Conceptual Model of National Intellectual System for Children Safety in Internet Environment

Rasim Alguliyev
Information Technologies Institute of ANAS, Baku, Azerbaijan
E-mail: director@iit.science.az

Sabira Ojagverdieva
Information Technologies Institute of ANAS, Baku, Azerbaijan
E-mail: allahverdiyevasabira@gmail.com

Received: 18 January 2019; Accepted: 19 February 2019; Published: 08 March 2019

Abstract—The article presents a conceptual model for the national intellectual system aiming the safety provision of the children in Internet environment. The structural components and work principles of the proposed model are explained. This model employs web-analytics, data sanitization (cleaning) technology, expert systems, text mining, clustering and classification methods, content filtering and etc. to protect children from harmful information in virtual environment. By using data sanitization methods, the study presents a conceptual model for obtaining more important, useful and age-corresponding information from internet resources and preventing harmful information.

Index Terms—Children safety in Internet, data sanitization, conceptual model, harmful content, Internet addiction of children, intellectual system.

I. INTRODUCTION

Information technologies in nowadays encompass almost all activity areas of people and influence the changes in their standard of living. The tendency of rapid development of Internet fosters this process. The application of wideband internet allows users to send and receive data of large volume with high speed [1]. The volume of information is growing. Conventional information processing tools are not capable to deal with this information flow and the only solution is to employ Big Data technologies [2]. Nowadays Big Data is used in all areas of information technology (computer networks, information security) [2]. It is also estimated to be aimed at families and children. Thus, parents always try to choose the most optimal ways for their children's safety and health.

Children and adolescents actively use the Internet, and most of them behave being affected by the interactive media culture. They use social media, implement mobile downloads, and engage in a variety of games on the Internet. Most teenagers tend to share their personal information (name, family information, place of residence, school they study, contacts, interests etc) in web diaries or blogs and social networks. It's common for kids to frequently update their profile and status on social networks, share their daily experiences, online experiences and their location, and chat with friends. New participatory platforms create favorable conditions for the socialization of young people, however the information they share make them available for everybody. Consequently, it leads to the contradictions between the representation of children online and the protection of privacy.

Internet presents unique opportunities for obtaining unlimited information in a short time interval and plays an important role in transforming them as an active member of information society. However, at the same time, it causes users to be prone to cyberthreats [3,4].

The children are active in networks as much as the adults. The placement of information of various content in Internet causes children to encounter threats in virtual environment as well [5]. The wide spread of information such as pornography, torture, brutal scenes, the propaganda of drugs, alcoholic drinks, terrorism, bad habits and etc. in virtual space and social media in particular inains undesirable moral and ethical characteristics in people [6]. Threatening and violation children and imposing information-psychological impact on them by ill-intentioned persons is often encountered [7]. This fact creates several problems related to information security alongside the advantages of using Internet [5]. The problems caused by such threats demand the solution of not only technological, as well as social issues.

It is known that, the world culture, ethical actions, religious views, habits and etc. are continuously transmitted into virtual space. Internet has entered the socio-economic life of people. It is impossible to protect each citizen accessing the Internet from spams and cyber attacks. The ways for protection against threats concern
regular citizens, experts and responsible persons. It is not sufficient to carry out awareness-raising works in order to protect children against information harming their health and psychology [8]. At this stage, the measures taken are in form of recommendations. Currently, there exist numerous software packages and security systems oriented towards the solution of problems of a secure use of Internet [9,10].

The article analyzes the perspectives of implementing data sanitization according to specified rules for the provision of the safety of children in Internet environment at national level and presents proposals for implementing suggested sanitization policy. A conceptual model of the national intellectual system is proposed for the purpose of preventing the harmful content from Internet.

II. RELATED WORK

Nowadays, children cannot imagine themselves without the Internet. In this regard, the major challenge for parents is preventing their children from unsafe information on the Internet. A huge number of devices with the Internet access makes it more difficult to control children each moment they become online. Currently, many software and security systems that address children's safe use of the Internet are available. However, despite the use of such systems, there is a great need for creating more robust software tools and mechanisms for ensuring the safety of children in the network.

Parental control application blocks unwanted web content, restricts presentation time, and bans the use of hazardous applications, making children to surf on computers and mobile devices safer [11]. At present, children are actively using almost all devices with the Internet access, subsequently modern parental control systems should manage on with this situation. Nowadays, different systems ensuring children’s safer use of the Internet on computers and other mobile devices are available [11]. For example, software “Qustidio” filters the content obtained from the Internet online regardless of the browser [12]. The software is easy-to-install and guides HTTPS traffic as child security software. It defines the timeframe for the devices ensuring regular functioning of the Internet, web browsers and networks, and controls the spatial and time settings. The disadvantages of this software incudes its high price and performing the social monitoring only via Facebook.

Net Nanny is another system designed to ensure the safety of children. This application is supported by Windows, MAC and Android operating systems [11]. It has high filtration capacity and controls the applications of Android OS and the time spent on the Internet. It also controls mobile phones. The advantage of the systems includes its capacity to restrict the time spent on the web browser. The notification is sent to the parent's phone warning him/her when a child wants to access the website. It has some restrictions on iOS. This web-based multi-platform approach enables parents to monitor the activity of their children on the web, however it has not been updated in recent years. One of the disadvantages includes its high price and the spatial issues, besides it has no opportunities for time control [13].

Symantec Norton Family provides online safety of children. It can be installed on all devices connected to the network (computer, mobile phone, iPad, etc.) [11,14]. It does not apply any restrictions on devices or the number of children applying here. The software also has spatial tracking capabilities [11]. However, all editions of the software are paid. It does not filter browsers not supporting HTTPS. It has also some problems related to social media tracking. Another software called “Kaspersky Safe Kid” has high potential and no limitations on the number of users and equipment on the application. It effectively performs Web filtering and has an extended notification system [15]. Its disadvantage includes restrictions on IOS, besides the content filtering is performed for certain sites. Social monitoring is available only for Facebook and VKontakte users.

Circle with disney is easy to install and controls each device connected to the camera [11]. It performs web content filtering based on a specific category. Sleep time automatically interrupts the Internet. The software is quite expensive. Adjustment to devices is difficult. It has minimal accountability capabilities. Circle with disney security device controls easy mobile applications. It enables parents to filter content and can generate a separate time limit for each device connected to the house network [11].

Clean router performs parental authentication and control for each device on the network [16]. It also provides secure authentication and sets the Internet time according to the devices. It reports about the blocked sites clicked by the user. Additionally, it sends a report on a daily activity (statistical information) on request. The disadvantage is that it does not specifically generate user-friendly reports. It is not capable to filter some porn sites. The analysis of the aforementioned systems shows that there is no any system that fully ensures children's safety on the Internet. Each system has certain deficiencies. Moreover, children and adolescents with a clear understanding of computer software are able to easily interfere the abovementioned software and systems. They can access harmful web sites through anonymous proxy sites and other browsers. They may even turn off the safe search engine.

Given these, the proposed system is estimated to be deployed to the provider. This system cannot be interfered by children. The system is linked to parents, schools, responsible persons, and so forth. It always provides extensive information on the web activities of children (web pages they visit, their access time, the time they spend there, etc.).
III. THE ROLE OF SANITIZATION POLICY IN INFORMATION SECURITY

Various measures are taken at national level in order to protect children in network environment. These measures are carried out in accordance with the policy based on social norms, customs and national-moral values of population as well as the legislation. At first, the web-pages must be analyzed in order to detect harmful content in internet environment. However, the complexity and variability in the structure of web-pages complicates the conduct of analytical processes [17]. It is possible to collect statistics regarding the information in web-pages by employing web-analytics tools in order to analyze those pages. These tools include meters and log analyzers. A meter – is an external tool allocated with small-size JavaScript code in order to collect statistics on a webpage. Such tools are usually of 1-2 kb size and do not harm the size as well as the design of web-page. While a user enters the web-page, the browser uploads an image linked to the meter and as a result, the meter located in the web-page for information collection becomes active and records traffic. Information collected in meter is entered to database [18]. Log analyzers – collect log files in the server (server journal) of website location within a specific period via internal registry tools. Registry files are deemed as a primary source of information collected in web servers (log files) regarding the traffic in web connection, behavior of users and websites visited by them [19]. Log analyzers analyze this information and store in internal archive. In this vein, pages reflecting the statistics and state of website are automatically created [20].

Web-analytics – previously, registry files have been used for eliminating problems in the system; however, recently, they bear larger importance in preventing or mitigating problems caused by threats [21]. Children spend substantial amount of time in global network and willingly or unwillingly download numerous contents pertaining to various topics. Sometimes their “surfing” in different websites for hours may lead to safety violations. Hence, the importance of information safety in network and the safety of children in particular has become a primary concern. Log files collected in web server can be used for the purpose of observing and evaluating the behavior of children in network environment as well as determining their interests. Moreover, it is possible to determine the degree of addiction of children to Internet with the help of log files.

Data Sanitization (DS) – is one of the methods of protection against threats from Internet network. The gist of this method is the prevention of audience’s access to vulnerable, personal, age-restricted, confidential, not recommended etc. contents, in other words, the application of specific limitations on contents [22]. DS is usually used during the editing of published materials, daily online media, cloud technologies and in other issues [23].

All sectors are subject to changes and reforms in the process of the formation of national e-government. The provision of information security can be listed among these reforms alongside the development of the regulation of mutual inter-agency communication, the creation of the classification of state services as well as the implementation of a unified technical architecture and software and hardware platform in e-government environment [24]. Such that, e-government must function sustainably, must be reliable and immune to threats. That is, e-government must provide information, energy and etc. security. The presented concept is considered as an infrastructure in e-government platform and is a network in the national segment of Internet environment. This proposed system is one of the multiple layers of e-state such as e-education, e-government, e-court, etc.

The security and sanitization policy of selecting harmful content from information traffic in Internet is variable. This policy varies among countries and depending on the filtering policy of contents. Content filtering considers the functions carried out by special filters included in servers in order to provide the information security. Depending on the sanitization policy, contents deemed as harmful may be banned not only for children, but for adults as well. In this case, special rules are imposed on filters.

The system of automatic cleaning of harmful content at national level is a special program module applied in corporate networks and is potentially developed according to a specific hierarchy starting from each family. This system is considered to apply sanitization policy at national level in e-state platform. The system blocks information harming child psychology, behavior, health and etc. Moreover, it neutralizes threats towards children and teenagers, is able to provide the security for minors by cleaning those threats and carries out the forecasting of potential threats in future.

![Hierarchic model of Traffic Sanitization Policy](image)

Based on above-mentioned, this system can be seen as
an integral part of e-state security. Such that, conducted sanitization policy is a projection of security policy in e-state environment. The cleaning of Internet traffic entering the country on organizational as well, as on individual basis is a part of sanitization policy and carried out as below (figure 1):

The process of sanitization in state agencies consists of 4 layers such as national, provider, corporate (academia, education, university) and user level active in corporate environment. The process of sanitization of Internet resources carried out by citizens on individual basis is carried out at 3 levels: national, provider and individual.

High-quality sanitization is not always attainable in the process of sanitization at national level. In this case, the sanitization process is conducted at the legislative level – for everyone in general and children, teenagers or others are not distinguished.

The age cohort of children and teenagers are not taken into consideration while cleaning the content related to those. The sanitization can be done at 3 levels: weak, medium and high. The sanitization at individual level mainly pertains to weak sanitization, in most cases, people do not consider the problem of content sanitization. The sanitization at medium and high levels is only carried out in state agencies and education institutions. As the sanitization requires specific software and additional hardware, it is costly and hence, individual Internet users are interested in content sanitization only in critical situations.

IV. A CONCEPTUAL MODEL FOR SANITIZATION OF INTERNET TRAFFIC

The model presented in Figure 2 is developed based on the traffic sanitization of webpages obtained from Internet. The proposed conceptual model encompasses rich Internet environment with all sorts of contents (harmful-harmless, useful-useless, etc.) on one hand and children and teenagers on the other hand. It is important to evaluate the relations between those and coordinate the situation correctly. According to the model demonstrated in Figure 2, filtered webpages are presented to the audience consisting of children by considering their age, interests, health, knowledge level, psychological state and other features. In order to provide the functioning of the proposed conceptual model as an intellectual system, some hierarchically ordered components are used: traffic separator, traffic collector, user authentication and etc. By using the data sanitization method, the model is considered to prevent harmful information and make specific decisions.

It is known that, children using Internet vary according to their interests and age groups. These differences are observed in social networks as well as in relation to digital objects. Children using Internet are divided into following age groups in some studies. The division to age groups is carried out in accordance with rules or acts adopted in healthcare and governments (under age 6; 6-12 years; 12-16 years; 16-18 years) [25].

It can be concluded from the division that, presented information must have various features for child audience. The algorithm applied for the purpose of corresponding information evaluations simplifies the workload when web-content is divided into categories. Some categories such as time, topic, focus of interest, physiological capabilities, type of content, parental control and psychological profile of a child are added to considered categories. Indicated categories can be added or changed according to national, religious, racial, ethnic, geographical and other features. It is possible to conduct analysis by categorizing information from Internet via traffic separators and grouping them according to content.

V. A CONCEPTUAL MODEL FOR SANITIZATION OF INTERNET TRAFFIC

The traffic separator to be employed in the model – is used for content analysis or the analysis of the content of information and the study of texts and graphic information. It is the formal observation and statistical procedure conducted on the quantitative indicators of this information [26]. Traffic separators carry out the separative and divisive functions of the content in some sense. The study utilizes traffic separators for the purpose of the grouping of web-content according to its type. It is considered to utilize traffic separators to collect inquiries during the access to the system and to determine the semantics according to the type of web-content. The wide spread of multimedia files such as advertisement, audio, video, image and etc. and online services enrich the content of these pages.

The changes in the design of a page for the purpose of additional content alongside the primary content of the pages make the structure of those more complex. The inclusion of various additional elements enriches the structure of web-pages on one hand and complicates the
analysis of those on the other. In some approaches to the analysis of web-pages, the page is divided into framed parts via horizontal and vertical lines [26]. In other approaches, the content-analysis of web-pages is based on the clustering methods according to the topic [27].

The study envisages to divide multimedia resources according to their types via the content separator in these pages in order to detect required pages: text-type information, image files, audio and video files.

The sanitized multimedia traffic filtered through the separator filter is collected and there emerges a need for clustering of information based on the topic. Therefore, the process reaches the next step. Sanitized multimedia reaches the entrance of traffic collector. Multimedia traffic entering from Internet is sanitized, aggregated and structured. The sanitization of web-pages is carried out at the technological level. Figure 3 describes the division of the content of web-pages according to entering information features. The information is divided into 3 parts: information with low importance, important information and harmful information.

The blockage of additional pages with advertisement, spams and other content is carried out via auxiliary JavaScript and Flash softwage packages. If there exist banned tags, content propagating violence and horror, pornographic information, words, argotic expressions, images or videos on pages, the closure of those pages and the blockage of inappropriate advertisements are considered [28].

The source of contents of various type received from Internet is detected during the filtering and stored to data treasury. Some tools of the intellectual analysis can be utilized in this case (text mining and etc). According to predefined rules, banned web-pages or pages with harmful information content are detected and stored into “banned data treasury”. Parental control is required in order to strengthen the control over the information with low importance. Information with low importance is a part of general information and is deemed harmless for children audience. The biometric features of the information gathered as a result of the analysis and classification must be considered before presenting those to children.

Biometric technologies provide the transparency, accuracy as well as the safety during the presentation of various types of services in a modern world. At the same time, a secure access to various objects can be obtained by using biometric technologies. Biometric technologies is based on the science of biometrics considering the individual and unique characteristics of a person [29, 30].

Biometrics encompasses the recognition system of humans based on one or numerous biological or behavioral characteristics [31]. Biometric characteristics include the shape of hand, fingerprints, the eyelid, voice parameters, elements of face and etc. It is divided into two categories based on physiological and behavioral features [32, 33].

Biometric information is used for managing the access features and access control in the field of information technologies. The biometric system can function in three modes [32]:

1. Registration – is a process of registration in the database of objects;
2. Verification – biometric parameters and identifiers are presented to biometric system during biometric verification. The sample is compared with the record corresponding to this identifier. This verification determines the correspondence to pre-developed biometric template or standard;
3. Identification – biometric information is obtained and compared with samples in biometric base thereafter. A template with the highest correspondence level is considered analogous to presented sample.

The biometric identification system is a process of comparison of the information presented by the user with existing information database. The biometric system is a system of image recognition and consists of the following elements [34].

- Sensor module – accepts biometric data of an individual;
- Feature extraction module – processes biometric data received for calculating the feature parameters;
- Comparison module compares the values given in a template and comparison value is calculated.

Fig.3. The Scheme of Content-based Separation of Web-pages

The source of contents of various type received from Internet is detected during the filtering and stored to data treasury. Some tools of the intellectual analysis can be utilized in this case (text mining and etc). According to predefined rules, banned web-pages or pages with harmful information content are detected and stored into “banned data treasury”. Parental control is required in order to strengthen the control over the information with low importance. Information with low importance is a part of general information and is deemed harmless for children audience. The biometric features of the information gathered as a result of the analysis and classification must be considered before presenting those to children.

Biometric technologies provide the transparency, accuracy as well as the safety during the presentation of various types of services in a modern world. At the same time, a secure access to various objects can be obtained by using biometric technologies. Biometric technologies is based on the science of biometrics considering the individual and unique characteristics of a person [29, 30].

Biometrics encompasses the recognition system of humans based on one or numerous biological or behavioral characteristics [31]. Biometric characteristics include the shape of hand, fingerprints, the eyelid, voice parameters, elements of face and etc. It is divided into two categories based on physiological and behavioral features [32, 33].

Biometric information is used for managing the access features and access control in the field of information technologies. The biometric system can function in three modes [32]:
• Decision making module – identifies the user based on the comparison values calculated in comparison module.

Children are allowed to enter the system based on the template developed by considering biometric characteristics of those. A positive or negative decision can be made by the system in this case. A harmful impact to a human body by various modules in computer must be considered during the biometric identification. The solution of various scientific problems is required during the functioning of this intellectual system and the adoption of specific decisions.

VI. ON POSSIBILITIES AND PROBLEMS OF INTELLECTUAL SYSTEM

The infrastructure of the conceptual model is constituted of the tools supporting the decisions of administrators as well as parents. The systematic approach is employed during the solution of issues requiring certainty, uncertainty, accuracy, comparison and risk during decision-making [34]. The consideration of criteria corresponding to specific issues is important during accurate decision-making.

Web-contents separated according to biometric characteristics – are stored in database as information tables, diagrams and reports. This web-contents can be used in various decision-making processes. The primary components of those are given below:

– Specification of suspicious web-content transmitted to users by an administrator;
– Specification of the most important web-content;
– Transmission of the most important web-content to a user in future requests without conducting the analysis.

It is important to take into consideration all possible indicators while solving safety issues of children in Internet. There is a necessity to carry out the specification of harmful web-pages in the environment of reception, comparison and analysis of large-size information during the decision-making by an administrator. The system presented in the article is an open system; it can be integrated to international system and function in worldwide information network by considering national features. The integration of this intellectual system to Hotline, Stopline international networks and their coordinated functioning is envisaged. In this case, the rapid detection of online crime events, rapid receipt of required information and etc. issues can be solved.

The conduct of national sanitization policy and the use of decision-making systems is necessary in this case. The proposed intellectual system can apply full blockage or partial blockage by evaluating the harmful content. If harmful content is detected during the repeated control of specific web resources (according to a specific time interval), a partial blockage is carried out by the system.

National Information Center (NIC) informs citizens regarding threats from Internet and shows the methods for protecting their children against these threats. The center considers to organize consultancy hours, provide educative materials and hold trainings in order to protect children against these threats. It is envisaged to create a web-page of NIC, educative materials, hotline and etc. The provision of environmentally sustainable information to children and teenagers is deemed as one of the primary goals of NIC.

Intellectual system has the features of customization. The customization is not carried out at the highest level. However, as the system transitions from the highest level to lower levels, customization can be carried out and the information harming national interests and etc. can be prevented. Corporate policy and corresponding sanitization policy is carried out at low levels. When a child or a teenagers utilizes Internet individually, the latter becomes customized, that is, the filtering is related to a specific child (by considering the name, surname, biometric features and etc. of a child). This intellectual system improves gradually, adopts high-level features and “becomes” smart as the time passes. That is, by controlling the actions of each individual, the system is capable to classify the information on websites accessed by a child in a specific timeframe and her attitude towards the information by analyzing the log files. It divides children into groups conditional in their activities, classifies them and marks these groups with different colors. The intellectual system classifies children based on their individual information as below:

- Green color – a child accesses websites corresponding to his/her age (well-behaving, smart etc.);
- Yellow – he/she is prone to network games (naughty, etc.);
- Red – he/she is active in network and accesses websites not corresponding to his/her age (games, etc.)

The intellectual system keeps the contact of an individual with Internet under control and evaluates the information of a child or a teenager based on statistical information collected and their behavior. In this vein, it makes decision by determining his/her classification group. The following can be adhered to such decisions:

1. Analyzing the online behavior of a child in a specific time interval and determining his/her category of classification;
2. Notifying a parent, teacher and other responsible persones regarding the online behavior of a child (for instance, the transition of a child from green color to yellow color category and etc.);
3. Determining the classification group of a child based on medical features and establishing contact with a doctors on this basis;
4. Notifying school psychologist regarding a child’s...
behavior in accordance with minor’s problems.

The analysis of web-pages for the purpose of children’s safety is among the important issues to be solved. The mentioned system envisages the utilization of text mining, clustering, classification methods. It is important to establish expert systems by considering the biometric features of children.

The architecture of the intellectual system considered is based on the client-server technology and constructed with hierarchic order. While transitioning to a lower level, the next level is considered to be a client if the previous level was server. The process continues till the last connection. Eventually, coordinated information exchange is carried out among all databases.

VII. CONCLUSION

The research has revealed that, the Internet in modern world is a global network consisting of not only interesting and important information, but also harmful and dangerous web-pages. There exists a need for special methods and tools for searching, detecting and obtaining harmless web-pages for children in existing information chaos.

The application of the proposed conceptual model of traffic sanitization is paramount for the protection of children against harmful information in the environment of e-state. The proposed conceptual model can be used during the solution of problems emerged while providing the safety in Internet environment, education system etc. by considering the age group of children.

The traffic sanitization is estimated to result in the formation of the secure Internet environment. This means the protection of children and adolescents from the information space. Protection of children in the Internet is the duty of not only the parents. This issue is part of every state’s information security policy. The protection of children and adolescents from harmful information in virtual space is one of the key issues in the world of terrorism, political tension, and network crime. Therefore, the proposed intellectual system is important for the physical and psychological health of the future generations.

REFERENCES

[11] Best Parental Control Software Featured in This Roundup: online: https://www.pcmag.com/article2/0,2817,2346997,00.asp
[16] Clean router. online: https://cleanrouter.com/features/

using code, 2010, analysis, Online: https://www.google.com/patents/US20130124684


[28] Perevozchikova M.S., Sapegin A.H. Control methods over the access of pupils to computer resources // “Komm”, 2010, № 10, pp. 56-60.


Authors’ Profiles

Rasim M. Alguliyev. He is director of the Institute of Information Technology of Azerbaijan National Academy of Sciences (ANAS) and academician-secretary of ANAS. He is full member of ANAS and full professor. He received BSc and MSc in electronic computing machines from the Azerbaijan Technical University in 1979. He received his PhD and Doctor of Science (higher degree after PhD) in Computer Science in 1995 and 2003, respectively. His research interests include: Information Security, E-government, Data Mining, Big Data, Online Social Network Analysis, Cloud Computing, Evolutionary and Swarm Computation, and Scientometrics. He is author more than 571 papers, 4 monographs, 4 patents, several books.

Sabira S. Ojagverdieva graduated from Applied Mathematics faculty of Baku State University (BSU). Since the same year, she began working Institute of Information Technology of ANAS. Her area of interest includes information security, on Internet child protection, data sanitization and Data Mining technologies. She carries out scientific research on “Ensuring children’s safety on Internet” in the field of information security.