, we have one article in the pro-active, and many in the pro-passive. Articles asset balance are formed by the decomposition of the analytical elements of the asset at an additional economic characteristic (see Fig. 6). In liabilities on the contrary, entities are arranged in groups on economic grounds.

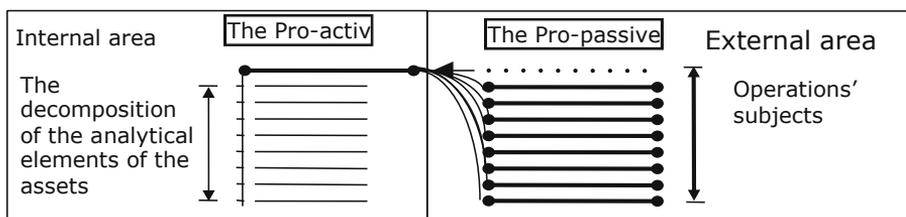


Fig. 6. Forming articles on pro-asset.

As a result of these operations, we have general balance sheet structure in the context of articles (see Fig. 7).

The Pro-active Sign		The Pro-passive Sign	
Fixed assets	+	Share capital	-
Intangible assets	+	The financial result (profit)	-
Current assets	+	Creditors	-
		Debtors	+
<hr/>		<hr/>	
Total the balance	+	Total the balance	-

Fig. 7. The overall structure of the balance-in the context of articles.

Share capital refers to the calculation of the founders. The financial result concerns the special participant of economic mutual relations - to the market. The market determines the margined income of commercial operations.

Feature of pro-balance is negative value of a result of a pro-passive. The debt receivable is reflected in a pro-passive with positive value. Total assets are equal to total liabilities in absolute numbers, respectively, when added up give a zero.

The multidimensional provides the ability to compare events of the reporting period with planned indicators. Comparison occurs at the reporting stage. The comparison method has several features. The first feature is that the multidimensional uses both positive and negative values of the measure. Thus, revenues will be positive, costs negative; accounts receivable will have a positive measure value, accounts payable will have a negative measure value. The same feature is also characteristic of the planning system in a multidimensional accounting model.

The second feature is that the results of the lines in OLAP can only be obtained as a result of data addition, but not subtraction. In order to calculate deviations from the planned values in the OLAP report, the plan data is inverted - replacing the sign of the number with the opposite. Positive numbers will be negative after that, and negative numbers will be positive. After that, adding the planned values to the actual ones in the OLAP reports gives a deviation from the planned values. It is worth noting that in such reports the under fulfillment of the revenue plan is shown with a negative value, and the savings in expenses with a positive one. Conversely, an over fulfillment of the plan for incomes is reflected in a positive value, and an excess of costs with a negative value (see Fig. 8). This result coincides with the characteristic of the results obtained in the verbal assessment of "bad - good" in most cases.

Cumulative indicators for the month and individual events can also be planned. Thus, it is possible to reflect obligations arising from the terms of concluded contracts in multidimensional accounting.

Planning can be carried out with varying degrees of detail in the context of classification signs. The level of detail depends on the management tasks and affects the detail of reports on plan completion.

We conducted a series of experiments to determine the possibilities of reporting based on multidimensional accounting data. As a result of these experiments, the possibility of informing interested users by providing them with specific data marts — aggregated accounting information data marts — was proved.

Line Names	Plan (inverted)	Fact	Difference
<b>Income</b>	<b>-115 779 661,02</b>	<b>103 200 000,00</b>	<b>-12 579 661,02</b>
Wholesale sales of light oil products	-106 779 661,02	96 000 000,00	-10 779 661,02
Storage services for oil products	-9 000 000,00	7 200 000,00	-1 800 000,00
<b>Cost of sales</b>	<b>104 322 033,90</b>	<b>-89 784 000,00</b>	<b>14 538 033,90</b>
Wholesale sales of light oil products	103 050 847,46	-83 520 000,00	19 530 847,46
Storage services for oil products	1 271 186,44	-6 264 000,00	-4 992 813,56
<b>Expenses (other)</b>	<b>3 195 391,04</b>	<b>-3 644 900,00</b>	<b>-449 508,96</b>
Auto services for the delivery of materials, fuels and lubricants, finished products	1 562 881,36	-2 100 000,00	-537 118,64
Lease of industrial buildings and structures, land	15 368,54	-34 000,00	-18 631,46
Payments in favor of employees	667 400,00	-665 000,00	2 400,00
Railway services for the delivery of materials, fuels and lubricants, finished products	500 000,00	-480 000,00	20 000,00
Land tax	90 512,00	-10 000,00	80 512,00
Mandatory accruals on payments in favor of employees	201 554,80	-188 000,00	13 554,80
Security services	60 450,00	-78 000,00	-17 550,00
Environmental pollution charges	9 822,60	-4 500,00	5 322,60
Landline phone	1 600,00	-1 000,00	600,00
Insurance services	27 317,00	-31 000,00	-3 683,00
Electricity	53 400,00	-53 400,00	0,00
Cellular communication	5 084,75		5 084,75
<b>Total result</b>	<b>-8 262 236,08</b>	<b>9 771 100,00</b>	<b>1 508 863,92</b>

Fig. 8. Report on the implementation of the plan of financial results.

A simplified example of planning report based on a cash flow statement in OLAP is presented in Fig. 9.

Row names	Plan (inverted)	Fact	Difference
<b>Balance Storage Initial</b>	<b>-300 000,00</b>	<b>339 346,94</b>	<b>39 346,94</b>
Opening balance	-300 000,00	339 346,94	39 346,94
<b>Sales</b>	<b>-116 930 000,00</b>	<b>109 405 317,11</b>	<b>-7 524 682,89</b>
Inflow	-116 930 000,00	109 405 317,11	-7 524 682,89
Outflow	0,00	-4 188 336,87	-4 188 336,87
<b>Purchase</b>	<b>90 000 000,00</b>	<b>-86 985 409,79</b>	<b>3 014 590,21</b>
Inflow	0,00	102 675,54	102 675,54
Outflow	90 000 000,00	-87 088 085,33	2 911 914,67
<b>Transfer between cash desk and bank</b>	<b>0,00</b>	<b>0,00</b>	<b>0,00</b>
Inflow	-47 000 000,00	45 956 599,00	-1 043 401,00
Outflow	47 000 000,00	-45 956 599,00	1 043 401,00
<b>Payments for taxes and fees</b>	<b>3 800 000,00</b>	<b>-3 829 601,50</b>	<b>-29 601,50</b>
Outflow	3 800 000,00	-3 829 601,50	-29 601,50
<b>Pension Insurance Calculations</b>	<b>1 415 000,00</b>	<b>-1 415 039,94</b>	<b>-39,94</b>
Outflow	1 415 000,00	-1 415 039,94	-39,94
<b>Calculations on social med. insurance</b>	<b>22 000,00</b>	<b>-21 937,22</b>	<b>62,78</b>
Inflow			0,00
Outflow	22 000,00	-21 937,22	62,78
<b>Pay to staff</b>	<b>15 400 000,00</b>	<b>-15 150 360,02</b>	<b>249 639,98</b>
Outflow	15 400 000,00	-15 150 360,02	249 639,98
<b>Financial operations</b>	<b>3 800 000,00</b>	<b>-14 163,81</b>	<b>3 785 836,19</b>
Inflow	-199 000 000,00	207 381 996,54	8 381 996,54
Outflow	202 800 000,00	-207 396 160,35	-4 596 160,35
<b>Settlements with founders</b>	<b>42 000,00</b>	<b>-41 716,50</b>	<b>283,50</b>
Inflow	-23 000 000,00	23 000 000,00	0,00
Outflow	23 042 000,00	-23 041 716,50	283,50
<b>Claims</b>	<b>0,00</b>	<b>-1 491 979,22</b>	<b>-1 491 979,22</b>
Inflow	0,00	900,00	900,00
Outflow	0,00	-1 492 879,22	-1 492 879,22
<b>Payment for settlement and cash services</b>	<b>164 000,00</b>	<b>-164 073,50</b>	<b>-73,50</b>
Outflow	164 000,00	-164 073,50	-73,50
<b>Total</b>	<b>-2 587 000,00</b>	<b>630 382,55</b>	<b>-1 956 617,45</b>

Fig. 9. Report on the implementation of the plan of cash flow.

The use of multidimensional accounting opens up broad prospects for the development of tools and forms for presenting reporting data to end users in the form of aggregated accounting data mart. Instead of reporting forms, users will receive databases with a multidimensional structure containing aggregated accounting information. The user can create standard reporting forms or any other forms at their own discretion. Aggregation (consolidation) of data will allow you to keep a commercial secret. The development of accounting reporting in the form of data mart is more promising than the now widespread XBRL (eng. eXtensible Business Reporting Language). The structure of the proposed categories, analytical features and their values in the multidimensional accounting reflects the content of economic processes without specific "accounting" terms, which increases the transparency of accounting, the clarity of accounting information by end users, and its usefulness in making informed management decisions.

Users of the of aggregated accounting data mart will have the opportunity to deeply comprehensively study the economic activities of the organization. They will be able to determine cause-effect relationships in his development and to develop more informed management decisions. This level of depth of work with financial information corresponds to the level of "knowledge" in the DIKW pyramid.

## **5. Conclusion**

Development of a technique of multidimensional accounting has allowed to look at conceptual bases and technology of the business accounting in a new fashion. Multidimensional accounting eliminates rupture between development of the business accounting and modern information technology, allows to use all advantages OLAP to handling of the considered data.

Multidimensional accounting based on the most general terms - a category, it uses a multidimensional classification instead of hierarchical. This suggests that the multidimensional accounting refers to the traditional accounting as a general to specific. Accordingly, the concept of balance (the pro-balance) and double-entry accounting disclosed in the multidimensional accounting, and is applicable to conventional accounting.

Multidimensional accounting improves the accounting systems to the level of "Information" in the DIKW.

We have identified the presence of the categories of entity and object in the semantic content of accounts. But it also found that the accounts contain signs related to the object and entity. It is these signs define the structure of accounts. Accounts contain the reference to base categories of accounting and value of the selected signs of these categories. The name of the accounts contains value of the selected signs related to the basic categories. But they have not reference to the appropriate category, though implies their participation. This phenomenon is analogous to what in linguistics is called the "hided verbal message". Consequently, the accounts of accounting are complicated in their content and unite categories and their attributes. This shows the hierarchical system of accounts as well as the hierarchical structure of sub-accounts. In our opinion, the hierarchical system of accounts defines the rigidity of the accounting system, and it is the cause of the constant struggle of different groups of users to reflect their information requests in the structure of accounts.

Multidimensional accounting allows to form the consolidated bookkeeping. For this purpose, it is necessary to unite Economic event data sheet of the organizations of the consolidated group, to designate them as «internal area», and provide unity of analytical directories. Financial results from operations between these entities also concern "internal area" of accounting.

In the future, multi-dimensional account can be used to organize the planning and budgeting. With the entry of additional attributes in it will be possible to combine different types of accounting: financial accounting, managerial accounting, tax accounting, cameral accounting, etc.

The use of aggregated accounting data mart to inform interested users can increase the completeness, accuracy of the transmitted information. This reduces the risk of misstatement of reporting data. The use of aggregated accounting data mart will significantly affect International financial report standards (IFRS). For a small volume and simple structure of accounting information a multidimensional accounting can be realized in the environment of the spreadsheet Excel.

## References

1. Lev, B., Gu, F.: *The End of Accounting and the Path Forward for Investors and Managers*. Wiley, Hoboken (2016)
2. Gu, J., Zhang, L.: Data, DIKW, Big data and data science. In: 2nd International Conference on Information Technology and Quantitative Management, ITQM (2014)
3. <http://www.systems-thinking.org/dikw/dikw.htm>. Accessed 21 Sept 2019
4. [http://indico.cern.ch/event/276139/contribution/59/attachments/501004/691997/DWH\\_and\\_BI\\_Improve\\_strategic\\_decision\\_making.pdf](http://indico.cern.ch/event/276139/contribution/59/attachments/501004/691997/DWH_and_BI_Improve_strategic_decision_making.pdf). Accessed 21 Sept 2019
5. Drury, C.: *Management and Cost Accounting*, 8th edn. Cengage Learning (Emea) Ltd., Andover (2012)
6. Van Der Poll, H.M.: *Towards a classification framework for accounting information*. Submitted in fulfilment of the requirements for the degree doctor commercii (financial management sciences). University of Pretoria. Republic of South Africa (2007)
7. Van der Poll, H.M., Gouws, D.G.: Issues surrounding the classification of accounting information. *S. Afr. J. Econ. Manage. Sci. (SAJEMS)* 12(3), 353–369 (2009)
8. Fairfield, P.M., Sweeney, R.J., Yohn, T.: Accounting classification and the predictive content of earning. *Acc. Rev.* 71(3), 337–355 (1996)
9. Hedish, N.D.: Account classification and principle codification. *Acc. Rev.* 34(4), 660–662 (1959)
10. Bedford, N.M.: The foundations of accounting measurement. *J. Acc. Res.* 6(2), 270–282 (1968)
11. Koltitz, D.: *Financial Accounting. A Concepts-Based Introduction*. Routledge, New York (2017)
12. Maciucă, G., Socoliuc, M.: The role of accounting system classification in the optimization of international harmonization process. *USV Ann. Econ. Public Adm.* 2(18), 200–206 (2013)
13. [https://www.researchgate.net/publication/327217155\\_Classification\\_and\\_Factors\\_Influencing\\_Accounting\\_Systems](https://www.researchgate.net/publication/327217155_Classification_and_Factors_Influencing_Accounting_Systems) 2018. Accessed 21 Sept 2019
14. Foskett, D.J.: Construction of a faceted classification for a special subject. In: *Proceedings of the International Conference on Scientific Information*. National Science Foundation, pp. 867–888 (1959)
15. Schmalenbach, E.: *Der Kontenrahmen*. Leipzig, G.A. Gloeckner (1927)
16. McCarthy, W.E.: The REA accounting model: a generalized framework for accounting system in a shared data environment. *Acc. Rev.* (3), 554–578 (1982). <https://www.msu.edu/~mccarthy4/McCarthy.pdf>. Accessed 21 Sept 2019
17. Shaposhnikov, A.A.: *Classification models in accounting*. Finance and Statistics (1982)
18. Kovalev A.E. *Multidimensional Model of Accounting*. In: *Advances in Intelligent Systems and Computing*. Digital Science 2019. Springer. pp. 314–324. <https://doi.org/10.1007/978-3-030-37737-3>.
19. Shvetsov, V.I.: *Database*. Internet University of Information Technologies. Moscow (2009)
20. <https://www.cfin.ru/vernikov/kias/vest.shtml>. Accessed 21 Sept 2019
21. Thomsen, E.: *OLAP Solutions. Building Multidimensional Information Systems*, 2nd edn. Wiley Computer Publishing, Hoboken (2002)
22. Schär, J.F.: *Buchhaltung und Bilanz: auf wirtschaftlicher, rechtlicher und mathematischer Grundlage, für Juristen, Ingenieure, Kaufleute und Studierende der Privatwirtschaftslehre mit*

- anhängen über "Bilanzverschleierung" und "Teuerung geldentwertung und bilanz". Springer, Heidelberg (1922)
23. Sokolov, V.Ya.: Accounting as the sum of the facts of economic life. Master, INFRA M., Moscow (2010)
24. Tsygankov, K.Y.: Double entry, Ledger accounts and their alternatives. Siberian FinancialSchool, no. 2, pp. 123–136 (2008)
25. Civil code of the Russian Federation, part 1., in the edition of the Federal law of 2010/05/08. N 83-FZ.

# Development principles for preparing financial reporting in the context of digitalization

Elena Dombrovskaya [0000-0002-3298-4337]

Financial University under the Government of the Russian Federation,  
Moscow, Russia, 125993

[https://doi.org/10.33847/2686-8296.2.2\\_5](https://doi.org/10.33847/2686-8296.2.2_5)

Received 16.10.2020/Revised 01.11.2020/Accepted 21.12.2020/Published 29.12.2020

**Abstract.** The information technology development and the digital economy have a significant impact on the business units' information system. The impact has resulted in a significant increase in the amount of business data used. With increased information availability, development of new processing methods and algorithms, the approaches to financial reporting need to be adjusted and modified. All changes to the basic principles of financial reporting have been included in the new version of the Conceptual Framework for Financial Reporting. It will come into force on January 1, 2020 for all business units developing their reporting policies under IFRS. The paper overviews the amendments and innovations introduced by the new version of this document. The Conceptual Framework dialectics has been reviewed based on all document versions since 1989. We analyzed the impact of digital technology on the transformation of accounting methods and principles and on the preparation of financial statements. The paper is of practical importance and helps to assess various areas of system development of the IFRS methodology in the context of digital economy.

**Keywords:** Conceptual Framework, Financial reporting, Digitalization, Principles.

## 1. Introduction

The system of international financial reporting standards is based on uniform approaches and principles determining the logics and goals of building such system, its key elements, their features and interaction. Conceptual Framework for Financial Reporting, or IFRS Framework, establishes the logics for building up the reporting information system. It determines the most important basic concepts of IFRS system. Not being a standard, the Conceptual Framework is the basis for all standards and deals with the key aspects of financial reports.

The approaches to financial reporting have changed with the changes in the economic environment; they reflected the processes in the global and national economies [1]. Principles of financial reporting in the public sector have been developed [2]. The realities of recent years included the new information technology, namely the digital one, the intensive development of which led to creating a brand-new concept of "digital economy". It is represented by various levels and has a significant impact on the business units' information system. A digital environment is being formed, which encourages development of platforms and technology, as well as effective interaction between various market players and economy sectors (areas of activity).

Digital technology has provided an unprecedented amount of business information used. As the result, for the last few years, the amount of information created has exceed more than twice that created for the entire human history [3].

The result was the growth of analytical and forecast value of reporting information about the business units' activity, the intensive use of which brought essential and substantive changes in the basic system of financial reporting [4].

Principles of valuation of accounting objects have undergone major changes [5]. In this regard, considering the dialectics of principles for preparing financial reporting is of great interest.

## **2. Methodology for preparing financial reporting according to the International Accounting Standards Board (IASB)**

Adopted by the International Financial Reporting Committee in 1989, the first version of the document under consideration was entitled "Statement of Principles for Financial Reporting". In 2001, after the reorganization of the Committee into the International Accounting Standards Board, the document was adopted as a backbone for IFRS. It was entitled "Conceptual Framework for the Preparation and Presentation of Financial Statements" [6].

The problems of development of the principles for financial reporting are touched upon in the works of many authors. Hendriksen and van Breda noted that the relevance is the most important qualitative characteristic of financial reports [7]. Needles and Powers Ethics defined that ethics is an important problem in preparing financial reports [8]. "Users of these reports must depend on the good faith of the people involved in their preparation. Users have no other assurance that the reports are accurate and fully disclose all relevant facts".

The work on amending the Conceptual Framework to update and bring it closer to the approaches used in the US GAAP was finished in 2010 with the document entitled "Conceptual Framework of Financial Reporting" adopted. Structurally, the new Conceptual Framework included the text of the previous document version with some amendments and two new chapters: "Objective of General Purpose Financial Reporting" and "Qualitative Characteristics of Useful Financial Information". That is, the new document was not adopted in the final version, and work on it has continued.

In addition to a shorter title, the most important amendments to the Conceptual Framework as of 2010 included clarification of the financial statements' purpose and the changes in its qualitative characteristics. In particular, the objective of the financial statements, which is to provide useful financial information, has been preserved in the new version; however, the composition of users and the content of the reporting information have been adjusted. Existing and potential investors, lenders, and other creditors were listed as its users. In the 1989 version, the financial statement data included information about the financial standing, financial results, and changes in the financial standing. In the 2010 document, their list is not detailed, and the purpose of financial reporting was described as providing users with useful financial information to make decisions about earmarking resources for the relevant enterprise. At the same time, the focus was on the fact that users need information about the future cash flows with the enterprise.

Amendments were also made to the composition of the financial statement's qualitative characteristics. In the 1989 version, they included clarity, relevance, reliability, and comparability. The Conceptual Framework of 2010 identified two groups of qualitative characteristics: the basic ones and those improving the usefulness of the reporting information. The basic ones were relevance and truthfulness, the improving ones were comparability, verifiability, timeliness, and understandability. As for the changes in qualitative characteristics, it should be noted that the reliability has been replaced by veracity: prudence (caution) was removed from the list of financial statement characteristics with a new characteristic of "verifiability" introduced. Only one limitation was left, that is the cost of providing useful financial information. Therefore, with the growing need for forecast information in the financial statements, there were higher veracity and verifiability standards.

In March 2018, the International Accounting Standards Board released a new, revised version of the Conceptual Framework. The main reasons for adopting the new concept document were the inconsistency with modern thinking, the need to restore

some previously excluded important provisions, and the lack of clarity. All these reasons were mentioned by those criticizing the 2010 version of the Conceptual Framework. The expansion of public information scope in the context of the digital economy development resulted in another transformation of the requirements for the same, with new version of the Conceptual Framework for Financial Reporting adopted.

The new Conceptual Framework is more clearly structured; there are more chapters. Although the new version includes all the issues listed in the old document version, the latest format approved in 2018 is easier to search information. We can assume the 2018 Conceptual Framework has become complete as compared with the previous version, which was an intermediate stage of the work on the document.

The composition and number of objectives as defined for the Conceptual Framework have been significantly reduced in the new version. The IASB excluded detailing in defining the goals and objectives of the Conceptual Framework, which was present in the 2010 version, mainly due to the targeted use of IFRS by various users. The integration of objectives shall be recognized as an indicator of the maturity of the IASB's position.

Therefore, on the one hand, the influence of subjective factors in the IASB's activities is growing. On the other hand, it is obvious that the diversity of business units' activities does not always fit into the rigid principles established by the Conceptual Framework. This is especially true for the situation with the growth of information about business units in the context of digitalization. In particular, Big Data allows for processing virtually unlimited amounts of information.

The Conceptual Framework's section dealing with the qualitative characteristics of financial reports has been radically revised in the 2010 version (Macve R. 2015). In 2018, no major amendments were made thereto. As before, qualitative characteristics have been divided into fundamental ones, previously called basic, and those that increased the usefulness of information. Disclosing the contents of these characteristics has been preserved in the 2010 format. The most important change in this chapter was that prudence has been re-added to the list of financial statement characteristics (Efimova, O., Rozhnova, O. 2019). Prudence is related to reporting neutrality and means that caution shall be used when using judgments under conditions of uncertainty. The use of prudence means that assets and income shall not be overstated, and liabilities and costs shall not be underestimated. This is done to prevent underestimating assets or income, or overestimating liabilities or costs. At the same time, the need for an asymmetric approach to different items is denied, while recognizing asymmetric requirements in some standards to choose the most relevant information.

Therefore, the restoration of prudence, traditionally used in the recognition and evaluation of accounting items, demonstrates consistency and continuity of the principles of financial reporting under IFRS. At the same time, an increase in the amount of information processed inevitably results in higher requirements for its qualitative characteristics, toughening individual principles to prevent information distortion and provide an objective basis for its auditing.

As a result of the increased amount of information processed, its limits have expanded in terms of using both reporting and forecasting data. In this regard, the IASB acknowledged in the 2018 version that the information usefulness might be reduced because of a high uncertainty of the estimates. In such cases, preference should be given to information that was less significant, but more clearly defined in the estimates. A landmark point was restoring the concept of substance priority over form in the Conceptual Framework, which was excluded in 2010. Adhering to these conceptual approaches has one goal: to truthfully present information in financial statements.

The terms and definitions used in the new version of the Conceptual Framework have also been amended. In particular, the concepts of assets and liabilities have been updated. References to compulsory confidence in the inflow (outflow) of economic benefits were excluded from their definitions while recognizing reporting items. This allows extending the list of items included in the statements, which is very

important for generating reliable information about the resources of the business unit in the context of big data processing.

At the same time, this will improve the accuracy of information about the assets and liabilities of the company, because the coverage of recognized objects will increase. The predictive nature of financial reporting, which would reflect a broader list of facilities, would also improve. The list of objects reflected in the financial statements will also be expanded by reflecting all the rights that the reporting entity has to benefit from the use of the asset and the obligation to transfer the benefits.

An important provision of the Framework is the evaluation of elements of financial reporting. The conceptual framework is established by two groups of estimates - "historical value" and "present value." The fair value may be fair value, reflecting the market value; the cost of use, representing the discounted value of future cash flows generated by the object; the fair replacement value, which is the price of acquisition of the equivalent asset at the valuation date. Fair value assessment requires the reporting entity to undertake serious work related to the collection, processing and analysis of various information. Large amounts of data must be used. The need to use predictive values complicates the valuation even when it comes to determining historical value. For example, the acquisition cost of fixed assets (IAS 16) or rights to use a leased asset (IFRS 16) includes the cost of liquidating the asset at the end of its use. In order to determine the capitalized value, it is necessary to establish: cost amounts,

Which will allow for its liquidation, which is expected to take place in many years, the establishment of a discount rate, and the subsequent continuous monitoring of possible changes in these figures, as well as the date of liquidation, will be necessary. It is digital technologies that provide relevant and reliable information and reduce the risks of non-recognition of objects in financial reporting.

The relevance and reliability of the reporting information depends to a large extent on the quality of the source data, their adequacy, compliance with the task to be solved, reliability, as well as the speed and depth of their analysis, including their classification into different groups. In this case, it is no longer limited to financial information, information of a non-financial nature is also necessary, it helps to obtain a complete picture of events and to verify financial data. In this way

At the same time, real achievement of relevance and reliability of reporting information, timely and truthful reflection only of real assets and liabilities, income and expenses, becomes possible only in conditions of application of digital technologies. It is the development of these technologies, such as advanced analytics tools using BIG DATA, that allows the fundamental provisions of the Financial Reporting Framework to be implemented in practice.

### **3. Data and Methodology**

The Russian economy which is highly dependent on oil and gas exports is now faced with a decline in Petro-dollars because of falling oil prices. Under new circumstances, Russian leaders have to decide how to reduce the energy resource dependence of the economy and, therefore, pay close attention to tourism and hospitality as potential drivers of the national economy. Given Russia's significant tourism potential in the world market, it is necessary to divert the country's economy from product dependence, identifying the main problems of competitiveness and financial support for the tourism industry in digital age in terms of increasing its export potential and ensuring a high level of attractiveness of tourist destinations in the formation of an effective program for the development of inbound tourism.

The authors used such methods as theoretical modeling and analysis. Official statistics and analytical reports of Statista, Inc., UNWTO, WTCC and Russian federal authorities served as the information basis of the study.

#### **4. Transformation of the Companies' Information System in the Context of Digitalization**

Corporate information systems have undergone significant changes in the context of automation and digitalization. These transformations have a significant impact on the development of an accounting model that allows for solving complex information processing tasks. A special place among modern information systems is taken by innovative systems that are based on the use of artificial intelligence technology, big data processing, the emergence of new analytical tools and cloud services.

Digitalization of economic processes has presented new requirements for identifying accounting and analysis objects and structuring accounting and analytical information using cloud technology. This happened due to the fact that processing of large amounts of data and its analysis can significantly increase the efficiency of production and commercial cycle processes. The ability to provide digital products and services has become vital in the context of the 2020 economic crisis that is aggravated by the coronavirus pandemic.

As digitalization processes continued to develop, new hybrid and modifiable accounting items emerged in corporate practice: digital assets, electronic money, virtual digital property, smart assets, smart contracts, blockchain, etc. This list is updated regularly as the information technology develops. Methods advance to allow evaluating new accounting objects: intellectual human capital, customer base, innovative products, etc.

In the context of digital technologies, there is a tendency to converge, deepen relations with users of reporting information. Working with them is individualized, the company is involved in the interests of investors and creditors, sensitive to their preferences. In these conditions, the flexibility of the reporting information is important, and it is possible to fill it with the data that is currently needed by a particular user. The format of financial reporting is increasingly expanded by non-financial information. Digital technologies save transaction costs and provide new potential for the development of a reporting information system. In this regard, it makes sense to assess the feasibility of the general user concept of financial reporting and its replacement with individual reporting. Financial reporting should not be driven only by the requirements of laws, regulations and standards. This inevitably leads to the depreciation of reporting information and formalizes the process itself. Admittedly, digital technology has provided unprecedented openness to today's business, and financial reporting has ceased to be the sole source of business information. The content of reporting should be based on information requests and expectations of specific users, provide them with real tools to solve their problems.

The speed of information exchange in the modern economy has a huge impact on the functioning of a corporate information system. Information sharing speeds up in conjunction with the expansion of circle of stakeholders, i.e. persons who are involved and interested in economic processes. To maintain communication and to preserve control when managing an economic entity, one needs to master modern digital tools. They allow to quickly monitor the processes occurring in the company's management systems. XBRL is such a digital tool, which is a format for transmitting regulatory, financial, and other reports.

Digitalization of the economy has significantly changed the nature of competition. From costs, competition has shifted to creativity. Creativity, innovation, non-standard thinking can give competitive advantages to companies. Creativity is evident not only at the technological and organizational level. Accounting and financial reporting are becoming creative. The development of a creative approach in accounting and financial reporting is related to the complexity of business processes,

the absence of appropriate regulations for new accounting objects and new technologies. Creativity in financial reporting is manifested in the use of an accountant's professional judgment to report innovations. The purpose of creative financial reporting is to overcome the limitations of existing accounting standards, to include forecast estimates and really necessary and useful information in the reporting. The conceptual basis of financial reporting should introduce into the normative plane both creative accounting, professional judgment of the accountant, and observance of professional ethics.

Expansion of information boundaries of financial reporting in conditions of digitalization, inclusion of forecast information and information resulting from professional judgment of the accountant, bring to the fore the problem of subjectivity and risk of fraud in reporting. The value of financial information under such conditions can be significantly reduced [9]. In order to protect the interests of users, it is necessary to establish reasonable boundaries for the creativity of financial reporting. The concept of accurate and honest reflection of information continues to be the philosophical basis of accounting. The return to the Financial Reporting Framework of the discretion requirement underscores this. However, objective strengthening of creativity in reporting requires guaranteed observance of the right of users to quality information. A fair reflection of creative reporting information can be considered a reflection that does not mislead users or lead to financial losses.

The largest consequence of digitalization for financial reporting is the use of blockchain technology. Many companies are already introducing blockchain into their resource planning systems to manage purchases. Blockchain technology allows safe registration of transactions, ensures transparency and reliability of information. However, its implementation in companies is hampered by the complexity and lack of a regulatory framework. Accountants have yet to understand the new trend and assess its impact on financial reporting.

## **5. Principles for Preparing Financial Statements in the Context of Digitalization**

Digital transformation of the information space involves not only changing the technology of obtaining, storing and using financial information but also adjusting the methods for collecting and processing information in accounting as well as its systematization in financial statements. These processes require developing and supplementing the system of principles for reflecting accounting information and for preparing financial statements.

An important principle in the preparation of financial statements is the uncertainty management principle. It involves building up a company's threat and risk management system. This allows for identifying and evaluating risk parameters and indicators that appear based on accounting and management data. As a result, financial statements will include a comprehensive assessment of threats to the company's development. All stakeholders will be able to respond promptly to such threats; that is, implementing the uncertainty management principle will provide information support for financial statements and the focus on further assessment of conditions under which threats and risks appear.

The principle of openness should be mandatory in the preparation of financial statements. This principle implies the reliability and verification of information, the use of a convenient report format, the timely publication of reports, the studying of interests of various stakeholder groups, and taking into account the identified expectations for reporting information.

Implementing the adaptability principle will ensure the ability to adapt to transformation, shifts and modifications of the company's reporting information system, taking into account external and internal requirements.

The comprehensiveness principle in the preparation of financial statements means that all subsystems and elements of the company's activity must be intricately linked to each other. The use of this principle will allow to comprehensively identify the company's functioning problems that relate to economic, social, scientific, technical, environmental, legal and other areas, and to reflect them in statements.

The predictability principle is based on mandatory taking into account the probable nature of real processes in the corporate development to ensure sustainable development and achievement of high performance through forecasting income and expenses.

The principle of self-development and self-learning implies the ability to develop the business entity based on creating its information system in line with the requirements of the knowledge and innovation economy. It involves developing a large communication network between all interested parties.

The information security principle is based on mandatory creating and maintaining the information and accounting environment security. It involves the ability to use "smart contracts".

The visualization principle is based on creating an informative and user-friendly form for presenting reporting information.

The principle of close cooperation with various groups of interested market players. This principle involves building up reporting information taking into account the expectations of all stakeholders.

The system consistency principle means providing for a coordinated interaction between all subsystems of the reporting information preparation system.

The synergy principle considers the need for cooperation between companies and stakeholders and influencing the accounting and reporting information system, aiming at changing the performance through the emergence of new qualities of such system.

## **6. Results**

When it comes to financial reporting in a digitalized economy, the most important thing is to ensure its qualitative characteristics, not to lose user confidence in the reporting information. Therefore, recognition of reporting items is linked to the fundamental characteristics of the reporting information: relevance and reliability. The IASB did not take the path of increasing the financial information uncertainty and riskiness, which could manifest themselves in the context of lower limits for recognizing individual items. Quality content of financial statements remains the most important task for informed economic decision-making.

Digitalization of the economy has had a significant impact on the reporting information system. Increased communication with users and the development of non-financial reporting require greater flexibility and user-oriented financial reporting. This can be facilitated by the concept of the individual usefulness of financial reporting, aimed at meeting the interests of specific groups of users. This approach will stop the impairment of financial reporting information that is currently taking place.

Digital technology has made creativity a competitive advantage. Financial reporting has also become creative, based on the professional judgement of the accountant and the observance of professional ethics. These issues should be included in the Financial Reporting Framework in order to enable the legal field to make a decision on the generation of reporting based on professional experience and expertise. To limit the risk of fraud in financial reporting, it is necessary to establish the boundaries of creativity. The main criterion may be the absence of financial losses for users of information.

The use of blockchain technologies in financial reporting is subject to serious

reflection in the future. Technical difficulties and lack of regulatory frameworks have so far limited the ability to assess the impact of this innovation on financial reporting. But it is likely that these issues will be the subject of professional debate in the years ahead.

## 7. Conclusion

The purpose of this paper was to review the dialectics of the Conceptual Framework for Financial Reporting and to assess the impact of the economy digitalization on the development of financial reporting principles. The analysis of the various versions of the Conceptual Framework and, primarily, that of 2018, showed that amendments made to the latest version have made the document more convenient. Structuring by chapters better reflects the document content and makes it easier to find the necessary information. Work is underway to formulate and summarize the Conceptual Framework objectives, the number of which is reduced to three. With significant expansion of the information processed in the context of economy digitalization, as well as the influence of the subjective factor needed more rigorous reporting approaches. Therefore, prudence has been re-added to the list of financial statement characteristics. This was done to prevent distortion of the reporting information and increase its relevance. The use of forecast information in the statements shall not reduce its veracity. Preference shall be given to information based on more specific estimates.

Further development of the Financial Reporting Framework should take into account the processes taking place in the economic and technological environment. They are related to changes in the reporting information system, promotion of non-financial reporting, orientation to specific information requests of users, expansion of the scope of application of professional judgment, ensuring reliability of financial reporting in conditions of digital transformations.

The study could not cover all existing problems. But consistent work to address them will make significant progress in the area under consideration and determine the role and place of financial reporting in a changed world.

## References

1. Henderson S. et al. Issues in financial accounting. Pearson Higher Education AU (2015)
2. Kachkova, O. E., Vakhrushina, M. A., Demina, I. D., Krishtaleva, T. I., Sidorova, M. I., Dombrovskaya, E. N., Klepikova, L. V. Developing the Accounting Concept in the Public Sector// European Research Studies Journal, 21(1), 636-649, (2018)
3. Flower J., Ebbers G. Global financial reporting. Macmillan International Higher Education (2018)
4. Efimova, O., Rozhnova, O.: The corporate reporting development in the digital economy. In: Antipova, T., Rocha, A. (eds.) Digital Science. DSIC18 2018. Advances in Intelligent Systems and Computing, vol. 850, pp. 71–80. Springer, Cham (2019)
5. Alexander, D., Bonaci, C., Mustata, R.: Fair value measurement in financial reporting. Procedia Econ. Financ. 3, 84–89 (2012)
6. Macve R. A Conceptual Framework for Financial Accounting and Reporting: Vision, Tool, Or Threat?. Routledge (2015)
7. Hendriksen, E., van Breda, M.: Accounting theory. Richard D. Irwin. Homewood, IL (1992)
8. Needles, B., Powers, M.: Principles of Financial Accounting. South-Western, Cengage Learning (2011)
9. Tassadaq F., Malik Q. A. Creative Accounting & Financial Reporting: Model Development & Empirical Testing //International Journal of Economics and Financial Issues, 2, 544-551, (2015)
10. Dombrovskaya E. Development of the conceptual framework for financial reporting in the context of digitalization. In: Antipova, T., Rocha, A. (eds.) Digital Science. DSIC 2019. Advances in Intelligent Systems and Computing, vol. 1114, pp. 334–344. Springer, Cham (2020)

## **Aims and Objectives**

Published online by ICS two times a year, Journal of Digital Science (JDS) is an international peer-reviewed journal which aims at the latest ideas, innovations, trends, experiences and concerns in the field of digital science covering all areas of the scholarly literature of the sciences, social sciences. The main topics currently covered include: Digital Communications and Network; Digital Economics, Education, Engineering, Finance, Health Care.

The main goal of the journal is the effective dissemination of original incites/results generated by the human brain and presented/reflected in articles using modern information/digital technology.

## **Editorial Board**

**Editor-in-Chief** Tatiana Antipova, ICS,  
<https://orcid.org/0000-0002-0872-4965>

**Associate Editor** Julia Belyasova, Catholic University of Louvain, Louvain-la-Neuve, Belgium;  
<https://orcid.org/0000-0001-6983-2129>

## **Editors**

Abdulsatar Sultan, Catholic University in Erbil, Erbil, Iraq;  
<https://orcid.org/0000-0001-5090-5332>

Jelena Jovanovic, University of Nis, Nis, Serbia;  
<https://orcid.org/0000-0001-7238-6393>

Lucas Tomczyk, Uniwersytet Jagielloński, Krakow, Poland  
<https://orcid.org/0000-0002-5652-1433>

Natalya Sukurova, State University of Telecommunications, Kyiv, Ukraine  
<https://orcid.org/0000-0003-4297-1123>

Olga Khlynova, Russian Academy of Science, Moscow, Russia  
<https://orcid.org/0000-0003-4860-0112>

Omar Leonel Loaiza Jara, Universidad Peruana Unión, Lima, Peru  
<https://orcid.org/0000-0002-3262-709X>

Roland Moraru, University of Petrosani, Romania  
<https://orcid.org/0000-0001-8629-8394>

Tjerk Budding, Vrije Universiteit Amsterdam, Netherland  
<https://orcid.org/0000-0002-5343-7535>

Zhanna Mingaleva, National Research Polytechnic University, Perm, Russia  
<https://orcid.org/0000-0001-7674-7846>

Quang Vinh Dang, Industrial University, Ho Chi Minh City, Viet Nam  
<https://orcid.org/0000-0002-3877-8024>

## **Contact information**

**Website:** <https://ics.events>

**Email:** [conf@ics.events](mailto:conf@ics.events)

