

Rashid Q. Alakbarov<sup>1</sup>, Mammad A. Rashidov<sup>2</sup>, Tural I. Mustafayev<sup>3</sup>,  
Mahammadrasul M. Yagubov<sup>4</sup>

DOI: 10.25045/jpis.v08.i1.06

<sup>1,2,3,4</sup>Institute of Information Technology of ANAS, Baku, Azerbaijan

<sup>1</sup>rashid@iit.ab.az, <sup>2</sup>mamed.rashidov@science.az,

<sup>3</sup>tural.mustafayev@iit.ab.az <sup>4</sup>mrasul.yagub@iit.ab.az

## INTERNATIONALIZED TOP LEVEL DOMAIN NAMES: THEIR REGISTRATION AND PROBLEMS

*The article analyzes development stages and the current status of internationalized domain names (IDNs). The registration processes (IDNs) to ICANN is researched, some recommendations are made on the creation of top level domain names written in national alphabet.*

**Keywords:** IDN, ICANN, IDN ccTLDs, DNS, ARPAnet, Punycode, Unicode, ASCII.

### Introduction

Internet is the global network enabling the interaction of computers around the world and providing the exchange of information. With the emergence of this network, people got access to new great sources of information.

One of the most important advances in the development of the Internet is an elaboration of the multilingual system with the Internationalized Domain Names (IDN). The system founded by the Internet Engineering Task Force (IETF) aims at developing the requirements for the implementation of domain names through the use of symbols of the national alphabets. One of the other issues facing the group is to examine the current situation and making offers taking into account the technical and social aspects [1]. Given this important fact, domain names written in the national alphabet should be studied.

### Hierarchical structure of the domain names system

To understand the essence of the Internationalized domain names, it is worth having a brief glance at the history and structure of Domain Name System (DNS) and the Advanced Research Projects Agency Network (ARPANET). In the late 1960s, Advanced Research Department of the US Defense Ministry funded an experimental global computer network – ARPANET, which combines important research institutions of the country. The main reason for founding this network was to enable the government departments to use the costly computer resources. The number of organizations and users connected to the network was increasing day by day. Although, ARPANET made way for other networks (NSFNET, SBC, Sprint, MFS, UUNET, etc.) after a while, this growth was rapidly continuing.

In 1970, the coverage of ARPANET included a few hundred nodes. The information need for the interaction of the nodes was stored in a file. *Stanford Research Institute Network Information Center, (SRI-NIC)* was responsible for this file, which was called *Hosts.txt*. The nodes sending their data to the center via e-mail were synchronizing own specific data with a copy of this file. The changes were added to the main *hosts.txt* file twice a week, and the rapid growth of the structure, in turn, gave rise to the following problems [2]:

- the size of the file was rapidly increase with the expansion of the network, and this growth complicated the file to be handled;
- information exchange flow increased rather rapidly. Because, each new record had to be delivered to all nodes and periodically synchronized;
- addition of the same names to the system caused conflicts;
- any two nodes in *hosts.txt* file did not have the same name in any case;
- the data coming from the nodes with the same name completely stopped the operation of the system.

Accordingly, the structure created by ARPANET was morally outdated and the need for the establishment of a new system emerged. In 1983, researcher of the University of California (USA) Paul Mockapetris designed the Internet Domain Names System and developed *DNS*, which is currently being used, together with researcher of the Information Sciences Institute (USA) Jon Postel.

From January 1985, the registration of common domain names started, and in the same year for the first time, the top-level country codes in Latin were defined for the United States (.us), United Kingdom (.uk) and Israel (.il). In November 2000, sponsored the top-level domains (sTLD) (.mobile, .travel etc.) and in November 2009, country codes written in the national alphabet (Bulgaria (.bg-бг) and Georgia (.ge - გე) emerged. Since 2012, the *Internet Corporation for assigned Names and Numbers (ICANN)* has decided to create domain zones without limits, and permitted to establish domain names for geographical areas (lat, .arab), cities (.berlin, .paris, .moscow), brands (.hp, .canon) and topical themes (.sport, .music, .car etc.) [3].

The names on the Internet are of great importance. The name of the web-site, its domain zone and other conditions are of the factors affecting the efficiency of the Internet activity. The name of the website is written in the address window of the Internet browser.

For example, registering the top-level domain names .baku, .ganja, and .sumgait for big cities as Baku, Ganja, Sumgait and others is one of very important steps. This can motivate the population of these regions to access the virtual world. Furthermore, unofficial competition would occur among regional countries in this regard. This step may also prevent the problems that exist in the registration of domain names in Azerbaijan.

*DNS* is a distributed database. Its structure is hierarchical. This structure allows locally controlling separate sections of the network and at the same time, these segments become accessible over the network due to the use of client-server mechanism. In order to ensure reliability and efficiency for the system, *DNS* replication and caching is used. So-called *DNS Server* applications are functioning in the server side of the client-server mechanism. These servers contain the data about the domain names and the Internet Protocol (IP) address of all nodes in the segment. The client function is activated in all machines, and prepares and sends *DNS* requests to the server and delivers the obtained information to the user. The data is hierarchically stored in the *DNS* servers, and each part of the hierarchy is a subpart of the previous one (Figure 1) [4]. Each domain is the root of subsequently generated subdomain. A symbolic dot (“.”) is put in the highest part.

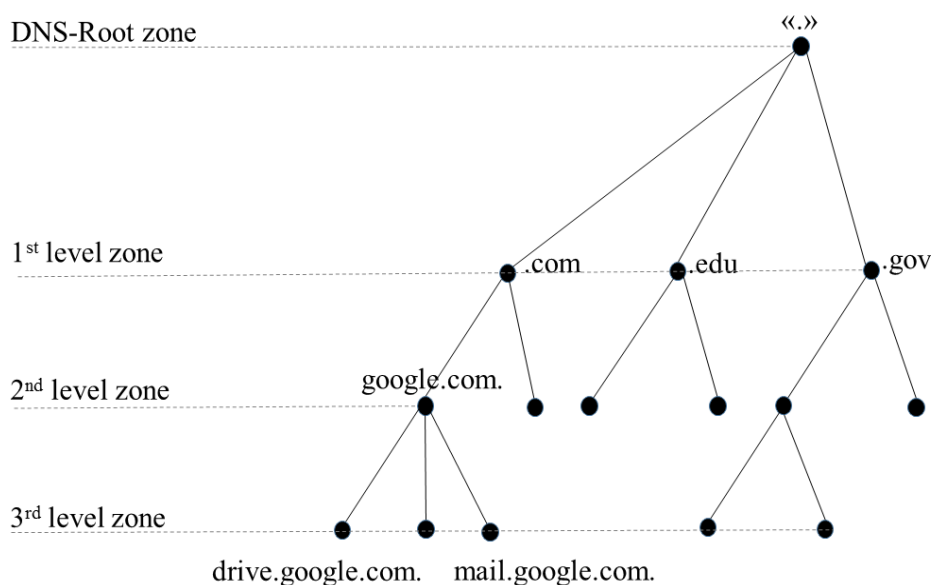


Figure 1. Hierarchical *DNS* system

The subsequent domains after the DNS root zone are called 1st (top) level domains. In the early use of *DNS*, 7 Generic Top Level Domains (*gTLD*) were created:

- *.com*- for the registration of commercial entities (e.g., *Hewlett-Packard*, *hp.com*);
- *.edu* - for the educational institutions (e.g., Harvard University, *harvard.edu*);
- *.gov* - for the domain names of governmental organizations (e.g., *NASA*, *nasa.gov*);
- *.mil*- for the domain names of military organizations (e.g., the United States Navy, *navy.mil*);
- *.net* - formerly used by the organizations to ensure the functioning of the network infrastructure (such as *NSFNET* (*nsf.net*) and *UUNET* (*uu.net*) like). Since 1996, it has been accessible for all commercial organizations as *.com* domain;
- *.org* - formerly used by non-profit organizations (e.g., *Electronic Frontier Foundation*, *eff.org*). since 1996, it has been accessible for all commercial organizations in *.net* domain;
- *.int* - for the registration of domain names of international organizations (e.g., *NATO*, *nato.int*).

Each domain or directory can be divided into smaller parts, which are called subdomains. The control of subdomains in DNS can be delegated to another institution. All domains and subdomains are recorded in the indexed for in the database of domain names. Thus, the domains at root zone or any other top-level unifies all the paths going to subdomains [5]. A new zone is generated together with each subdomain (Figure 2).

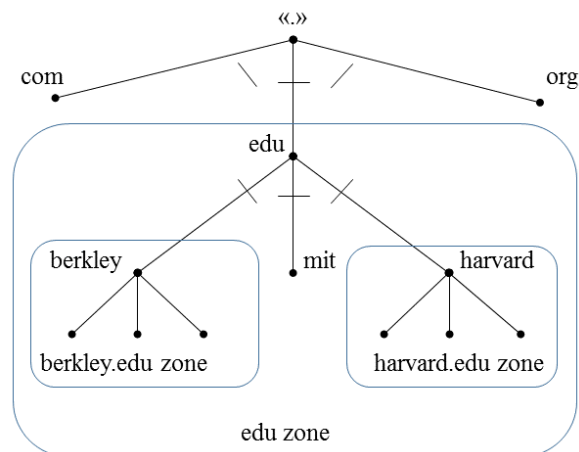


Figure 2. Hierarchy of subdomain zones

Delegating the control of the subdomains directly to the organizations themselves ensures the implementation of this mission a more convenient way. Hence, the organization controlling the subdomains is taking all responsibility.

Following this principle is very important for Azerbaijan too. The structure generated in this way facilitates the control and the distribution of the responsibility and leads to a more understandable structure. For example, we place an infrastructure of ANAs in the virtual space as a domain and subdomains (Figure 3):

*.az* - top-level domain for Azerbaijan.

*science.az* – an official web page of ANA registered in *.az* domain. The control of this domain and its subdomains to be developed is implemented by the administration of the Institute of Information Technology of ANAS.

*physics.science.az*– subdomain for the Institute of Physics of ANAS. This subdomain and the subdomains to be generated in the future are coordinated by administration of the Institute of Physics.

*nuclear.physics.science.az*- subdomain for the Laboratory of nuclear studies to be established at the Institute of Physics.

*uranium.nuclear.physics.science.az*- subdomain to be developed to store the data about the uranium. Obviously, if necessary, the administration of this subdomain may be assigned to any applicable department.

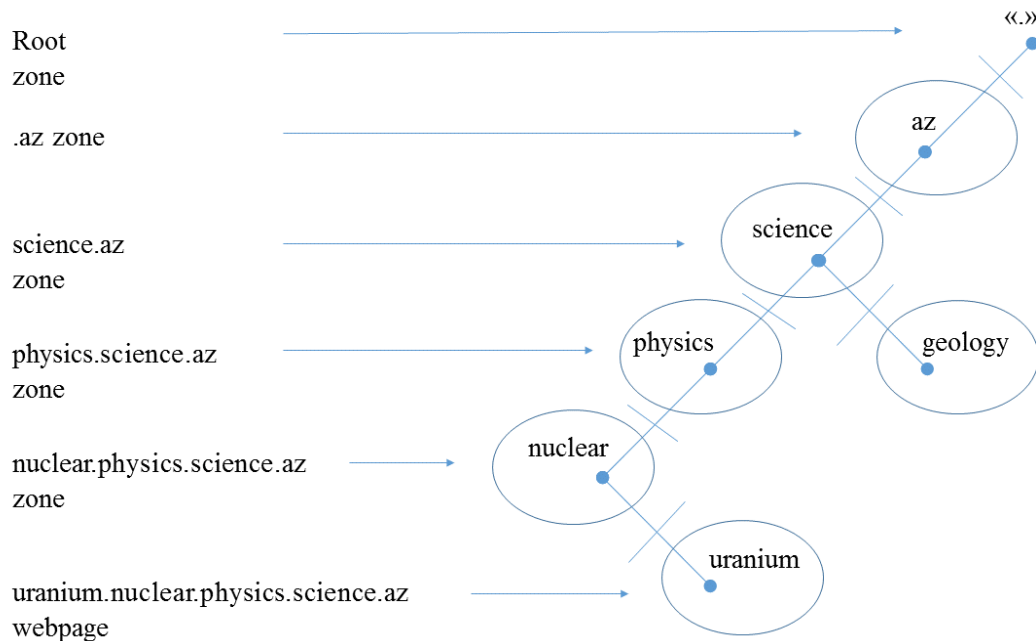


Figure 3. Subdomains of the domain *science.az*

### The development of the internationalized domain names

Globalization of the Internet and the diversity of the user profiles urge the domain names to be written in the national alphabet. Domain names written in the national alphabet are the domain names partially or completely written in the Arabic, Cyrillic, Chinese and others scripts, or in the Latin with different characters as the French, German, Azerbaijani and others. The DNS servers, which are used for converting the domain names into the network addresses in order to make the Internet resources accessible, have to use the *American Standard Code for Information Exchange* (*ASCII*), and this practical limitation has formed the initial standard for acceptable domain names. According to this standard, the number of *ASCII* characters is 37, and there are uppercase and lowercase Latin letters, figures from 0 to 9 and the symbol “-“ [6]. Nationalization of the domain names is the technical solution converting the domain names written in the national languages into *ASCII* text format to be registered with the *DNS* servers. The domain names written in the national alphabet are used only by the software packages and web-browsers capable to work with them and do not require any changes of the Internet infrastructure.

In fact, the Internet has rapidly spread all over the world going beyond the United States, and accordingly, the number of users not knowing the Latin characters began to increase. Due to the fact that the domain names can only be written in Latin letters, there was a serious obstacle to the use of the Internet resources. Obviously, the time came for a major change in the system of domain names. Thus, in December 1996, IDN was proposed by Martin Dürst and implemented in 1998 by Tan Juay Kwang and Leong Kok Yong with the supervision of Tan Tin Wee. After long discussions and protests, the Internationalizing Domain Names in Applications (IDNA) system was adopted as a standard and implemented for a number of top-level domain names [7].

The concept of internationalized domain name in *IDNA* was domain names containing the characters, “*IDNA To ASCII*” algorithm for which could be particularly implemented. In 1999, the *Internet Engineering Task Force* (*IETF*) developed a new *IDNA* group to replace the current IDNA protocol. In March 2003, *CANN* approved the development of the internationalized domain names

in the Internet based on *IDNA* standard. In June of the same year, *country code Top level Domains (ccTLDs)* were installed in *DNS* root zone.

*IDNA* was adopted standard as a mechanism in 2003 for the use of internationalized domain names containing *Unicode* characters. Under this mechanism, domain names were written in Latin characters together with diacritics(ñ, ä orü), or in non-Latin characters (e.g., Arabic, Chinese and Hangul). *ICANN* created a policy for the use of *IDNA* in 2003 on the basis of this system and in the same year registered a domain name *.jp* and some domain names in the subdomains of *.info* in March 2004. Developed policy was improved in November 2004 and in October 2009, *ICANN* adopted an order for the creation of top-level domains in native language scripts. Thus, for the first time, Egypt, Saudi Arabia and Russia registered their country code domains in national scripts in ccTLD DNS servers root zone.

Currently, 22 out of top-level domain names placed in the top-level domain zone and widely used are written in the alphabet used in different countries (Figure 4).

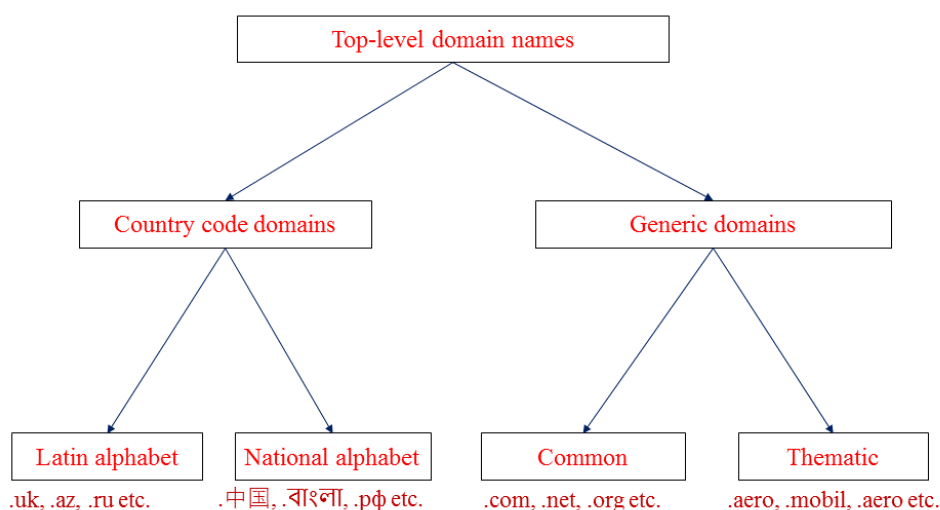


Figure 4. Distribution of top-level domain names

Although domain names support *Unicode* characters, the client programs, such as e-mail and Web-browsers, were limiting the use of these domain names. The problem was not associated with the *DNS* server and software packages, but with the network protocols used by the software packages. Because, the network protocols hosted only *ASCII* characters for domain names. To overcome these restrictions of the existing base, *IETF's IDNA* working group decided to represent the domain names in proper *ASCII* characters in order to provide the use of *IDN* in the web-browser and other software tools. *IDNA* defines how the domain names are written in *Unicode* characters and how they are represented in *ASCII*-based characters. *IDNA*-enabled applications support writing the domain names with *Unicode* characters and can use other *IDN* and *ASCII* versions. However, the applications that do not support the *IDNA* standard only require the domain names to be written in *ASCII* characters. Although all the characters are written in *Unicode* standard on computers, they are stored in *IDN DNS* servers based on *ASCII* standard. *Punycode* method (standardized representation method of *Unicode* character set) is used to perform the representation between the two standards. *Mozilla 1.4*, *Opera 7.11* and *Netspace 7.1* are the early software packages supporting *IDNA* standard. Other application software packages have additional *plugin* to support *IDN*.

*Punycode* is a way to represent *Unicode* with the limited character subset of *ASCII* supported by the *DNS* [7]. For example, “München” (German language spelling of Munich) is encoded as “xn - mnchen - 3YA” in *ASCII* characters through *Punycode*. This encoding syntax is defined in the document *RFC 3492* of *IETF*. The *IDNA* methodology encodes only select label components of



domain names in *ASCII* characters and in *Unicode* labels through *To ASCII* procedure (Figure 5). *Punycode* performs the encoding procedure in both directions.

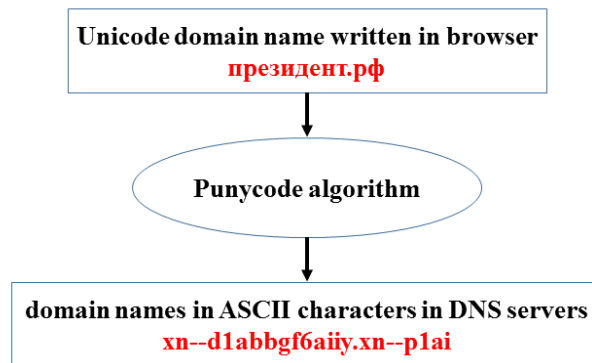


Figure 5. Working principle of *Punycode* algorithm

Currently, to purchase the domain names using the alphabet in top-level domain system *.az* is one of the most pressing issues. Hence, the important steps should be taken by the relevant government agencies for the creation of top-level domain names written in national alphabet and the process must meet international regulations and national legislation.

### Registration of top-level domain names in *ICANN*

As we know, at the moment, legal entities and individuals can apply for the registration of second-level domain names. However, only legal entities can apply for the registration of top-level domain names. In addition, *ICANN* simply requires using the non-Latin alphabet for the registration of top-level domain names written in the national alphabet. That is, none of the characters of the presented top-level domain name cannot be *ASCII* character. For example, if top-level domain name *.baki* is *ordered*, it will not be adopted. Because, the letters *a*, *b* and *c* of the domain name are determined by encoded by *ASCII* characters. The organizations, which cannot order top-level domain name written in national alphabet, can write the domain name in the national alphabetical in the subdomain of existing top-level country code domain. For instance, the domain name *məqalə.ict.az* is encoded by the *Punycode* algorithm and registered in DNS servers as *xn-mqal-v6bd.ict.az* [8].

The purchase of a new top-level domain name is more complicated process. It is important just to comply with the requirements of the registrar accredited by *ICANN* and pay fees. The candidate for the registration of a new *gTLD* domain, in fact, is responsible for the creation of a new registry and for the control of a part of the Internet infrastructure. Once *ICANN* announce the registration period the candidates (only legal entities) for the purchase of a new top-level domain name must be registered at *Application System (TLD Application System - TAS)*. Thus, he/she must pay a deposit of 5,000 USD. The application can be filled only after the payment made. In addition to the domain name, the applicant should mention general information about himself/herself and respond to the questions about his/her technical, organizational and financial resources in details. Meanwhile, the fee (180,000 USD) required by *ICANN* must be paid in order to start the evaluation process.

Furthermore, the explanation of this name in English must be indicated in the application form when purchasing internationalized domain name. After the submission period, *ICANN* publishes the orders ready for the evaluation with the data about the candidates within two weeks. The orders go through several stages as follows for the final decision (Figure 6) [9]:

- order submission period –a new top-level domain order must be prepared and submitted in written form before the deadline for the submission of the applications for top-level domain defined by ICANN;
- administrative check of completeness - complete verification of all information by ICANN based on submitted orders;
- early assessment - early assessment of the top-level domain request;
- extensive assessment –when the data about the early assessment is insufficient to decide for or against the applicant, additional research are carried out and the orders are evaluated more broadly;
- dispute resolution – any dispute about domain name or any other issue is resolved at this stage;
- submission of objections – receiving objections of individuals and entities related to the domains;
- competitiveness – once the domain name is submitted to ICANN this organization announces the domain name to the general public, and any legal entities wishing to purchase the domain name may submit a request. At this stage, a matter of competition between one or more legal entities intending to purchase the same domain name shall be resolved [10].

In absence of objections, the assessment process can last 9 months. Otherwise, if the early assessment fails and it is necessary to resolve disagreements, the process can take up to 20 months. After successful completion of all phases of the evaluation process, the candidates sign an agreement with *ICANN* and new domain is undergone the technical inspection before being added to *DNS* root zone. The agreement also includes a membership fee of 6250 USD each quarter [11].

Purchasing top-level domain names for Azerbaijan is recommended. However, the Azerbaijani alphabet contains Latin scripts, originally Azerbaijani letters (e.g., ğ, ə, ü etc.) are also available. However, *.baki*, *.xəzər*, *.şəki* and others domain names can be obtained.

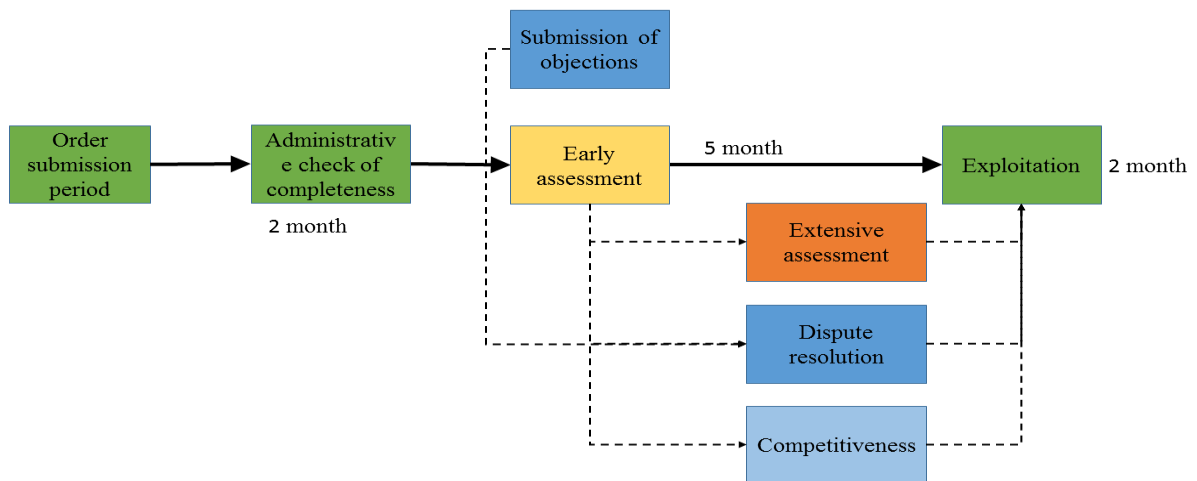


Figure 6. The algorithm for the implementation of documents for ICANN internationalized domain names

## Conclusion

In recent years, the Internet environment of Azerbaijan experiences qualitative and quantitative changes. Now, not only separate institutions, but also a large number of individuals are developing their personal sites and pages.

As mentioned, the process of registering domain names of different countries in national alphabets has been performed in the international community for a long time. Since 2009, the registration of the names of places and domain endings of these ending containing no more than 37 characters from-a to -z, including figures 0-9 and hyphen (e.g. *\*asia\** *baku*, *\*bank* etc.) is allowed. Although, this change is not so exciting for international companies, it is of great importance for the local authorities.

Therefore, relevant government agencies have to take the initial steps for the purchase of new names, such as *\*.baki*, *\*.naxçıvan*, *\*.şəki*, *\*.məktəb*, etc. The introduction of such new names provides the classification of local authorities in the virtual system, furthermore, facilitates the use of the Internet by local users who do not know foreign language.

## References

1. Gasimova R.T. Domain problems and solutions in the Internet, Baku, Information Technology Publishing House, 2012, 164 s.
2. Postel J. Computer Mail Meeting Notes, RFC 805, Information Sciences Institute, USA, 1982, 6 s.
3. Kasanova M.B. The Internet Corporation for Assigned Names and Numbers in the Internet Management Facility // News of economics, law and sociology, 2012, No 4, pp. 164-168.
4. Albitz P., Cricket L. DNS and BIND 5th Edition, United Kingdom: O'Reilly Media, 2006, pp. 640.
5. <https://www.icann.org/en/system/files/files/iana-functions-18dec15-en.pdf>
6. Rosenbaum R. Using the Domain Name System To Store Arbitrary String Attributes // Digital Equipment Corporation, USA, 1993, <https://webcache.googleusercontent.com>
7. Faltstrom P., Hoffman P., Costello A. "Internationalizing Domain Names in Applications (IDNA) / RFC 3490", Cisco, IMC & VPNC, University of California, USA, 2003.
8. <https://en.wikipedia.org/wiki/Punycode>
9. <https://www.newgtlds.icann.org/en/applicants/agb>
10. Gasimova R.T., Comparative analysis of the geographical top-level domains on the Internet // Problems of Information Technology, 2011, No7, pp.18-23.
11. Alguliyev R.M., Gasimova R.T. // Development of data mining system of domain names, Problems of Information Technology, 2011, pp. 29-36.