

System approach for Digital History

Nikolay Bystritskiy^{1*}

¹LMSU, Institute of Oriental Studies RAS, Moscow, Russia

Abstract: Present advanced capabilities for information storage and a clear presentation, uncover the possibility of accumulation in digital form large volumes of historical data. For this purpose, it is necessary to establish the theoretical foundations and principles of formalization and presentation of historical knowledge. This paper presents the description and experience of practical implementation of the developed methods for formalization and analysis of historical data sources.

1 Introduction

The utilization of modern advances in information technology has already produced impressive results, and not only in the field of mathematics and natural sciences, but also in applied fields related to the systematization of a large amount of data (management, finance, law, etc.). As for the humanities, the effective use of the IT capabilities faces a number of fundamental problems: a weak degree of formalization, the difficulty of objectification and verification of information, the problems of its linkage and differentiation [1]. The fascination with the issues of digitizing and preserving textual and graphic sources, rather than solving methodological problems, led to the fact that we can talk more about the Digital libraries, rather than Digital Humanities. Considering this, some researchers even note a certain crisis in this area [2]. To overcome the current situation and transform the Digital Humanities into a full-fledged area of professional scientific research, it is necessary to conduct theoretical and practical research, in which the possibilities of formalizing and mathematizing certain aspects of the knowledge would be formulated and verified.

2 Problem

As practice shows, for historical data procession and analysis, mathematical or formal logical methods are not enough, but it is necessary to find postulated theoretical consistent patterns that need to be verified on real factual material. As the experience of related disciplines (primarily in the field of creating automated control systems) shows that the

* Nikolay Bystritskiy: nb@rusistin.ru

most effective methodology for operating with poorly formalized data is to combine the creation of full-scale databases with the construction of research models for their theoretical understanding.

Today, modern historians come to the fore the task of creating a verified historical hypernarrative, verified by reliable databases [3]. This involves the creation and integration of resources necessary for the work of the historian across the entire spectrum of traditional areas of research, such as bibliographic, prosopographic, factographic, chronology, geographic databases, ensuring the systematization and processing of the main groups of narrative sources. It becomes clear, that the specifics of history as a scientific discipline requires the elaboration of the theoretical foundations and principles of formalization, visual representation and scientific analysis of the factual data, with the subsequent verification of the developed approaches in solving practical problems.

Despite the numerous attempts of research groups and the existence of developments of individual institutions, in the field of historical research there is still insufficient systematization of implemented projects. At the present stage, a significant difficulty that hinders the development of Digital History is insufficient attention to the generalization of technical and methodological solutions obtained by colleagues in the course of work on different projects [4, 5]. In this regard, we note that the development of universal methodological approaches to formalizing and informatively describing historical data is of high research value and relevance.

The proposed approach to digitalization of historical knowledge is aimed at a systematic solution to the following problems:

1. Identification of typology and classification of information knowledge elements about the historical process.
2. Creation of a technique of knowledge elements formalization and their terminological registration.
3. Reduction to a single view and comparison of various forms of messages (factual knowledge) about one event/action.
4. Formalization of factual knowledge in the form of a rigid structural scheme: acting object A – action – passive object B – circumstances.
5. Choice of a single description format for chronological and spatial data, including the non-trivial type (interval, fuzzy, etc.).
6. Supply of factual knowledge with spatial and temporal characteristics (using geographical coordinates and chronological Julian day)
7. Identification of mutual relations between factual knowledge entities (temporary, causal, typological, genetic, etc.).
8. “Type” and “rank” characteristics setting of factual knowledge for the possibility of their subsequent classification and filtering.
9. Processing types of variable information (various versions, hypothetical reconstructions, etc.).
10. Determination of reliability (probabilistic) parameters of factual knowledge and their sources.
11. The development of the methodology for operating imaginary events, which, without having reality, can have a real impact on the course of the historical process.
12. Suggesting applications for the visual representation of knowledge data sets, their grouping, linking and comparison (visualization).

3 Approach

Here we will dwell in more detail on the developed system approach to the classification and formalization and analysis of elements of historical knowledge.

3.1 Typology of historical knowledge elements

The subject of history is a person and his actions [6]. Therefore, the main producing object (actor) is a person and his social configurations - groups and organizations. Natural objects, as well as the operations of recently emerging intelligent digital devices, also have a noticeable effect on person's activity. All of them are participants (objects) of the historical process. Understanding that the elements of knowledge represent information about the instances of reality constituted the historical process, we first need to identify and typologize them [7].

The historical process itself is a complex interweaving of time-unfolded actions and natural processes. The action *<actus>* is purposeful (target-setting) in essence and is carried out by the so-called acting object (person, group, organization). Unlike action, the natural process *<processus naturalis>* is fundamentally not targeted (for example, natural, transcendental or random phenomena).

Messages that are part of historical sources, in most cases, tell about the results of human activity or natural forces [8]. Based on this, historical events *<obvenientia>* represents of the results of actions/untargeted processes. It is assumed that the event has a point (instantaneous) duration and can be specified by temporal and spatial coordinates [9]. An action/process has a spatial-temporal duration and can be represented as a vector starting with an initiation event and ending with an end event. The latter is a determining event, since it is precisely characterizes the analyzed state of a historical object.

Ontologically, the main types of actions/processes are divided into main categories: in form – a) material or b) information; by content – a) output; b) receipt; c) stay; d) transformation. In the direction of action, distinguish:

- active (Aa) - one or more active objects acting outside ($Oa \rightarrow Op$);
- mutual (Ai) - symmetrically interacting objects ($Oa/p \leftrightarrow Op/a$);
- returnable (Ar) - one object whose actions are directed at itself ($\leftarrow Oa/p \leftarrow$);
- passive (Ap) - undergoing external action ($Op \leftarrow Oa$).

Based on this classification, actions operands are established that combine similar types of actions. The actions operand defines a group of actions that can be linguistically expressed through lexical synonyms (verbs) with intersecting or including one other semantic fields.

3.2 Characteristics of factographic knowledge and their formalization

3.2.1 Actions characteristics

For the convenience of operating with knowledge about historical actions/events and with the aim of structuring them, they should be formalized and distributed into subgroups: domestic policy (unrest and civil war, state, economic and legal), foreign policy (war, embassies, pacts), religious, cultural, biographical, natural (astronomical and climatic phenomena, disasters).

Undoubtedly, understanding the role of historical action/event in the global process and establishing the degree of its impact on the subsequent course of history seems promising. To formalize the role, one should use such an evaluation characteristic as the rank (significance) <gradus> of a historical action/event. Applying such a technique, a scientist has the ability to classify an action/event into a specific category, ranging from universal significance (super-important, rank 0) to personal significance (negligible, rank 9).

It should be noted that most actions are complex, and can include chains of action at a lower level (for example: war – campaigns – battles – attacks, etc.). Thus, upon reaching a certain level of detail, the number of analyzed actions will reach such huge volumes that it becomes necessary to resort to the services of Big data technologies [10, 11].

3.2.2 Connections

Naturally, various connections can be seen between actions and processes data: temporal, spatial, causal, etc., called inter-actions connections <nexus interactus>. According to their orientation, the connections are divided into direct ones – from the past to the future (for example, causal) and reverse ones – from the future to the past. By their sign, they are positive (“through”, “achieve”) and negative (“contrary”, “avoid”). The length of the connection is determined by the size of the interval (temporal and/or spatial) between the event of the end of the first action and the event of the beginning of the next associated with it. Depending on the affiliation of the connections to a particular mode of time relative to the current object, three types of inter-actions connections can be distinguish: realized, actual and potential. The content of the historical past of the object is determined by the totality of the implemented connections that connect actions that are already in the past. Actual connections are characteristic of the present, that is, those connections whose implementation has already begun, but has not yet been completed, and which combine events of the past, on the one hand, and the future, on the other. Finally, the future of the object is made up of potential connections, the implementation of which has not yet begun, since they combine the proposed future actions.

3.2.3 Causal and targeted connections

An important element in the analysis of historical data is the creation of a methodology for setting and describing causal connections <nexus causalis> between the actions (events) of the historical process objects [6]. The results of the action-cause presented by the final event are the initial factors for the action-effect. Causal connection is characterized by intensity – a parameter that reflects the degree of influence of the cause on the effect. Each cause can act as a stimulus, i.e. to favor the appearance of the investigation is a contributing (positive) causal connection, and to complicate its onset is an obstructing (negative, inhibiting) connection. If for a process-effect, causes have a direct effect, then for an action-effect, causes also affect the target setting of an active object [12].

The specificity of the determination of human activity lies in the fact that, along with the causation of subsequent actions by the previous ones (determination by the past), there is also the determination of the future, that is, the targets and intended results of the activity. In this case, the targeted connection <nexus destinatus> is formed between the event-target and the action-means.

Finally, inter-actions connections can be composite when actions can be both in causal and targeted connections with each other.

3.2.4 Temporal and spatial characteristics

A separate task is the adequate fixation of temporal and spatial characteristics [13, 14]. For an unambiguous description of the temporal characteristics of historical events, it is considered expedient to use the so-called chronological Julian day (a sequential count of the day, starting from 12:00 p.m. January 1, 4713 BC Julian calendar), widely used in astronomy and computer science [15]. The time interval acts as a typical temporal characteristic of the action. For describing the relationships between time intervals in science, the apparatus of interval time logic and temporal networks are used [16–22]. In many cases, we know the dates of historical events approximately, and sometimes there are several options for dating one event. An effective approach is the use of interval algebra with fuzzy boundaries, which uses probabilistic methods [23–25].

A similar approach could be used to formalize the spatial characteristics of events. To describe spatial data, it is natural to apply the traditional geographical coordinates widely used in science and technology, and to consider a vector (spatial interval) as a typical characteristic of an action. Based on it, using the apparatus of spatial interval logic and spatial networks, it is possible to build spatial paths of any complexity [26–28].

3.2.5 Veracity

A special set of important theoretical questions is related to the probabilistic nature of knowledge, including historical factual data and hypotheses. For a researcher, first of all, it is necessary to formalize and verify the reliability of certain historical sources [29]. Each message of a historical source that carries factual information has a certain veracity <testatus> associated with the subjectivity of the author, the degree of his awareness, dependence on other sources, genre features, etc. With regard to archaeological and other “silent” sources, the assessment of the veracity of research hypotheses and interpretations comes to the fore. Thus, each element of historical knowledge has its own veracity indicator, which should be identified by scientists in the course of analysis. For research empirics, the indicator of knowledge veracity should take values in the interval from absolutely reliable (1.0) to absolutely unreliable (0).

The veracity indicator is the most important attribute of historical knowledge, which is mostly derived from estimates of the reliability of information sources. Actual data about the action/event may be considered worthy of attention depending on the threshold values of the veracity indicator.

3.3 Reconstruction: hypothesis production and assessment

One of the main tasks of historical research is the reconstruction and explanation of the past – hypotheses proposition based on the construction of causal chains of actions. The usual research procedure involves several stages.

At the first stage, an analysis of factual material on belonging and involvement in the subject of research. Here, from the entire set of data, data sets reflecting the studied objects, actions, connections and their characteristics are distinguished. Further, a detailed analysis of all elements of the data set and reconstruction of inter-actions connections are carried out. Particularly responsible for the hypothesis is the construction of causal relationships that form a chain of actions. In continuation, the researcher provides an explanation of the hypothesis. At the final stage, the researcher must evaluate the hypotheses – determine its

credibility (veracity) and formulate a logical, factual and historiographic justification of the hypothesis. The main components of this procedure can be efficiently digitalized.

The practical issues of evaluating hypotheses, as well as establishing indicators of reliability and significance (rank), can be solved using logical methods for analyzing the integrity, connectivity and consistency of the totality of historical data, as well as using expert ranking and mathematical probabilistic methods [30, 31]. Their optimal implementation needs serious study on large data sets of specific historical material. Here, methods of collective expert assessments are applied, which involve determining the degree of agreement of expert opinions on specific data parameters. To assess the degree of agreement on a particular parameter, the following indicators are calculated: the variance of the estimates, the standard deviation of the estimates, the coefficient of variation of the estimates and the Kendall concordance coefficient [32]. The values of these parameters allow us to conclude the heterogeneity of expert opinions and form the most convincing estimates.

The veracity assessment of competing hypotheses may be carried out depending on the cumulative characteristics of the veracity of the totality of actions/events and the connections included in them. Based on the cumulative indicator, we are able to compare the veracity of hypotheses and make a choice regarding the most plausible.

4 Implementation

For a long time the author with colleagues are creating a database of digital information of sources on Byzantine history [33, 34]. The accumulated data are experimental material for validating the presented approach. The developed database includes following integrated data: prosopographic (historical objects – participants of the historical process), geographic (spatial data), factographic (actions/events) schesiographic (connections). As an illustrative implementation example, we will consider the hypothesis creation and analysis procedure by constructing options for action chains based on the study of 4 database records.

Action 1. The attack of Russia on Constantinople in 860

Time: 2035337-2035338 (June, 18 860); veracity of dating = 0.9

Place: 41.0167 ° N; 28.9667 ° E (Byzantium> Constantinople)

Type: foreign policy

Rank: 3 (regional)

Active object (Oa): army of Rus'

Passive object (Op): city population

Action (A₁): MIL IMP (military attack)

Sources: Cumont F. *Anecdota Bruxellensia. T. I.* (Gand, 1894). P. 45; Φωτίου Ομιλίαι (Θεσσαλονίκη, 1959). Σ. 29–52; *The Russian Primary chronicle: Laurentian text.* (Cambridge, 1953). P 58–61.

Action 2. Emperor Michael III hastily returned to Constantinople and gathered an army

Time: 2035320-2035350 (June 860)

Place: 41.0167 ° N; 28.9667 ° E (Byzantium> Constantinople)

Type: foreign policy

Rank: 3 (regional)

Active object (Oa): Byzantine army

Passive object (Op): army of Rus'

Action (A₂): MIL MNS (treat)

Sources: Cumont F. Op. cit.; Φωτίου. Op. cit.

Action 3. Patriarch Photius and Emperor Michael III perform a divine service of deliverance from the invasion

Time: 2035320-2035350 (June 860)
Place: 41.0167 ° N; 28.9667 ° E (Byzantium> Constantinople)
Type: religious
Rank: 3 (regional)
Active object (O_a): city population
Action (A₃): RCM ORA (prayer service)
Sources: Φωτίου Op. cit.; Theophanes Continuatus. (Bonae, 1838). P. 1–81; Kuzenkov, P.V. “Campaign of 860 to Constantinople and the first baptism of Rus’ in medieval written sources.” In *The oldest states in Eastern Europe 2000*. (Moscow, 2003). P. 3–172.

Action 4. The retreat of Rus’ from Constantinople

Time: 2035350-2035411 (July-August 860)
Place: 41.0167 ° N; 28.9667 ° E (Byzantium> Constantinople)
Type: foreign policy
Rank: 3 (regional)
Active object (O_a): army of Rus’
Action (A₄): MIL RCM (retreat)
Causes (C₄): 1. panicky religious fear (0.3); 2. approach of imperial troops (0.7); 3. goal achievement (0.1)
Sources: Cumont F. Op. cit.; The Russian Primary chronicle.

Based on selected database entries, we can build a number of hypotheses.

Hypothesis 1. Action 1 → Action 2 → Action 4.

Description: the retreat of the troops of Rus’ was a consequence of the threat of an attack by the Byzantine army.
Results of analysis: veracity of sources - 0.8, logical consistency - 0.7.
Assessment of the hypothesis veracity: rather reliable - 0.56.

Hypothesis 2. Action 1 → Action 3 → Action 4.

Description: the retreat of the troops of Rus’ was the result of fear of divine powers.
Results of analysis: veracity of sources - 0.5, logical consistency - 0.7.
Assessment of the hypothesis veracity: rather unreliable - 0.35.

Hypothesis 3. Action 1 → Action 2 → Action 3 → Action 4.

Description: the retreat of the troops of Rus’ was a consequence of the threat of an attack by the Byzantine army and, then, of fear of the divine forces.
Results of analysis: veracity of sources - 0.5, logical consistency - 0.2.
Assessment of the hypothesis veracity: extremely unreliable - 0.1.

Hypothesis 4. Action 1 → Action 3 → Action 2 → Action 4.

Description: the retreat of the troops of Russia was the result of fear of divine forces and, then, the threat of an attack by the Byzantine army.
Results of analysis: veracity of sources - 0.5, logical consistency - 0.7.
Assessment of the hypothesis veracity: rather unreliable - 0.35.

As a result, we are able to compare hypotheses and select the most plausible.

5 Further applications

Aggregation of historical data obtained from various sources into a single formalized database opens the way to the use of digital methods for their grouping, systematization, establishment of temporal, spatial and other connections. Further use of analytical tools makes it possible to scrupulously look at the topology of events, their conditioning, will provide a deeper understanding of social processes dynamics. A systematic analysis of sets

of actions, taking into account all the connections, will reveal the potential for solving many current tasks.

5.1 Possibilities for clarifying dates and routes

By analyzing the dating and localization of events in the complex, it is possible to build chains of actions in time and space, according to the characteristics of which we can validate the veracity of dating and localizations. For example, comparing the value of the relationship between the lengths of the spatial and temporal durations of an action or connection, with a confidence range of speeds of movement or forwarding of correspondence, it is possible to come to the conclusion about the correctness of dating and/or localization. Their significant difference suggests dating or localization error. The work of modern researchers already use this technique to clarify the chronology of events and travel routes [35].

5.2 Search for implicit regularities

A detailed analysis of a large number of actions and connections data sets allows determining typical configurations of actions and the connections between them (action patterns). Such patterns make it possible to identify certain regularities, including hidden ones. In addition, the found deviations from typical configurations in other data sets may indicate incompleteness or inaccuracy of the data, or about the uniqueness of this situation.

5.3 Predictive capabilities

Many scientists and public figures noted the educational and prognostic potential of history. It can be fully realized only by determining certain patterns of society development [36]. In this vein, the identification and analysis of action patterns will allow us to create a set of typical options, the parts of which can be used to assess current and future situations, create expert decision-making systems for difficult conditions.

5.4 Automatic narrative generation

With the development of intelligent referential and search engines, immersion applications and games, the role of developed user interfaces, which include narrative generation systems, is growing. The source for such systems is information from structured data banks, and, for example, in the field of history, the one described in this paper can be considered. Research teams of the leading universities and corporations carry out development in the prospective field of automatic narrative generation [37, 38].

5.5 Scientific visualization

Today, visualization used in many scientific fields, since it allows to present the current research data in the form of the most convenient for the researcher to understand, conduct and evaluate their analysis and evaluation. Visualization methods are effective for history where processes are relatively slow and the life span of one generation is not enough to comprehend global phenomena. The most effective forms of historical visualization are

timeline, dynamic map (“live map”), space-time cube, 3D reconstruction, networks. Formalized datasets act as sources for all historical visualization tools [39].

5.6 History of IoT and AI activity

Modern social life is becoming more complicated – gradually new participants are introducing in it – smart devices of the Internet of Things (IoT) and Artificial Intelligence (AI). Such smart devices can act as creators of their own events and participate on a par with a person in creating joint events that shape the past of humankind and affect the future. Consequently, they become co-authors of the history. The approach described above can also be implemented to account for the activity of smart devices [40].

6 Conclusion

Based on the experience, briefly summarized above, the system approach to the acquisition and analysis of digital historical data demonstrates its effectiveness. Together with the data arrays accumulated with its support, it can serve as a real backbone for the development of a whole range of digital tools of a historian. Further enhancement of this approach and its wide scientific expertise will enable the emergence of new useful applications.

References

1. K.V. Khvostova, *Novaya i noveishaia istoria*, **3**, 66–78 (2007)
2. M. Taller, *Historical informatics*, **1**, 5–13 (2012)
3. S.P. Karpov, *Problems of the methodology of studying and teaching modern history. Proc. int. conf. October 29, 2009.* (Moscow, 2009).
4. N.I. Bystritskiy, *Vizantiiskii Vremennik*, **71**, 286–295 (2012)
5. N.I. Bystritskiy, *Vizantiiskii Vremennik*, **101**, 290–298 (2017)
6. M. Bloch, *Apologie pour l'histoire ou Métier d'historien* (Paris, 1952)
7. N.I. Bystritskiy, *Natural Science Methods in the Digital Humanitarian Environment*, 19–22 (Perm, 2018)
8. P. Ricœur, *Histoire et vérité. [History and Truth]* (Evanston, Ill., 1965)
9. C.S. Jensen, et al., *Temporal Databases: Research and Practice* (Berlin, 1998)
10. N.I. Bystritskiy, *Supercomputer technologies in science, education and industry*, 81–87 (Moscow, 2012)
11. P. Manning, *Big data in history* (N.Y., 2013)
12. P. Kuzenkov, N. Bystrickij, *Proc. of the 23rd Int. Congress of Byzantine Studies. Belgrade, 22-27 August 2016. Them. Session of Free Comm*, 231-232 (Belgrade, 2016)
13. E.Y. Kandashina, L.V. Litvintseva, D.A. Pospelov, *Predstavlenie znanii o vremeni i prostranstve v intellektual'nykh sistemakh* [Presentation of knowledge about time and space in intelligent systems] (Moscow, 1999)
14. D.J. Peuquet, *Representations of Space and Time* (N.Y., 2002)
15. P.V. Kuzenkov, *Hristianskie hronologicheskie sistemy* [Christian chronological systems] (Moscow, 2015)

16. J.F Allen, Communications ACM, **26:11**, 832–843 (1983)
17. J.F. Allen, Int. J. Intelligent Syst., **6:4**, 341–355 (1991)
18. J. Bentham, *The Logic of Time* (Dordrecht, 1983)
19. J. Bentham, El. News J. Reasoning about Actions and Change. **2** (1998)
20. A. Galton, *The Stanford Encyclopedia of Philosophy* (2008)
21. E.F. Karavaev, *Osnovania vremennoi logici [The base of time logic]* (Leningrad, 1983)
22. P. Ohrstrom, P.F.V. Hasle, *Temporal Logic from Ancient Ideas to Artificial Intelligence* (Dordrecht, 1995)
23. S. Badaloni, M. Giacomin, *APIA99*, 155-165 (Springer, 2000)
24. D. Dubois, H. Prade, IEEE SMC, **19**, 729–744 (1989)
25. H. Ohlbach, *Forschungsbericht. Institute for Informatics* (Munich, 2005)
26. *Handbook of spatial logics.* (Springer, Dordrecht, 2007)
27. A. Galton, *Foundations of Geographic Information Science* (London, 2003)
28. S.O. Varosian, D.A. Pospelov, Izv. AN SSSR. Tech.Cybernetics, **5**, 86–99 (1982)
29. S.P. Karpov, *Proc. int. conf. "Can history be objective?"* (Moscow, 2012)
30. A.K. Alikberov, Voprosy Filosofii, **2**, 16-26 (2018)
31. G.A. Gevorkian, *Veroyatnoe i dostovernoe znanie [Probable and reliable knowledge]* (Erevan, 1965)
32. J.H. Boose, B. Gaines, *Knowledge Acquisition Tools for Expert Systems* (N.Y., 1988)
33. P.V. Kuzenkov, N.I. Bystritskiy, *Proc. of the 22nd Int. Congress of Byzantine Studies, III*, 302-303 (Sofia, 2011)
34. N.I. Bystritskiy, P.V. Kuzenkov, Vizantiiskii Vremennik, **103**, 13–31 (2019)
35. E.S. Grishin, Historical informatics, **3**, 106–122 (2018).
36. I.D. Kovalchenko, *Metody istoricheskogo issledovaniia [Methods of historical research]* (Moscow, 1987).
37. A. Nanetti, S.A. Cheong, Asian Review of World Histories, **4:1**, 3-34 (2016)
38. A. Nanetti, S.A. Cheong, M. Filippov, *Proc. Int. Conf. on Culture and Computing. Kyoto, Ritsumeikan Univ., Sept. 16-18, 2013.* 104-110 (Los Alamitos, CA, 2013).
39. N.I. Bystritskiy, "History and Computer" Association information bulletin, **44**, 78–81 (2015)
40. N.I. Bystritskiy, Research Result. SHR. **4:4**, 52–66 (2018)