

UDC 004.652.4:930.25

Y.V. Prokop¹, PhD,
O.G. Trofymenko², PhD, Associate Professor,
S.M. Voronoy¹, PhD, Associate Professor¹O.S. Popov Odessa National Academy of Telecommunications, Odessa, Ukraine
yulia13.prokop@gmail.com

voronoy@onat.edu.ua

²National University "Odessa Law Academy", Odessa, Ukraine
egt@ukr.net

Structure of the geographical block for archival information retrieval systems

There is a growing public demand for online archival information retrieval systems that would have an effective means of quickly identifying documents and finding information on specific topics, in particular, for the selection of archival files by geographical feature. A hierarchical structure of the geographical block for a unified archival information retrieval system has been developed. The proposed structure has a clear hierarchy and allows a search not only by names of settlements but also by districts, regions/provinces, and all subordinate administrative units since the historical subordinations for settlements are defined. Compared to the available search implementations in online archival databases, the structure of the geographical block developed in this work does not limit its scope for searching archival data either by the territorial or thematic orientation of the database. By establishing administrative links between the district/uyezd, regions/provinces, and settlements that have been subordinate to them over a historical period, it enables archival documents to be found in not only large cities and towns but also small villages. The scientific novelty of the results is a comprehensive solution to the problems of searching archival data by a geographical filter, which will allow archivists to optimize the accounting of documents, and users to retrieve the information effectively when accessing archival institutions online. Unlike previously known archival Internet database implementations on the websites of state archival institutions and international private companies, the proposed hierarchical structure of the geographical block for archival information retrieval systems takes into account numerous aspects of geographical marking of archival documents, which allows providing a wide range of geographical filters requirements. The use of the proposed structure of the geographic block in unified national archival information retrieval systems will enable not only to select data in a particular database of a specific local archive but also to provide interaction between archives and ability to search archival files in all archives attached to the corresponding unified search engine at once.

Keywords: *information retrieval system, archival database, geographic search, archival documents marking, archive data search.*

DOI: 10.31474/1996-1588-2020-1-30-90-100

Introduction

Nowadays, there is a significant increase in interest in historical, local lore, and genealogical research, and there is an increasing public demand for access to primary sources stored in archives and libraries. During the rapid development of technology and informatization of society, one of the criteria for the effectiveness of research is the speed of the search for information on specific topics.

Many developed countries now have archival information retrieval systems (IRS) (or are in the process of being formed them). They provide various levels of search capability, allowing researchers to select documents based on one or other selection criteria. The most common filters offered by information retrieval systems are keyword search in files and fund names, by period, and by the topic group. Some

archives provide remote access to thematic databases where users can search for surnames, which is indispensable in genealogical and biographical research.

Some countries have implemented nationwide archival IRSs, which accumulate data from regional archives and provide general search. For example, the Polish system <https://szukajwarchiwach.pl> stores information from many local archives related to various settlements in the country. For such a level of modern archival IRS, as for any advanced archival information systems, the criterion for the selection of documents on a geographical basis is relevant. However, despite its importance, most archives offer to search for files only by a keyword, which may be, in particular, the name of the settlement. In this case, it is difficult and even mostly impossible to take into account aspects of different names of settlements in various historical periods – and it is difficult to

predict what variant was used in the name of a particular archival file. A similar problem may arise because of the unpredictability of the language in which the header of the file was entered into the database. The search becomes more problematic because different regions often have towns with the same names – and researchers are not always able to separate correct entries in the search results. Equally important is the need to select information not in a particular settlement, but in uyezd, district, province, and all settlements that have been subordinated to them in a specific historical period. However, without the administrative link between the province, uyezd/district, and the relevant settlements, such a search cannot be performed.

The object of study is the archival information retrieval system, which provides a search capability, allowing researchers to select archival documents based on some selection criteria.

The subject of study is the structure of the geographical block of archival IRS.

The purpose of the work is to develop a clear hierarchical structure of the geographical block for archival information retrieval systems, which takes into account the regional peculiarities of the country structure at different historical periods.

To improve the search results in archival IRS, it is appropriate to include special geographical blocks in their composition, taking into account the hierarchical structure of modern and historical administrative and territorial units of the country, subordination, and all historical names of settlements. For this, it is necessary to investigate the practical implementation of searching files and information by a geographical feature on the websites of state archives and international private companies, and to analyze their specificity. This will allow to develop the optimal structure of a geographic block of a general archival IRS, which would fully satisfy the users, taking into account geographical features of the country's administrative structure at various historical periods.

Related Works

Researchers [1–4] have devoted their work to informatization of archival affairs in Ukraine and the post-Soviet space. However, their research is generalized, and their works do not refer to the possibility of geographical search in archival information systems at the level of storage units (files).

Modern researchers in other countries pay great attention to scanning archival documents, providing remote access to copies, and problems of indexing documents at the level of their content. For example, [5] proposed a new approach to digitizing archive records at the document content level. On the other hand, researchers are focusing on user needs and creating a simple and user-friendly interface [6]. An example of a database based on user needs

analysis is given in [7]. Article [8] emphasizes the expediency of studying the approach of historians to the search for archival documents: what criteria and filters they require and commonly use. The paper argues that using these search criteria, one can create convenient tools with quick access to the requested information. According to the authors of the work, the search for keywords is not enough for historians and structured metadata is required to produce effective results. Still, the study does not refer to the possible geotagging of archival data.

The study [9] especially refers to the handling of architecture-related archival sources. Its authors propose to interlink the archival sources with objects outside of the archive by using the method of geotagging. They suggested a new data model, which is based on a reference-plane system.

In turn, IT professionals point out the advisability of geotagging data even when browsing. For example, [10] states that everything we do takes place in a geographical context, and many requests for information on the Web have a geographic focus (it has been estimated that 13–15% of Web queries submitted to traditional search engines contain a place name). The authors point to common mistakes that can affect the results of a geographical search. They include the inability to distinguish between different instances of the same name, a lack of methods to rank and explore effects concerning their geographic relevance and the non-retrieval of resources, which are geographically relevant but use a place name different from that specified in the query, and so on. This research focuses on the search for data in the information resources on the Web, which have their specifics and differences from the search criteria in archival databases.

Researchers at Historical Geographic Information Systems (GIS) have identified the need for geotagging data and are looking for optimal ways to do so. Article [11] demonstrates the feasibility of using a geographical methodology and the study of spatial relationships for the reconstruction and reinterpretation of the human past. Work [12] shows the importance of geographical marking of information. The following geographical block structure is proposed: Place name (for example, "Odesa") – Location category (for example, "city") – Official geographical reference (for instance, longitude and latitude coordinates). It is claimed that these three elements establish a correspondence between informal geolocation (using place names and place categories) and formal (longitude and latitude). These elements, plus transparent relationships between geographical locations and time boundaries for places and their characteristics, are critical components of digital directories. At the same time, the study does not deal with the possibility of using the specified structure for archival information systems. The paper [13] proposes to use local (non-

official) names to describe the geography, as well as vague, non-administrative names.

Integration of geoinformation resources and free access to them using specialized infrastructures of spatial data, geo-portals, databases of shared use and web GIS are considered in work [14]. Article [15] explores the possibility of using the geoinformation system for archival documentation of an architectural project. The creation of a geographic database for the census is discussed in [16]. It proposes the following multi-level hierarchy: Country – Province – District – Subarea – City (rural or urban). The authors of the article [17] demonstrate the geographical marking and visualization of historical censuses of England and Wales by comparing historical directories and a modern address base.

Article [18] describes the geographic database that was designed based on the censuses of the South African population repeatedly conducted in the country from 1904 to 2001. It takes into account the change of settlements' names (including variations of transcription), as well as the names of the cities that have disappeared over time. The names of settlements are given in two languages (English and Afrikaans). The proposed database is designed to investigate historical demographics and urban dynamics rather than to store data from archival institutions and does not take into account certain specificities. Nevertheless, the experience of the creation of a database that can track historical changes over nearly a century is useful in developing a unified archival IRS, even given the rather specific administrative and territorial structure of South Africa.

Many works are devoted to the analysis of available Russian thematic archival databases. Some of them have geographic names within the structure fields. Still, most studies do not use the hierarchical structure of geography. In particular, the article [19] describes the characteristics of the database "Siberian and Far Eastern Book. Eighteenth Century – 1930: Consolidated Catalog", which includes, but is not limited to, a geographical reference to the publications in the database. The paper [20] analyzes local databases introduced in the State Archives of the Vologda Region. The possibility of the geographical search is indicated, and the following hierarchy of administrative subordination is proposed (for example, in 1782): Country – Governorship – Province – Uyezd – Volost – Settlement. However, this database has a narrow theme (dedicated to revisions) and a time limit from 1743 to the middle of the XIX century. It also has no web version. Article [21] proposes a methodological approach to cataloging information resources on the cultural, historical, and social development of small indigenous peoples of the Khanty-Mansiysk Autonomous Okrug – Yugra. A geographical area is among the required fields of the structure.

Thus, the analysis of studies on the selected topic showed that the problem of geotagging of data

for search by geographical filters in the development of a unified archival IRS is relevant. Researchers are exploring the use of geoinformation resources and designing geographic blocks of databases. However, their designs are mostly related to information systems in a narrow subject area: population censuses, digital directories, etc. Most surveys use geographic data at a particular historical moment and do not take into account changes in administrative subordination over a long historical period.

Material and methods

To analyze the specifics of the practical implementation of public archival Internet databases, the authors investigated the sites of archival institutions and private companies of various countries: Unified information retrieval system of Polish archives (<https://szukajwarchiwach.pl>), Electronic Catalog of the National Archives of Great Britain (<http://www.nationalarchives.gov.uk>), the Archival Information System of the National Archives of Estonia and the City of Tallinn (<http://ais.ra.ee>) and others.

Currently, only a few archives declare geographic search capability. Let us look at some examples of such search engines.

The site of the Russian State Historical Archive (<https://rgia.ru>) has a geographical index; it is possible to search for files by geographical name. Filling in the index with data on administrative units is likely to occur at the same time as entering and marking files without adhering to a clear structure. In particular, as of 01/13/2020, the Kherson uyezd of the Kherson province contained in the index only one subordinated township – the village Maximivka. Some settlements are directly subordinated to the province, not the district (uezd) as it should be. This makes the search for the district meaningless. Similar geographical indices have also been introduced at some other sites, such as the "Electronic catalog of state archives of the Moscow region" (<https://arch.mosreg.ru>) and "Information resources of the Pskov region" (<http://document.archive.pskov.ru>). It seems that they do not have a clear geographic block structure, which results in ineffective geographical filtering.

The only Ukrainian archive, the State Archives of the Kyiv Region (<http://dako.gov.ua>) introduced an information system with a selection of files by geography. However, a complete search (by city, township, and uyezd) works in the electronic catalog (<http://dako.folium.info>) only for the 1897 Census. The civil registry books can only be searched by locality. The general geographic catalog in this archive has the following structure: record number, heading of the geographical object (province), a subheading of the object (district/uezd), name of the settlement, document name, date of the document, fund number, fund name, inventory number, file number, sheet number. Studying the response of the

system to the search queries leads to the conclusion that the geographical information of each file is stored in a textual form in one table with the data on the documents. Consequently, standardized table normalization is not respected in this database, which results in poor search performance that cannot be considered to be a fully satisfactory result of the system's operation.

Today, unfortunately, the possibility of selection data by geography, providing by state archival institutions, does not meet the full needs of users. IRS private developers much more moved forward.

Miriam Weiner's website "Routes to Roots Foundation" (<http://www.rtrfoundation.org>) provides an opportunity to search for files containing basic genealogical information (vital books, censuses, etc.) by the name of Eastern European cities (Poland, Lithuania, Ukraine, Moldova, and Belarus). The database is based on modern names of settlements and administrative subordination only, search by old names or by districts/provinces is impossible. The famous American genealogical corporate project, Ancestry.com, preserves the historical names of the cities and only their current subordination: City – Region – Country. There are no villages in the database – only cities and towns.

The "All Galicia" database of the Geshet Galicia Internet Project (geshergalicia.org) has been dating since the late 18th century to the Second World War, a limited historical region – Galicia, consisting of part of modern Western Ukraine and southeastern Poland. The geographical block of the database contains a list of settlements directly subordinated to the country of their present location. Name variations (for example, Lviv/Lemberg) are provided. Distances between settlements are specified. The search block of the database does not allow the selection of data by voivodeship or uyezd/district but allows specifying the radius of distance from the selected settlement and search for all settlements that fall within the nominally defined circle. It was this approach with a circle of a certain radius that allowed the developers of the database to bypass administrative subordination and its change after the collapse of the Austro-Hungarian Empire in 1918. Another feature of the All Galicia database is that the documents included in this database refer only to relatively large settlements – the database does not contain small villages, so the size of the geographical block is small and does not require division of the entire list of settlements by districts.

The popular powerful JewishGen online project (jewishgen.org) spans territorial documents in many countries across continents. Users can specify the name of the settlement (modern or any historical), select the country (modern) and region (province or voivodeship of pre-Soviet times). The name of the settlement and its administrative and territorial

subordination are given in this database for four historical periods (fixed moments): 1) before the First World War (1900), 2) the interwar period (1930), 3) after the Second World War (1950) and modern. Subordination of administrative-territorial units is determined by four levels: 1) settlement, 2) uyezd (in the first period) or nominal region (in the second period), 3) province (in the first period) or republic (in the second period), 4) country. In the third and fourth periods, the second and third levels are not used, as all documents entered in the database relate to the first or second periods and there is no need for further detail.

Consider the levels and periods of the JewishGen geographical block structure on the example of the modern Ukrainian city of Tartakiv in the Lviv region, as it is set in this geographical block: 1) Tartakov – Sokal district – Galicia province – Austrian Empire; 2) Tartakov – Sokal district – Lvov province (voivodeship) – Poland; 3) Tartakov – Soviet Union; 4) Tartakiv – Ukraine.

Significantly, the number of settlements in the JewishGen database is not detailed; there are no small villages, which is probably due to the specifics of the documents submitted in the system. Also, administrative and territorial subordination is rather nominal (as can be seen in the example of Kropyvnytskyi), and modern names have not yet been updated.

Family Search (familysearch.org), which is actively cooperating with Ukrainian archives, created one of the world's most significant genealogy projects. The geographic block of their search engine has a hierarchical structure of four levels: 1) country, 2) region/province, 3) district/uyezd, 4) settlement. The database contains modern names and modern subordination (for example, Shpola – Shpolyanskiy district – Cherkasy region – Ukraine) and pre-Soviet (Shpola – Zvenigorodsky uyezd – Kiev province – Russian Empire). Old names are linked to pre-Soviet subordination. In particular, Kirov and Zinovievsk are subordinate to the Kherson province of the Russian Empire. Small settlements in the database are practically not listed. There are no disappeared cities, such as Novogeorgievsk (flooded in 1961 after the construction of the Kremenchuk hydroelectric power station).

IRS "Archive Links" was created in 2009 as part of the project "Jewish Roots" (j-roots.info) of the Lithuanian company "Genealogy & History" (Y. Prokop participated in the development of the database). The aim was to reconstruct the complete history of renaming and administrative subordination of all settlements during the period from the end of the XVIII century until now for the most accurate geographical marking of archival files. A time interval is specified for each name and each subordination in the database. Together, these intervals should fully cover the period from 1795 (for some settlements – earlier) to 2018 (last updated

geographical data). The difference between the IRS "Archival Links" and other genealogical systems is the considerable detailing of settlements to the level of villages, but the list of settlements in the database is far from complete. An example of historical administrative subordination of Bobrynets city (Kirovograd region), transliterated into English, is shown in Table. 1.

The developers of the "Archive Links" database have failed to implement the goal fully, as finding out all the subordinations of small settlements was a challenge that was almost impossible to solve. Particularly difficult for this is the historical period from 1918 to 1941. The developed structure of the geographical block does not imply discretion. However, some subordinations in specific periods were not defined. So it was decided to move the boundaries of known periods and other conventions, which distorts the picture of historical realities.

Table 1 – Administrative history of Bobrynets city in the "Archive Links" database

Bobrynets (Bobrynets district -> Kirovograd region -> Ukraine)	1992–2018
Bobrynets (Bobrynets district -> Kirovograd region -> Ukrainian SSR -> USSR)	1939–1991
Bobrynets (Bobrynets district -> Nikolayev region -> Ukrainian SSR -> USSR)	1937–1938
Bobrynets (Bobrynets district -> Odessa region -> Ukrainian SSR -> USSR)	1923–1936
Bobrynets (Yelisavetgrad uyezd -> Nikolayev province -> Ukrainian republic)	1920–1922
Bobrynets (Yelisavetgrad uyezd -> Kherson province -> Ukrainian republic)	1919–1919
Bobrynets (Yelisavetgrad uyezd -> Kherson province-> Russian Empire)	1865–1918
Bobrynets (Bobrynets uyezd -> Kherson province -> Russian Empire)	1795–1864

Therefore, an analysis of the practical implementations of various geographical blocks in the available online archival databases revealed the following peculiarities. First, the considered databases have some specialization, sometimes narrowly thematic and/or territorial. Secondly, the list of settlements included in the databases is not complete. For the most part, these are only cities, towns and large villages, and the introduction of small settlements into the databases is either not foreseen at all, or they are added only if the theme of the project is appropriate. Third, the documents and information they contain mostly relate to the period from 1800 to 1944.

On the other hand, a unified archival database of Ukraine, as well as the most other modern countries, should contain a variety of documents, often related to settlements, which in recent centuries have changed their names and administrative

subordination. The time of archival documents is determined by the specifics of a particular archive, is much wider than in genealogical databases, and extends to the present time. Therefore, it is hardly possible to fully transfer the experience of one or another database to develop the structure of the unified archival IRS.

An analysis of the geographic search implementation in the mentioned above search engines identified the need to develop a geographical block structure to ensure optimal geotagging (with the possibility of further search) of any archival document that would meet these requirements:

1. The presence in the database of all historical names of settlements. For example, the search for documents about the city of Kropyvnytskyi should be conducted by all the names that the city once had (Yelisavetgrad, Zinovievsk, Kirovograd, and Kropyvnytskyi), as well as permanent folk names. For example, the agricultural colony of Zelenopol in Yekaterynoslav province had the folk name Madler, and the colony of Novovitebsk in Kherson province was also called Zholtaya.

2. A clear hierarchy. Since the search should take place both by the name of the settlement and by district/uyezd or province/oblast, the options of historical subordination for all settlements should be defined.

3. Since in reality the process of filling the database with subordinate information is non-trivial and may take a long time, it should be possible to search by the specified radius of the distance around the locality in the database.

These requirements were taken into account when developing the structure of the geographical block, shown in Fig. 1. For this purpose, the authors use the term "locality" for an administrative unit (town, village, district, uyezd, etc.).

The main table of the geographical block will be the Locality table (Fig. 1). It has four fields: locality ID (LocalityID), level in the hierarchy of subordinates (Level), and geographical coordinates of latitude (Coord1) and longitude (Coord2). Possible levels are 0 – settlement, 1 – parish (volost), united territorial community, etc., 2 – district, uyezd, 3 – oblast, province, voivodship, 4 – republic within the country (for example, the Ukrainian SSR within the USSR), or province (for example, Galicia within the Austro-Hungarian Empire), 5 – country. It is important to set the coordinates (latitude and longitude) for settlements, for other administrative units they will be nominal.

The Names table (Fig. 1) will have, in addition to the NameID key field, the name in four languages – Ukrainian (NameUKR), Russian (NameRU), English (NameEN), and another language (NameOther) with deciphering what language (LanguageOther). The Ukrainian language is selected for the application of the specified structure in

Ukraine, and for other countries, this field will be used for their state language. The expediency of the Russian-speaking name is because much of the documents of the pre-Soviet period and many documents before 1991 in the archives of Ukraine, Moldova, Lithuania, and other countries of the former

USSR were compiled in Russian. However, in some regions, the documents may also be Polish, Hungarian, Romanian, etc. with their respective names. Name fields in Russian and additional language are optional.

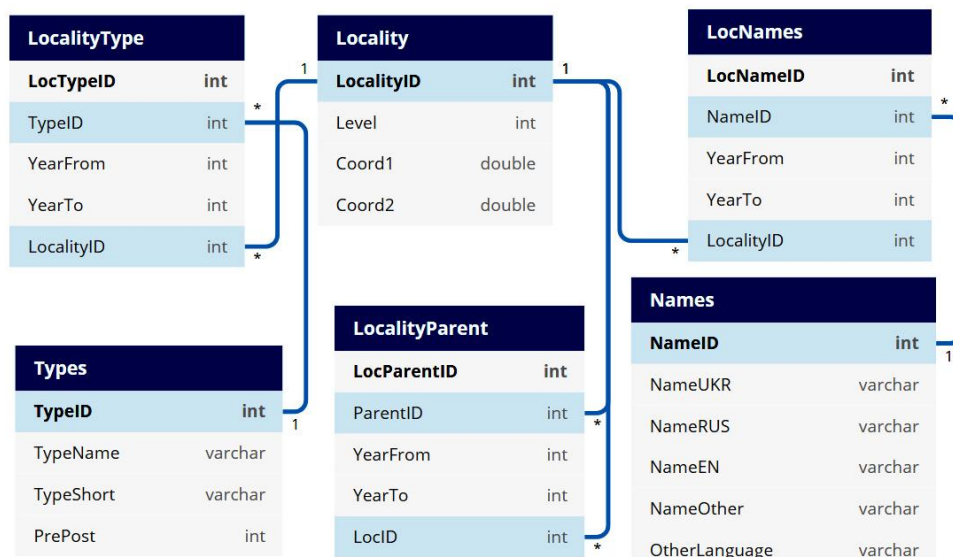


Figure 1 – The structure of the geographical block of an archival IRS

However, it should be borne in mind that the Russian name is an important component of the geographical block, so when filling the database, every effort should be made to assign Russian names to settlements in at least those regions that were part of the Russian Empire before 1917. Also, not only official names of settlements should be entered, but also folk names, as they are sometimes indicated in the documents.

It should be taken into account that the renaming refers mainly to settlements. The change of names for uyezds, districts, provinces, etc. was often accompanied by changes of administrative boundaries and the creation of new localities, so entering the database of renaming localities of levels 1-5 should be done with caution and only if there is the confidence that borders have not changed during renaming.

The LocNames table (Fig. 1) will bind all historical names to the same locality. In addition to the LocNameID key field, the table will have the following fields: NameID (Name ID from the Names table), as well as YearFrom and YearTo – name time limits.

The Types table (Fig. 1) contains information about the types of administrative units used in the database. In addition to the TypeID key field, it retains the full and abbreviated type names (TypeName and TypeShort) and the placeholder for type (0 – unused, mostly used for countries; 1 – used in front: city, village; 2 – used behind: province, district) that will be used to display the results correctly.

The LocalityType table (Fig. 1) specifies types for localities. It has a key field (LocTypeID), a locality index (LocalityID), a type index (TypeID), and the time interval during which the locality had this type (YearFrom and YearTo). Changing the type will affect the settlements (village, town, and city). For uyezd and district (in particular, Uman uyezd – Uman district), a new locality should be introduced when changing the type.

The purpose of the LocalityParent table (Fig. 1) is to establish the history of administrative records. In addition to the key field (LocParID), the table contains a locality index (LocalityID), a parent locality index (ParentID), and time limits (YearFrom and YearTo). Subordination is usually set as follows: level 0 locality is subordinate to either level 1 locality (village – parish) or level 2 locality (district or uyezd). In some cases, the level 0 locality can be subordinated to level 3-5 locality (big cities – provinces/oblasts, capitals – countries, etc.).

In all tables, YearFrom and YearTo fields are optional, but one should not leave these fields blank for good reasons, as this may cause errors when binding documents to the locality. To geotag the file (or another record) in the archival database, one must associate the file with the Locality field of the Locality table. Since the file may involve multiple localities, it is appropriate to create a separate table containing the file index and LocalityID to avoid many-to-many relationships.

Consider the data in these tables of the created structure on the example of the city of Oleshky (Kherson region). It had the types of settlement (until 1853) and the city of the Crimea (from 1854 to the present). Fig. 2 shows a diagram of the history of its renaming and administrative subordination. According to Fig. 2, the history of subordination on the scheme does not cover the entire existence of the city. However, this information is sufficient to mark archival files and other records in the database. To understand what the filled tables of such a structure might look like, we will include in them the data on the town of Oleshky and the administrative units to which it was subordinated, from Fig. 2. The Names table (Table 2 without two columns NameOther and OtherLanguage, which are empty in this example) must include all the names of the city and the administrative units to which it was subordinated.

In the Locality table (Table 3), we enter the levels of the locality indicated in Fig. 2, and latitude and longitude coordinates in radians of Oleshky.

Table 2 – Names

Name ID	NameUKR	NameRU	NameEN
14	Україна	Украина	Ukraine
15	СРСР	СССР	USSR
16	Українська	Украинская	Ukrainska
17	Російська імперія	Российская империя	Russian Empire
18	Херсонська	Херсонская	Khersonska
19	Таврійська	Таврическая	Tavriyska
20	Дніпровський	Днепровский	Dniprovskiy
21	Олешківський	Алешковский	Oleshkivskiy
22	Цюрупинський	Цюрупинский	Tsuryupynskiy
23	Олешки	Алешки	Oleshky
24	Цюрупинськ	Цюрупинск	Tsuryupynsk
25	Херсонський	Херсонский	Khersonskiy
26	Одеська	Одесская	Odeska

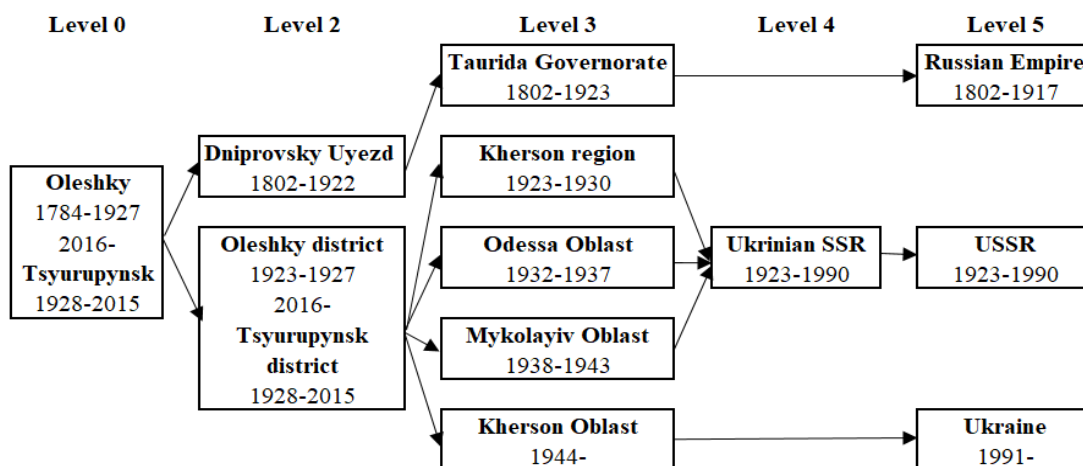


Figure 2 – History of renaming and administrative subordinations of the city of Oleshky

Table 3 – Locality

LocID	Level	Coord1	Coord2
9	5		
10	5		
11	4		
12	3		
13	3		
14	2		
15	2		
16	0	0,8303	1,704
17	2		
18	3		
19	5		

The LocNames table (Table 4) contains localities (LoclityID from the Locality table) and names (NameID from the Names table) and the time

limits when the names existed. A value of 0 in the YearTo field means that the name still exists. The Types table (Table 5) is filled with information about the types of possible administrative units. The LocalityType table (Table 6) contains the accordance between Locality (LocalityID) and Type (TypeID) with the appropriate time limits. In the LocalityParent table (Table 7), we set the subordination of one locality (LocalityID) to another (ParentID), specifying the appropriate time limits. After entering the data in tables 2-7, locality 16 will be ready to mark (bind) files or records in a unified archival database. In addition, the following tables identify localities 9 (Ukraine), 10 (USSR), 11 (Ukrainian SSR), 12 (Kherson region), 13 (Tavria province), 14 (Oleshkivsky/Tsyurupinsky district), 15 (Dniprovsky district), 17 (Kherson district), 18 (Odessa region), 19 (Russian empire).

Table 4 – LocNames

LocName ID	Locality ID	NameID	YearFrom	YearTo
1	16	23	1784	1927
2	16	23	2016	0
3	16	24	1928	2015
4	14	21	2016	0
6	14	22	1928	2015
7	14	21	1923	1927
8	15	20	1802	1922
9	12	18	1944	0
10	13	19	1802	1922
11	17	25	1923	1930
12	18	26	1920	1925
13	9	14	1991	0
14	10	15	1922	1990
15	11	16	1919	1990
16	19	17	1721	1917

Table 5 – Types

TypeID	TypeName	TypeShort	PrePost
1	страна		0
2	область	обл.	2
3	губернія	губ.	2
4	район	р-н	2
5	повіт	п.	2
6	місто	м.	1
7	радянська соціалістична республіка	РСР	2
8	автономна республіка	АР	2
9	воєводство	в.	2
10	край	край	2
11	містечко	м	1
12	волость	вол.	2
13	провінція	пров.	0
14	округ	окр.	2
16	село	с.	1
17	селище	с.	1

Table 6 – LocalityType

LocTypeID	LocalityID	TypeID	YearFrom	YearTo
1	9	1	1991	0
2	10	1	1922	1990
3	11	7	1919	1990
4	12	2	1944	0
5	13	3	1802	1923
6	14	1	1923	0
7	15	5	1802	1922
8	16	6	1854	0
9	17	14	1923	1930
10	18	3	1920	1925
11	19	1	1721	1917
12	16	16	1784	1853

Table 7 – LocalityParent

LocPar ID	LocID	Parent ID	Year From	YearTo
1	11	10	1922	1990
2	12	11	1944	1990
3	12	9	1991	0
4	14	12	1944	0
5	14	17	1923	1930
6	17	18	1923	1925
7	17	11	1926	1930
8	16	15	1802	1922
9	16	14	1923	0
10	15	13	1802	1922
11	13	19	1802	1917

The tables of the geographical block of the database must be filled in advance before binding files to the locality.

Results and discussion

The research made it possible to develop a hierarchical structure of the geographical block for a unified archival IRS, shown in Fig. 1. The proposed structure has a clear hierarchy and makes it possible to search not only by names of settlements but also by districts/uyezds, regions/provinces, and all subordinate administrative units due to the definition of historical subordination for settlements. That is why this structure is a significant step forward from previous research in the field of archival information systems.

Compared to the available archival search implementations in existing online archival databases, the geographical block structure developed in this work does not limit the scope of its use in databases to search archival data either by the territorial or thematic orientation of archives. It does not restrict the list of settlements that can be added to databases, and by establishing administrative links between the districts/uyezds, regions/provinces and the corresponding settlements that have been subordinated to them at a specified historical interval, allows to trace the history in archival documents, not only for large cities and towns but even small villages. The considered structure takes into account the positive achievements of private genealogical IRS and is based on an analysis of the search needs of historians and other researchers.

Conclusions

The paper proposes the structure of a geographical block for a unified archival IRS, which takes into account:

– the regional specificity of the country's structure at different historical intervals, which makes it possible to search not only for a particular settlement but also for uyezd, district;

- aspects of various names of settlements in different historical times;
- change of settlements' status and administrative subordination over a long historical period;
- level in the hierarchy of subordination and geographical coordinates of latitude and longitude for settlements, which reduces the complexity of finding the required archival documents;
- the possibility of having and searching archival documents of various historical periods, compiled in different languages, indicating in them the respective names of settlements in those languages;
- storing and processing not only official names of settlements but also folk names.

The scientific novelty of obtained results is a comprehensive solution to the problems of searching archival data by geographical feature, which will allow archivists to optimize the accounting of documents and execute requests. Unlike previously known archival Internet database implementations on

the websites of archival institutions and private companies of various countries, the proposed clear hierarchical structure of the geographical block for archival IRS takes into account numerous aspects of geographical marking of archival documents, which allows providing a wide range of search requests. The practical significance of obtained results is that the introduction of such a structure in archival IRS with the ability to search by geographical feature should significantly increase the efficiency of data selection and detection of documents, which will positively affect the effectiveness of historical, genealogical and other research. The joint filling and use of a common geographical block by archivists from different countries will increase its effectiveness. The use of the designed geographical block structure in the development of unified national archival IRS will allow collecting data not only in a separate database of a certain local archive but also to cooperate between archives and provide a search for archival files in all archives attached to the corresponding unified IRS at once.

References

1. Боряк Г.В. Десять років інформатизації архівної справи в Україні: проблеми, здобутки, перспективи. Міжвід. зб. наук. праць. 2002. Вип. 5: Архіви складова інформ. ресурсів суспільства. С. 9-18.
2. Васильченко М. М. Інформатизація архіву як соціально-комунікаційної структури. Вісник Харківської державної академії культури. 2010. Вип. 31. С. 199-208.
3. Колченко О.А., Брюханова Е.А. Основные направления информатизации архивного дела в контексте развития электронного общества. Вестн. Том. гос. ун-та. 2019. № 443. С. 114-118.
4. Копырина С.Н. История создания архивных сайтов России и тенденции развития сайтов федеральных архивов. Научный вестник. 2017. № 4(9). С. 18.
5. Colavizza G. Index-Driven Digitization and Indexation of Historical Archives. *Frontiers in Digital Humanities*. 2019. Vol. 6. P. 1-16.
6. Kilchenmann A., Laurens F., Rosenthaler L. Digitizing, Archiving and then? Ideas about the Usability of a Digital Archive. *Archiving 2019 proceedings*. 2019. P. 146-150.
7. Zhai H., Shi H., Zhai R. The Design and Implementation of Database on Library Management Information System. *Advances in Intelligent Systems Research: 2nd International Symposium on Computer, Communication, Control and Automation*. Paris: Atlantis Press, February. 2013. P. 836-836.
8. Torou E., Torou K., Torou V. Historical research in archives: user methodology and supporting tools. *International Journal on Digital Libraries*. 2010. Vol. 11. P. 25-36.
9. Breser C.J., Zedlacher S., Winkler R. The Principle of Geotagging. Cross-linking archival sources with people and the city through digital urban places. Lisbon. 2017. P. 205-221.
10. Christopher B.J, Ross S.P. Geographical information retrieval. *International Journal of Geographical Information Science*. 2008. Vol. 22 (3). P. 219-228.
11. Marti-Henneberg J. Geographical Information Systems and the Study of History. *Journal of Interdisciplinary History*. 2011. Vol. 42(1). P. 1-13.
12. Goodchild M.F., Hill L.L. Introduction to digital gazetteer research. *International Journal of Geographical Information Science*. 2008. Vol. 22, issue 10. P. 1039-1044.
13. Jones C.B., Purves R.S., Clough P.D. Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science*. 2008. Vol. 22, issue 10. P. 1045-1065.
14. Лурье И. К., Аляутдинов А.Р., Аляутдинов С. А. и др. Методы интеграции географических информационных ресурсов и обеспечения доступа к ним для решения научных и образовательных задач. *Интернет и современное общество: Труды XV Всерос. объедин. конф. Санкт-Петербург*. 2012. С. 73-78.
15. Bitelli G., Gatta G., Guccini A.-M. GIS and Geomatics for archive documentation of an architectural project: The case of the big Arc of entrance to the Vittorio Emanuele II Gallery of Milan (1877). *Journal of Cultural Heritage*. 2019. Vol. 38. P. 204-212.

16. Handbook on Geospatial Infrastructure in Support of Census Activities: Studies in Methods, United Nations, Series F, No 103, New York, 2009. 273 p.
17. Lan T., Longley P. Geo-Referencing and Mapping 1901 Census Addresses for England and Wales. *International Journal of Geo-Information*. 2019. Vol. 8(8), No 320. P. 1-14.
18. Giraut F., Vacchiani-Marcuzzo C. Territories and urbanisation in South Africa. Marseille, 2009. 80 p.
19. Лаврик О. Л., Мандринина Л. А., Рыкова В. В. Электронный сводный каталог сибирской и дальневосточной книги как исторический информационный ресурс. Библиосфера. 2018. № 2. С. 63-68.
20. Овчинникова Н.А. Архивный документ как источник формирования современных информационно-поисковых систем и электронных баз данных учреждений. Документ. Архив. Информационное общество. Материалы III Междунар. науч.-практ. конф. Москва. 2017. С. 379-385.
21. Филимонова Н.В., Хопияйнен О.А. Каталогизация информационных ресурсов о культурно-историческом и социальном развитии коренных малочисленных народов ХМАО-Югры. Вестник урovedения. 2016. № 2(25). С. 167-172.

References

1. Borjak, G.V. (2002), "Ten years of informatization of archival affairs in Ukraine: problems, achievements, prospects" ["Desyat` roki`v i`nformatizaczi`yi arkhivnoyi spravi v Ukraini` : problemi, zdobutki, perspektivi"], *vol. 5, issue "Archives are a component of information resources of society"*, pp. 9-18.
2. Vasylychenko, M.M. (2010), "Informatization of the archive as a social-communication structure" ["I`nformatizaczi`ya arkhivu yak soczi`al`no-komuni`kaczi`jnoyi strukturi"], *Visnyk of Kharkiv State Academy of Culture*, vol. 31, pp. 199-208.
3. Kolchenko, O.A., Bryukhanova, E.A. (2019), "The Main Directions of Archiving Informatization in the Context of Electronic Society Development" ["Osnovny`e napravleniya informatizaczi`i arkhivnogo dela v kontekste razvitiya e`lektronnoho obshhestva "], *Tomsk State University Journal*, vol. 443, June, pp. 114-118.
4. Kopyrina, S.N. (2017), "History of Russian Archive Websites and Trends Development of the Federal Archives sites" ["Istoriya sozdaniya arkhivny`kh sajto`v Rossii i tendenczi`i razvitiya sajto`v federal`ny`kh arkhivov "], *Scientific Bulletin*, vol. 4(9), p. 18.
5. Colavizza, G. (2019), "Index-Driven Digitization and Indexation of Historical Archives", *Frontiers in Digital Humanities*, vol. 6, pp. 1-16.
6. Kilchenmann, A., Laurens, F., Rosenthaler, L. (2019), "Digitizing, Archiving... and then? Ideas about the Usability of a Digital Archive", *In Archiving 2019 proceedings*, pp. 146-150.
7. Zhai, H., Shi, H., Zhai, R. (2013), "The Design and Implementation of Database on Library Management Information System", *In Advances in Intelligent Systems Research: 2nd International Symposium on Computer, Communication, Control and Automation*, Paris: Atlantis Press, February, pp. 836-836.
8. Torou, E., Torou, K., Torou, V. (2010), "Historical research in archives: user methodology and supporting tools", *International Journal on Digital Libraries*, vol. 11, pp. 25-36.
9. Bresler, C.J., Zedlacher, S., Winkler, R. (2017), The Principle of Geotagging. Cross-linking archival sources with people and the city through digital urban places, Lisbon, pp. 205-221.
10. Christopher, B.J., Ross, S.P. (2008), "Geographical information retrieval", *International Journal of Geographical Information Science*, vol. 22(3), pp. 219-228.
11. Marti-Henneberg, J. (2011), "Geographical Information Systems and the Study of History", *Journal of Interdisciplinary History*, vol. 42(1), pp. 1-13.
12. Goodchild, M.F., Hill, L.L. (2008), "Introduction to digital gazetteer research", *International Journal of Geographical Information Science*, vol. 22, issue 10, pp. 1039-1044.
13. Jones, C.B., Purves, R.S., Clough, P.D. (2008), "Modelling vague places with knowledge from the Web", *International Journal of Geographical Information Science*, vol. 22, issue 10, pp. 1045-1065.
14. Lourie, I.K., Alyautdinov, A.R., Alyautdinov, V.A., Osokin, S.A. (2012), "Methods of geographical information resources integration and access to them for scientific and educational purposes" ["Metody` integraczi`i geograficheskikh informaczi`onny`kh resursov i obespecheniya dostupa k nim dlya resheniya nauchny`kh i obrazovatel`ny`kh zadach"], *In Proceedings of the XV All-Russian Conference "Internet and Modern Society"*, Saint Petersburg, pp. 73-78.
15. Bitelli, G., Gatta, G., Guccini, A.-M. (2019), "GIS and Geomatics for archive documentation of an architectural project: The case of the big Arc of entrance to the Vittorio Emanuele II Gallery of Milan, by Giuseppe Mengoni (1877)", *Journal of Cultural Heritage*, vol. 38, pp. 204-212.
16. Handbook on Geospatial Infrastructure in Support of Census Activities: Studies in Methods, United Nations, Series F, No 103, New York, 2009, 273 p.

17. Lan, T., Longley, P. (2019), "Geo-Referencing and Mapping 1901 Census Addresses for England and Wales", *International Journal of Geo-Information*, vol. 8(8), No 320, pp. 1-14.
18. Giraut, F., Vacchiani-Marcuzzo, C. (2009), *Territories and urbanisation in South Africa*, Marseille, 80 p.
19. Lavrik, O.L., Mandrinina, L.A., Rykova, V.V. (2018), "Electronic union catalog of the Siberian and Far Eastern books as a historical information resource" ["E`lektronny`j svodny`j katalog sibirskoj i dal`nevostochnoj knigi kak istoricheskij informacziorny`j resurs"], *Bibliosphere*, vol. 2, pp. 63-68.
20. Ovchinnikova, N.A. (2017), "Archival document as a source of formation of modern information retrieval systems and electronic databases of institutions" ["Arkhivny`j dokument kak istochnik formirovaniya sovremenny`kh informacziunno-poiskovy`kh sistem i e`lektronny`kh baz danny`kh uchrezhdenij"], *In III scientific-practical conference "Document. Archive. Information society*, Moscow, September 28, pp. 379-385.
21. Filimonova, N.V., Khopiaynen, O.A. (2016), "Cataloguing information resources on cultural, historical and social development of the indigenous peoples of KhMAO-Yugra" ["Katalogizacziya informacziorny`kh resursov o kul`turno-istoricheskom i soczial`nom razvitii korenny`kh malochislenny`kh narodov KhMAO-Yugry"], *Bulletin of Ugric Studies*, vol. 2, pp. 167-172.

Надійшла до редакції 05.04.2020

Ю.В. ПРОКОП¹, О.Г. ТРОФИМЕНКО², С.М. ВОРОНОЙ¹

¹ Одеська національна академія зв'язку ім. О. С. Попова, м. Одеса, Україна

² Національний університет «Одеська юридична академія»

yulia13.prokop@gmail.com, voronoy@onat.edu.ua, egt@ukr.net

СТРУКТУРА ГЕОГРАФІЧНОГО БЛОКА ДЛЯ АРХІВНОЇ ІНФОРМАЦІЙНО-ПОШУКОВОЇ СИСТЕМИ

В умовах зростання інтересу до історичних, краєзнавчих та генеалогічних досліджень зростає суспільний попит на створення архівних інформаційно-пошукових систем, які б мали ефективні засоби швидкого виявлення документів та пошуку відомостей певної тематики, зокрема, для відбору архівних справ за географічною ознакою. В статті пропонується ієрархічна структура географічного блока для систем пошуку архівної інформації, яка враховує регіональні особливості устрою країни в різні історичні періоди. Завдяки встановленню адміністративних зв'язків між районом/повітом, областю/губернією і відповідними населеними пунктами, система при використанні цієї структури дозволяє здійснювати пошук архівних даних не лише за назвами населених пунктів, а й по повітах і губерніях та всіх підпорядкованих ним адміністративних одиницях. Наукова новизна результатів полягає у комплексному вирішенні проблем пошуку архівних даних за географічною ознакою, що дозволяє архівістам вдосконалити облік документів, а користувачам оптимізувати пошук даних при онлайн-звертанні до архівних установ.

Ключові слова: інформаційно-пошукова система, архівні бази даних, пошук за географічною ознакою, маркування архівних документів, пошук архівних даних.

Ю.В. ПРОКОП¹, Е.Г. ТРОФИМЕНКО², С.М. ВОРОНОЙ¹

¹ Одесская национальная академия связи им. А. С. Попова, г. Одесса, Украина

² Национальный университет «Одесская юридическая академия»

yulia13.prokop@gmail.com, voronoy@onat.edu.ua, egt@ukr.net

СТРУКТУРА ГЕОГРАФИЧЕСКОГО БЛОКА ДЛЯ АРХИВНОЙ ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ

В условиях роста интереса к историческим, краеведческим и генеалогическим исследованиям растет общественный спрос на создание архивных информационно-поисковых систем с эффективными средствами быстрого выявления документов и поиска сведений определенной тематики, в частности, для отбора архивных дел по географическому признаку. В статье предлагается иерархическая структура географического блока для систем поиска архивной информации, учитывающей региональные особенности устройства страны в разные исторические периоды. Благодаря установлению административных связей между районом/уездом, областью/губернией и соответствующими населенными пунктами, система при использовании этой структуры позволяет осуществлять поиск архивных данных не только по названиям населенных пунктов, но и по уездам и губерниям и всем подчиненным им административным единицам. Научная новизна результатов заключается в комплексном решении проблем поиска архивных данных по географическому признаку, что позволит архивистам усовершенствовать ведение учета документов, а пользователям – оптимизировать поиск данных при онлайн-обращении в архивные учреждения.

Ключевые слова: информационно-поисковая система, архивные базы данных, поиск по географическому признаку, маркировка архивных документов, поиск архивных данных.