

The structure of historical intelligent systems knowledge base

Hubarevich Nastassia

Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
stasia@tut.by

Vityaz Siarhej

The Center for the Belarusian Culture, Language and Literature researches of the National Academy of Science of Belarus
Minsk, Belarus
vit@history.minsk.by

Grigyanets Romuald

The State Scientific Institution "The United Institute of Informatics Problems of the National Academy of Science of Belarus"
Minsk, Belarus
griganec@bas-net.by

Abstract—The paper considers the approach to the formalization of historical knowledge on the basis of historical sources, which possess variability, incompleteness and unreliability of knowledge, OSTIS technology is used as a technological basis, and knowledge is represented in the form of semantic networks with a basic set-theoretic interpretation.

Keywords—digital history, knowledge representation, methods of knowledge processing

I. INTRODUCTION

In modern science there are most popular researches at the junction of two or more sciences. One of these areas includes the Digital history, which is gaining increasing development in foreign, Russian and Belarusian science.

Computer technologies came to the historical researches like the separate field in the 80s, together with the creation of the association "History and computing" in Britain, when the history reseachers began to realize and identify themselves separately from the general area "Computer and the humanities". The main accent in both areas was made on information support of the humanitarian researches and, in particular, historical researches. Gradually the foreshortening was changed to instrumental support of research, and using the mathematical and statistical methods in humanitarian research. And was fixed the new term "History computation".

Later, it became evident the insufficiency and limitations of statistical and mathematical research methods for humanitarian, and especially historical knowledge. First of all, this was due to the fact that these methods had an application mainly in economic and demographic research in the field of history. "Computing" field showed itself as excessively rigid, "instrumental", remote from interests and researches of the majority of historians [1]. Especially when reseachers began to use the computation methods and modern information technologies for analyzing weakly structured text sources, the need to develop another, source-oriented approach, which would take into account the specifics of historical sources with their "irregularities" and "blurring" of data was realized [2]. In this regard, there was a turn-out from methods to sources, which was marked by an increase in interest in the problem

of creating information resources in the field of history. In science this direction was called Digital history.

Digital history is defined as a broad field of applied development for improving the work of historians - researchers and teachers - based on computer technologies, to facilitate users' access to digitized historical resources, to increase their visibility and visual representation, to expand the possibilities of working with data [3].

Currently, distinguish four most priority areas in the digital history evolution [4]:

- The area, which is connected with virtual reconstruction, spatial visualization and use of GIS technologies in historical research [5], [6], [7].
- The second one is associated with the creation of big online resources containing materials on history with open databases, with the possibility of third-party replenishment and editing [8]
- The third area is connected with usage interactive hypermedia technologies, nonlinear information presentation (interactive narrative). [9], [10].
- The fourth is associated with the resources of collective authorship (Web 2.0) elaboration.

The further evolution of each area poses the question of the historical knowledge representation and the methods of knowledge processing, considering the particular qualities of historical sources. In questions related to the historical knowledge representation it is necessary to pay attention to the following problems:

- The problem of incomplete, fuzzy and contradictory knowledge representation. For example, the different naming of the organization in different sources and even in the single one.
- The problem of the temporal characteristics of historical objects representation, such as the changes in tracing of streets, the street names, house numbering etc.
- Problem with the exact dates of the historical situations and events.
- Problem of the automation of the process of correlation the different pieces of information from various sources

with each other. For example the data of street names and its building numbering is contained in various sources, and we need to reconstruct all addresses of each building on the street during the whole time of its existence.

- The problem of the multilingualism of sources. So the sources on the history of the city of Minsk are presented in Russian, Belarusian, Polish, rarely German, French and Latin, and there is a need to information correlation.
- The problem of automatic knowledge base filling from the partially structured sources on natural language.

II. PROPOSED APPROACH

The paper presents an approach to the formalization of information from historical sources. As a technological basis for the implementation, *OSTIS technology* is used [11]. Systems based on this technology are called *ostis-systems*, and they consist of a knowledge base that stores all information about the entities described, and knowledge processing machines, a team of agents that perform operations on knowledge.

The knowledge base of the system is a semantic network, where the vertices of the network represent concepts, and the edges indicate the relationship between concepts. And there are some different between the sign of entity (sc-element) and the identifier (name) its entity in various external languages. This separation allows, without changing the description of the entity itself and its connections with other entities in the system, to change the identifier language used only for interaction with the user of the system.

In the historical knowledge base the main identifier for the of entities assigned by the name of the first mention in the source, and to present the changes in entity names, the temporal decomposition is used. The variability of these names is formalized as synonymy. For example the organization BSSR Council of People's Commissars have as the the temporal parts the names "BSSR Supreme Soviet", "Supreme Soviet of the Republic of Belarus", "National Assembly of the Republic of Belarus". And the synonyms of the first name will be Sovnarcom, SNK etc.

At present time in the historical knowledge base with the following degree of completeness, the following *subject domains (SD)* are identified and described:

- *SD of artifacts* describes all historically valuable and artificially created material entities as a result of a purposeful activity.
- *SD of urban planning* describes immovable monuments of history and culture.
- *SD of persons and social communities* considers a person and all arising from his activity communities of people.
- *SD of ideas* describes compiled on the basis of purposeful activity results.
- *SD of historical actions and events*
- *SD of historical sources*, which describes sources, the information from which formed the basis of the knowledge base.

The most of concepts in the system are the temporal entities, which have the temporal characteristics, such as duration in time, the time of the beginning and the time of ending.

All changes of entities and their relations with other entities are represented as situation – structure that contains at least one temporal entity. The combination of several temporal entities and the connections between them in the structure means that at some point in time all these entities in their interrelationship existed simultaneously.

Situations are divided into the following classes:

- *One-time situation* – the situations which mostly represented the moment of the emergence of some non-material entity, which means that from this moment this entity in the form of some idea will always exist. For example, the situation of authorship between a person and his work;
- *Completed situation* – situation which had in the past the time of the beginning and ending also ;
- *Non-completed situation* is a situation that began in the past but has not been completed yet, such as the existence of some person or building.

The time of the situation existence can be set by:

- *The time point*, which has discrete and arbitrarily exact values. For example, September 1, 1939 4.40 AM and September, 2 1945 4.02 PM – the begin and the end date of the World War II.
- *The period of time* – duration between two time points, such as in the sentences "Minsk City Theater was built in the XVII century and destroyed in the 60s of the XX century".

This approach to the formalization of knowledge in historical systems makes it possible:

- Eliminate duplication in the knowledge base of elements that denote the same entity;
- Present fuzzy time boundaries of the duration of events preserving the possibility of machine processing and analysis of information;
- Specify temporary data with any necessary degree of detail, in any chronological system, and be able to automatically translate and compare dates from one system of chronology to another;
- Describe the temporal properties of not only specific entities, but also whole situations with any necessary number of elements;
- Simply and flexibly make changes into information about entities. By introducing or deleting some information about entities, there is no need to rewrite the situation with the participation of this entity;
- Convenient search by association. That is, we can query the system for any time slices relative to any objects in the knowledge base, we can request specific types of relationships of a given entity with a certain class. For example, we can inquire about the state of some building

at a particular time, find out all the tenants in this time period, and then find out the other addresses of these tenants in the city.

Every situation in knowledge base by the relation source* connected with the specific document, where the information about the represented situation was given.

As sources of information are used: Memorial books of Minsk province (1845-1917)[12], Minsk hadbookes (1926, 1967)[13], book "The Minsk main street" (1880-1940) [14]

III. EXAMPLES OF IMPLEMENTATION

Consider concrete examples of the implementation of these solutions in the historical knowledge base of the city of Minsk.

The figures 1 and 2 demonstrates an example of the street temporal decomposition, where there are different names in different time periods are represented. The second figure demonstrate the formalization of the period of existence of the concrete street name.

The building has the following characteristics: geolocation (coordinates on the map), building number at different time periods, owners, residents of this building, the organizations in it, architectural style of the building at different time periods, type of building, construction material, architectural elements. The description can be supplemented with any other necessary information.

As can be seen in figure 3, the building called "Przhe-liaskovski House" from 1896 to 1912 had number 18, during the period 1913-1938. the building had number 14, and in the period from 1939 to 1951 it was number 12.

There is no need to restore and enter manually the building address. The system have the knowledge about the date of changing the street names and the date of changing the numbers of each building, and will be able not only restore the addresses of concrete building automatically for the entire period, was described in the system, but also to search for the building at the address in exact period of time.

The organization characteristics in the system are the structure, composition, belonging to the classes of organizations etc. Example on the figure4

Persons are characterized in the system by the following signs: date of birth and death, place of work, study, residence, positions held, participation in organizations, class affiliation, major achievements, authorship, etc.

The description of any concept of the real world in the system makes it possible to present other information about the object in the form of drawings, video fragments, etc. For a person, these can be his portraits, voice recordings, for buildings - plans, measurements, sketches and so on. Such representation of knowledge about the history of the city gives wide opportunities of knowledge processing.

The knowledge about the concrete time of the persons employee in the particular organization, we can automatically

find out the name of the organization in any given period, the buildings it was located, the co-workers of the person in this organization. That gives wide opportunities not only for nonlinear presentation of information, but also broad opportunities for its analysis for the presence of implicit relationships between entities. It gives the possibility to find out all places in the city connected with any person in knowledge base. Agents will correlate the time of work of the person in the organization with the time of location of concrete organization in different buildings, all places of persons living and will give the answer, as on the figure 5.

IV. CONCLUSION

The proposed approach to the presentation of historical knowledge makes it possible to fully reflect information from sources, taking into account their incompleteness, inaccuracy and fragmentation, and also to present historical knowledge in a formal form accessible to machine processing.

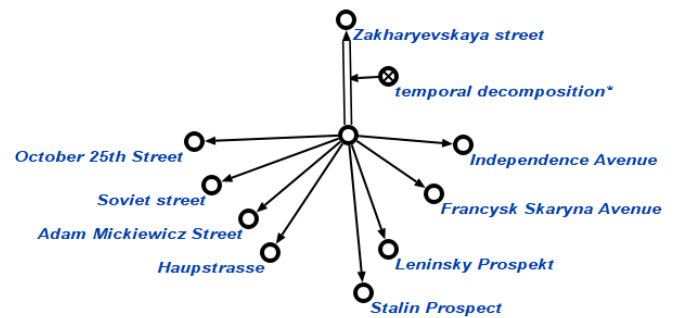


Figure 1. An example of decomposition of Zakharevskaya street.

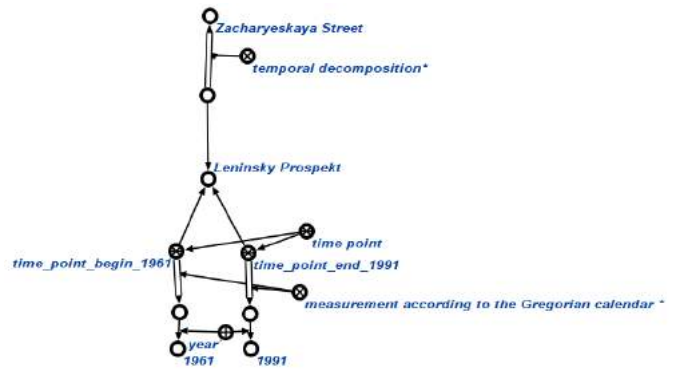


Figure 2. An example of dating a temporal part of the street.

REFERENCES

- [1] Garskova I.M. Analysis of the historiography of historical informatics as a scientific. *Kharkiv historiographical collection*, 2010, vol.10, pp. 138-171
- [2] Garskova I.M. Historical informatics and quantitative history: continuity and interaction. *Analysis and modeling of socio-historical processes*, Moscow, 2006, pp. 63-64.
- [3] Borodkin L.I. Historical informatics: overload? *Bulletin of the Perm university*, 2011, Iss. 2(16), pp. 5-11.

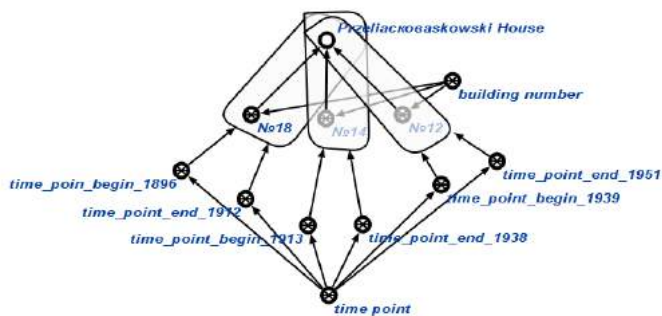


Figure 3. An example of dating a temporal part of the street.

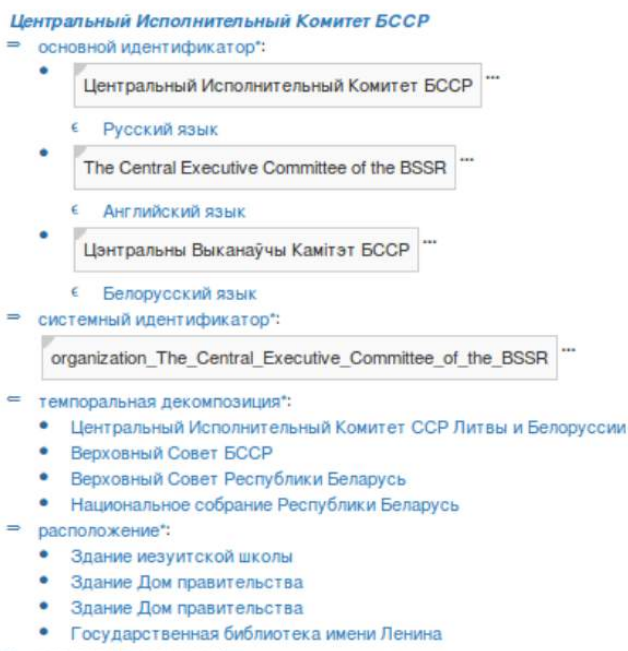


Figure 4. An example of dating a temporal part of the street.

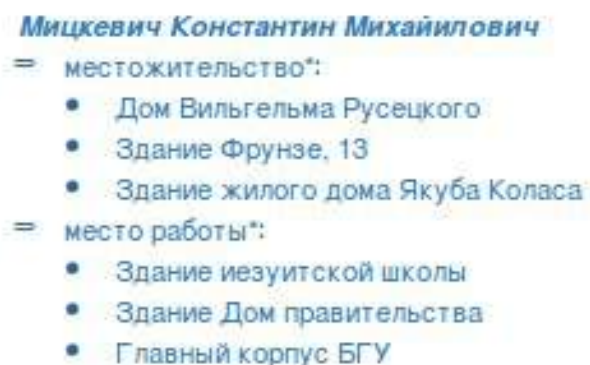


Figure 5. An example of dating a temporal part of the street.

[4] Arthur P. Exhibiting history: the digital future. *Journal of the National Museum of Australia*, 2008. Vol. 3, No 1

[5] Bryukhanova I. E. A., Rygalova M. V. Historical GIS on-line: a Review of Foreign and Russian Projects. *Journal of the Altai State University*, 2016, No 2, pp.56-59

[6] Nemtinov V. A., Morozov V. V. Virtual modeling of objects of cultural and historical heritage with using GIS technologies. *Bulletin of the Tambov State Technical University*, 2011, vol.17, No 3, pp. 709-714

[7] Ivakin Yan A., Potapichev Sergey N. Geochronology tracking — specialized GIS-tool for history researches. *Historical Information Science*, 2016, No 1-2, pp. 3-11

[8] Makewell Terry The National Museums Online Learning Project Federated Collections Search: Searching Across Museum And Gallery Collections In An Integrated Fashion. Available at: <https://www.museumsandtheweb.com/mw2008/papers/makewell/makewell.html>

[9] The Labyrinth Project. Available at: <https://dornsife.usc.edu/labyrinth/about.html>

[10] The Big History Project. Available at: <https://www.bighistoryproject.com/>

[11] The IMS.OSTIS website. Available at: <http://www.ims.ostis.net>

[12] Memorial books of Minsk province (1845-1917) Available at: <http://infobelarus.nlb.by/7n3119cd/7n3119cd.html>

[13] Minsk. Brief handbook Available at: <http://minchanin.esmasoft.com/books/minsk67/>

[14] Marakou L. The main street of Minsk (1880-1940). Book 1. Minsk, Publisher 'Mastackaja litaratura', 2012. 260P.

СТРУКТУРА БАЗ ЗНАНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ ПО ИСТОРИИ

Губаревич А.В. *, Витязь С.П.***, Григянец Р.Б.***

Белорусский государственный университет информатики и радиоэлектроники

**Центр исследования белорусской культуры, языка и литературы НАН РБ
***ОИПИ НАН РБ
Минск, Беларусь

В работе рассматривается подход к формализации исторического знания на основе исторических источников, обладающих вариативностью, неполнотой и недостоверностью знаний. В качестве технологической основы используется технология OSTIS, и знания представляются в виде семантических сетей с базовой теоретико-множественной интерпретацией.