Intelligent Data Curation in Virtual Museum for Ancient History and Civilization

Desislava Paneva-Marinova[1] [0000-0001-5998-687X], Jordan Stoikov[1], Maxim Goynov[1], Detelin Luchev[1] [0000-0003-0926-5796], Radoslav Pavlov[1], Lilia Pavlova[2]

[1] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
[2] Laboratory of Telematics, Bulgarian Academy of Sciences, Sofia, Bulgaria
dessi@cc.bas.bg, jstoikov@shieldui.com, goynov@gmail.com, dml@math.bas.bg, radko@cc.bas.bg, pavlova.lilia@gmail.com

Abstract. The virtual museum is an advanced system managing diverse collections of digital objects that are organized in various ways by a complex specialized functionality. The management of digital content requires a well-designed architecture that embeds services for content presentation, management, and administration. All elements of the system architecture are interrelated, thus the accuracy of each element is of great importance. These systems suffer from the lack of tools for intelligent data curation with the capacity to validate data from different sources and to add value to data. This paper proposes a solution for intelligent data curation that can be implemented in a virtual museum in order to provide opportunity to observe the valuable historical specimens in a proper way. The solution is focused on the process of validation and verification to prevent the duplication of records for digital objects, in order to guarantee the integrity of data and more accurate retrieval of knowledge.

Keywords: Database Management, System Architecture, Functionality, Data Integrity, Knowledge Retrieval, Data Validation, Record De-duplication, Cultural Heritage.

1 Introduction

For a long time, cultural heritage has been maintained in museums, galleries, libraries and research laboratories, where not everyone was able to access this wealth. Digital technologies that have been developed during the past couple of years introduced new solutions of documentation, maintenance and distribution of the huge amounts of collected material. Among these new technologies are virtual museums, which have already proven their worth as a contemporary conceptual solution for access to and attractive presentation of cultural archives. Virtual museums contain diverse collections of digital objects (such as text, images, and media objects) that are organized in various ways and are managed by complex specialized services such as content structuring and grouping, attractive visualization, advanced search (semantic-based search, multilayer

Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings. Vol. 9, Sofia, Bulgaria: Institute of Mathematics and Informatics – BAS, 2019. ISSN: 1314-4006, eISSN: 2535-0366

and personalized search, context-based search), resources and collection management, indexing, semantic description, knowledge retrieval, metadata management, personalization and content adaptability, content protection and preservation, tracking services, etc. Thus, the valuable cultural heritage wealth is accessible anytime and anywhere, in a friendly, multi-modal, efficient, and affective way.

However, these systems suffer from the lack of tools and services for intelligent data curation with the capacity to add value to data. In this paper, we propose a solution for intelligent data curation in a virtual museum in order to provide opportunity to observe and analyze valuable ancient history specimens in a proper way and in their historic context, so that some yet undiscovered treasured of the human civilizations be manifested. This solution more specifically is focused on the validation and prevention of duplication of newly added or existing records for digital objects.

Section 2 of this paper presents current concepts for virtual museum system architecture, tracking main functionality and services supporting users' needs. Section 3 includes a discussion on intelligent data curation issues. In section 4, a model of intelligent data curation service is described. The paper ends with some conclusions and further development plans.

2 Virtual Museum System Architecture

The virtual museum mainly contains service panels for *Museum content management*, *Museum content presentation*, *Administrative services* (see figure 1), jointed to a *Media repository* and a *User data repository*.

The *Museum content management* module refers to the activities related to basic content creation: add (annotate and semantic indexing), store, edit, preview, delete, group, and manage multimedia digital objects; manage metadata; search, select (filter), access and browse digital objects.

The *Museum content presentation* module supports objects and collections display. It also provides collections creation (incl. search, select/browse and group multimedia digital objects according to different criteria and/or context of usage), their metadata/semantic descriptions and attractive visualization, status of collection display. Content presentation module aims to provide access to all virtual museum services through wide range of contemporary technologies and devices – not only desktop PCs, but mobile phones, tablets, TVs, VR devices, etc. Interactive media technologies are used to provide best user experience within the content of the virtual museum.

The *Administrative services* panel mainly provides user data management, data export, tracking, and analysis services.

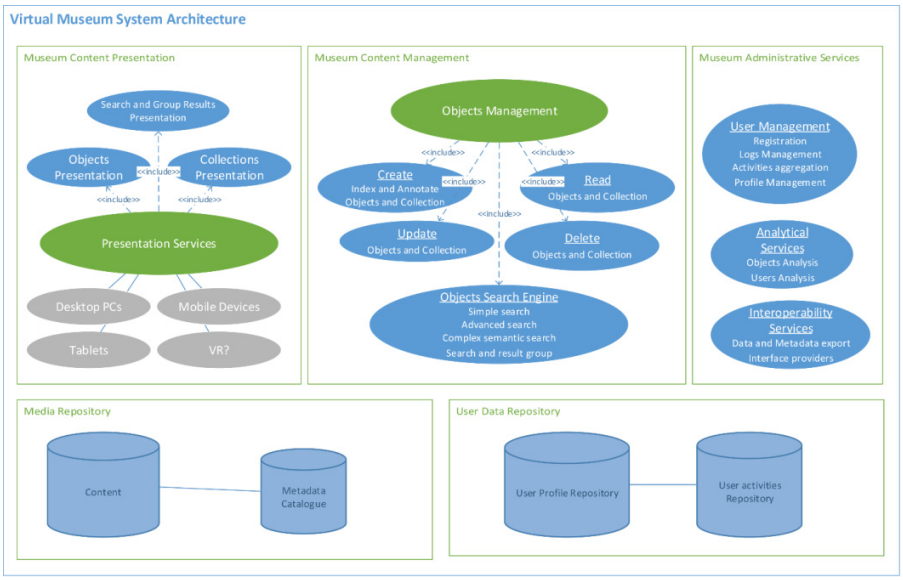

Fig. 1. Virtual museum system architecture

For every object all semantic and technical metadata are saved in the Media repository. These metadata are represented in catalogue records that point to the original media file/s associated to every object.

The User profile repository manages all user data and their changes.

2.1 Virtual Museum Functionalities in Details

Museum content management. The main part of the content creation process is the annotation and semantic indexing of digital objects in order to add them to the museum media repositories. The entering of technical and semantic metadata for the digital objects is implemented through different automated annotation and indexing services.

The technical metadata that could be expressed in Dublin Core or other standards are attached to every multimedia object automatically. They cover the general technical information, such as file type and format, identifier, date, provider, publisher, contributor, language, rights, etc.

An annotation template needs to be implemented for the semantic description of digital objects. The template provides several options for easy and fast entering of metadata:

- Autocomplete services (All used (already entered) field values are available in a panel for reuse.);
- Automated appearance of dependencies coming from the relations of the defined classes' (concepts) in an ontological descriptive structure valid for the museum object domain. All main relations and rules expressed in the

ontological structure are incorporated during the development of the annotation template;
- Bilingual data entering with automated relation between the relevant values in different languages (if it is applicable or necessary);
- Automated appearance of the number of the used field value, providing regular data tracking (if it is applicable or necessary);
- A tree-based structure of the annotation template. Only checked fields are displayed for entering metadata;
- Possibility for adding more than one media for one metadata description in order to create rich heterogeneous multimedia digital objects tracking (if it is applicable or necessary);
- Reuse of an already created annotation for new objects: the new media object has to replace the older one, the annotation is kept and the new object appears after saving;
- Automated watermarking of the image and video objects;
- Automated resizing and compression of the media objects (image or video);
- Automated identification of file formats;
- Automated conversion of the media object (audio, video, text) in a format suitable for Web-preview;
- Automated terms explanation (if it is applicable or necessary): After saving a new digital object in the museum media repository, a special machine traces for the appearance of dictionary terms in the object metadata description. If some terms are available, the machine adds links to their explanations. In the case of entering a new dictionary term, its presence in the available objects is discovered automatically and a link is added.
- Digital object duplication checks (similarities calculation): In order to avoid duplicate image objects a service that checks the similarity between images is provided. It uses an algorithm that caching images for optimizing their compare (see (Pavlov, Paneva-Marinova, Goynov, & Pavlova-Draganova, 2010)).

The virtual museum provides a wide range of search services, such as keyword search, extended keyword search, semantic-based search, complex search, search with grouping results, etc. Their realization is based on querying action to the metadata knowledge base. Moreover, five types of conditions for the results set are meant:

- "objects having or NOT characteristic c";
- "objects having value $=,\neq,\leq,\geq,<$or $>$v for characteristic c". In the search templates, the user could search digital objects with précised criteria.

The search services support content request and delivery via index-based search and browse of managed content and its description.

Museum content presentation. During the design of the content presentation services a profound analysis was made of content selection and preview possibilities in order to satisfy the user's needs. First we had to determine the preview possibilities of a separate digital object and its components and after that the preview of grouped objects (collection preview).

The visualization of the rich semantic description of the separate digital object is determined through hidden parts appearing in a new window after link selection. This possibility is used mainly for the long descriptions and for the dictionary terms. Parts of the descriptive data field are hidden, but their values are available for searching in special forms.

During the design of object grouping services, the main ontology classes of the object descriptive structure (viz. museum domain ontology) could be selected as object grouping criteria.

Every user can create his private collection of selected objects after search activity. Rich search possibilities (mentioned above) are available in order to assist collection creation. The user can write the collection's title and short description. He can also select its status: private or shared with other users. New objects for a collection appear automatically after their entering.

Custom collections in virtual museum are dynamic objects. They are criteria based, not list based. Every new object in the virtual museum will be automatically added to a collection if it meets the criteria defined for the collection. The owner/creator/follower of a collection can be notified when a new object is added and becomes part of the collection.

Every object and collection can be presented in interactive way using any browser compatible device – PC, phone, tablet, or TV. Content presentation services are based on responsive web technologies in order to satisfy the majority of the modern devices diversity and provide great user experience no matter if user consumes the virtual museum services through their phone, tablet, PC, or other smart device.

Administrative Services. The *Administrative services* panel mainly provides user data management, data export, tracking services, and analysis services. The user data management covers the activities related to registration, data changes, level set, and tracking activities of the user. The tracking services have two main branches: tracking of objects, tracking of user' activities. The tracking of objects spies on the activities of add, edit, preview, search, delete, selection, export to XML, and group of objects/collections in order to provide a wide range of statistic data (for frequency of service use, failed requests, etc.) for internal use and generation of inferences about the stable work (stability) and the flexibility of the work and the reliability of the environment. The tracking of users' activities monitors user logs, personal data changes, access level changes and user behaviour in the museum environment. The QlickTech® QlinView® Business Intelligence software could be used as an analysis provider. Therefore, it needs to be connected to the museum tracking services and objects data base by preliminary created data warehouse. It will provide fast, powerful and visual in-memory analysis based on online analytical processing and quickly answered multi-dimensional analytical queries (Codd, Codd, & Salley, 1993).

The export data services provide the transfer of information packages (for example, packages with digital objects/collections, user profiles, etc.) compatible with other data base systems. For example, with these services a package with objects could be transported in an XML-based structure for new external use in e-learning or e-commerce applications.

The authors actively try to develop workable solutions in the field of cultural heritage database management and presentation. The above described solution for virtual museum architecture follows authors' previous developments in the digital content management systems (viz.. digital libraries, digital repositories, galleries, etc.) for Bulgarian artworks and treasuries (see (Paneva-Marinova, Goynov, & Luchev, 2017), (Luchev, Paneva-Marinova, Pavlova-Draganova, & Pavlov, 2013), (Paneva-Marinova, Pavlov, & Rangochev, 2010), (Pavlova-Draganova, Paneva-Marinova, Pavlov, & Goynov, 2010) and (Goynov, Paneva-Marinova, & Dimitrova, 2011). These systems are successfully implemented to presents the valuable Bulgarian cultural heritage: Bulgarian iconographic art, Bulgarian ethnographic and folklore artefacts, medieval and early modern Bulgarian texts for saints in combination with ethnological data and visual sources, etc.

3 Intelligent Data Curation

In the modern era of big data, the curation of data has become more prominent, particularly for software processing high volume and complex data systems (Furht & Escalante, 2011). The term is also used in historical uses and cultural heritage digital assets and content management solutions, where increasing cultural and scholarly data from digital projects requires the expertise and analytical practices of data curation. In broad terms, curation means a range of activities and processes done to create, manage, maintain, and validate a component. Specifically, data curation is the attempt to determine what information is worth saving and for how long (Borgman, 2015). The essential elements of a powerful data curation tool are annotations, metadata, standards, models, databases, etc.

Moreover, data curation is intellectually intensive activity that is time consuming and requires a lot of dedicated resources. Taking into account the increasing role and amount of data, curation risks to be a bottleneck for any digital asset management or content management project in the long term. One of the challenges for the automation of data curation is difficulty in completing missing data and the level of granularity. Such a solution, however, looks practical because the data curation process is one of many iterations, consistency and includes complex data evaluation. The human and machine aspect need to be combined in order to solve the two most crucial data-integration problems: linkage of records (which often refers to linking records across disparate sources, referring to the same real-world entity) and schema mapping (mapping columns and attributes of different datasets).

One approach is to use a record to modularize curation processes. Splendiani (Splendiani, 2017) considers curation activities as functions in a "curation space" that is exemplified via a "curation record". The curation process is broken down to the following classes of operations:

- **Schema mapping:** Machine-assisted process to identify and map the similar attributes from different data sources together in one, unified data set. The same entities (e.g., events, studies, places) might be described by data sources of different origin in separate ways and in this case the usage of different schemas

and vocabularies (a dataset schema is generally an official description of the main attributes and the values that can be taken by them). For instance, one source may refer to a person's credentials by the means of two attributes (Name and Title), another source may use the terms Pers. Name and Royal Title, and a third might use PN and Rank, in order to address the same thing. The major activity in schema mapping is to set a mapping among those attributes. The problem may occur to be more challenging and may involve different conceptualizations especially in the cases when relationships in one source are represented as entities in another. Most often, in the ETL suites are used the most common schema mapping solutions that focus traditionally on the mapping of a small number of such schemas (usually less than ten) that deliver to users a suggested mapping that considers some similarity among column's name and the content of them. With the maturity of the big data stack, however, the enterprises have the power to easily acquire a huge number of different data sources and have at their service applications that can ingest data sources as they are generated. We can use an example from the pharmaceutical industry and the conducted clinical studies, where tens of thousands of studies and assays are conducted by scientists across the world, often using separate technologies and a combination of local schemas and standards. It is essential for the companies' businesses and is often required as mandatory by regulations and laws to use standardized and cross-mapping collected data. This approach has changed the main assumption of most solutions for schema mapping that the suggestions curated by users should be part of a manual process. In such a case the main encountered challenges are: (1) providing of automated solution that requires reasonable interaction with the user, meanwhile being able to map numerous schemas; and (2) designing of matching algorithms that are robust enough to accommodate different languages, formats, reference master data, and data units and granularity (Ilyas, 2018).

- **Standard setting:** Building a probabilistic machine learning model specific to the organization's domain and stakeholders based on answering a series of yes/no questions to whether two records are the same. Given enough feedback, a pattern is captured that is required to build and maintain logic in order to generate de-duplicated, master data.
- **Validation:** Throughout the human-guided process to build a machine learning model for mastering data, the user is able to see measurable outputs for each item of yes/no feedback provided. The feedback directly corrects the model. This calculation is culminated in the 'confusion matrix', indicating the precision, recall, accuracy, and F score of the model based on human feedback (tamr, 2019).

Further defining the data curation process, it is based on the organization and integration of data collected from various sources. Data curation includes "all the processes needed for principled and controlled data creation, maintenance, and management, together with the capacity to add value to data" (Miller, 2014). For example, in science, data curation may indicate the process of extraction of important information from scientific texts, such as research articles by experts, to be converted into an electronic

format (Blank, 2012). Using an efficient master data management solution can greatly facilitate the above-described process. However, a step further in the mastering process is, providing the ability to select a representative record, or golden record, for each set of duplicate, or grouped, data records derived from all data sources. For example, if a cluster of records is identified across systems for the same artist, but each record has a variation on the artist's name the human-guided machine learning approach merges those records and generate a single golden record using the most common values for each attribute - assuming that aligns with how the business perceives their data. The goal of the golden record is to consolidate and generate a single record of truth. This approach is especially applicable when update of existing record is required from several different data sources.

Summing it all up, the curation process can be expressed in terms of rules that embed "atomic operations" like extractors, transformations, etc. The rules can rely on abstraction/inferences for higher genericity and can also be used to produce meta-information (Splendiani, 2017).

4 Model for Intelligent Data Curation in Virtual Museum. Record De-duplication Issues

The linkage of records, the resolution of entity and the deduplication of records are only a few of the terms that describe the need for unification of multiple mentions or database records that describe the same real-world entity. If we consider the example in Table 1 (showing a single schema for simplicity), it is obvious that the records are about Alexander, but they look quite different (Walbank, 2019), (O' Brien, 2005), (Bosworth & Baynham, 2000) and (Green, 1991). Actually, all these records are correct or were correct at some point in time. It is easy for a well-qualified human to determine if such a cluster refers to the same entity, but it is hard for a machine to conduct this judgement. Therefore, more robust algorithms should be utilized to find such matches in the presence of errors, different styles of presentation and mismatches of granularity and references of time.

Table 1. Data unification at scale.

Name	Attribute	Title	Year
Alexander			356 – 323 BC
Alexander	the Great		356 – 323 BC
Alexander	of Macedon		356 – 323 BC
Alexander	III		356 – 323 BC
Alexander	III the Great		356 – 323 BC
Alexander	III of Macedon		356 – 323 BC
Alexander	(all atributes)	King of Macedonia	336 – 323 BC
Alexander	(all atributes)	Basileus of Macedonia	336 – 323 BC
Alexander	(all atributes)	Hegemon of Hellenic League	336 - 323 BC

Name	Attribute	Title	Year
Alexander	(all atributes)	Pharaoh of Egypt	332 – 323 BC
Alexander	(all atributes)	King of Persia	330 – 323 BC
Alexander	(all atributes)	Lord of Asia	331 – 323 BC

The issue is an old one. In the recent decades, the community that conducts researches has come up with many similarity functions, supervised classifiers in order to differentiate matches from non-matches, and clustering algorithms for collecting matching pairs in the same group. Current algorithms can deal with thousands of records (or millions of records partitioned in disjointed groups of thousands of records), similar to schema mapping. Taking into account the massive amount of collected dirty data – and in the context of the abovementioned schema-mapping problem, we face a number of challenges:

Challenge One: How to scale the quadratic problem (comparing every record to all other records, so computational complexity is quadratic in the number of records).

Challenge Two: How to train and build machine learning classifiers that handle the subtle similarities as in Table 1.

Challenge Three: How to engage humans and domain experts in providing training data, given the nature of the matches, which are rare in most cases.

Challenge Four: How to leverage the knowledge of all domains and previously developed rules and matchers in one integrated tool.

Talking about similarity, both the problems of schema mapping and deduplication occur after finding matching pairs (attributes in the case of schema mapping and records in the deduplication case).

Most of the building blocks can be reused and leveraged for both problems. Regarding correlation, most record matchers depend on some known schema for the two compared records; however, unifying schemas requires some type of schema mapping, even if incomplete. For this reason and many other, the solution at hand is for consolidating these activities and devising core matching and clustering building blocks for the unification of data that could: (1) be leveraged for different activities for unification (in order to avoid piecemeal solutions); (2) scale to a massive number of sources and data; and (3) have human in the loop as a guiding driver of the machine in building classifiers and applying the unification at large scale, in a trusted and explainable way. The idea is to use a human in the loop to resolve ambiguities when the algorithm's confidence on a match falls below a threshold (Ilyas, 2018).

When extracting data from different sources in cases of initial data upload or record updates, with large masses of data exists the risk of accumulating a lot of duplicate records. In this section will be presented a solution approach for deduplication. The first step is to look for mechanisms to enrich the data. In this way, extra fields can be added to each record which can assist in the deduplication process.

As a result of this step, each record has K attributes of information. The next step depends on the availability of training data. This consists of a collection of pairs of records which a human specifies as matches (i.e. duplicates) and a collection of pairs of records that are non-matches.

To this step fits a decision tree model in the following manner. For each attribute is requires a distance function, D (a1, a2), which specifies how far apart are any two values a1 and a2. In general, a distance function can be user-specified. However, for each character string attributes, Jacard and cosine similarity distance are popular metrics, and a human is asked to choose between these two. For numeric data are used arithmetic distance. For each attribute, is chosen a collection of split points based on dividing the training data into L equal sized buckets. Then, for each attribute it tries these L "split points", and avidly chooses the attribute and the split point that most accurately classifies the training data. In effect each of the L* K cases is a predicate of the form:

Attribute-I < split point => non-match

Attribute-I >= split point => match

And

Attribute-I >= split point => non-match

Attribute-I < split point => match

After that is selected the predicate that best fits the data at hand. With this "root node" chosen, continues the fit of the two second level nodes. It continues in this fashion until the benefit of additional levels is marginal or until a user-defined maximum depth, Max, is reached. In effect a decision tree model is fitted to the training data, with parameters D, L and Max.

If there is not enough training data active learning is used to get more. A "cluster review" process can also be employed. This step allows a human to review suggested matches and to correct ones that are in error. Hence, cluster review produces additional training data to refine the model used, and can be thought of as an active learning scheme.

So far are identified collections of records that it thinks represent the same entity, i.e. are duplicates. Consider one particular collection and resources that represents Alexander (356 – 323 BC) – a king (basileus) of the ancient kingdom of Macedon, as shown below.

Name:

Alexander the Great (Greek: *Αλέξανδρος ὁ Μέγας*, Bulgarian: *Александър Велики*)

Alexander of Macedon (Greek: *Αλέξανδρος ὁ Μακεδών*, Bulgarian: *Александър Македонски*)

Alexander III (Greek: *Αλέξανδρος Γ΄*, Bulgarian: *Александър III*)

Title with period of reign:

(Alexander) King of Macedonia (336 – 323 BC)

(Alexander) Basileus of Macedonia (336 – 323 BC)

(Alexander) Hegemon of Hellenic League (336 BC)

(Alexander) Pharaoh of Egypt (332 – 323 BC)

(Alexander) King of Persia (330–323 BC), etc.

Apparently, we need a canonical form for name, a resolution for several values for the title of the ruler that are attached to the name, and the recognition that we have several different periods of reign.

First are used user-specified column rules which define how to aggregate the column values in a cluster into a "golden value". Also supported are the options to "choose the most frequent value", "majority consensus", "keep all values" and "choose average

value". Based on applying these rules, each cluster of data is reduced to a simpler one with less multi-valued attributes.

Then, is examined each column, looking for patterns of values. For example, in the Alexander cluster, it removes the duplicate value "Alexander" and is left with:

III (The Third)

The Great

Of Macedon

Then, it assumes that longer strings are better than shorter ones, and forms candidate substitution rules, as follows:

III of Macedon (*Γ΄ό Μακεδών, III Македонски*)

III The Great (*Γ΄ό Μέγας, III Велики*)

The above described example is based on content units and their descriptive metadata, for which is in process the development of a virtual museum of ancient history and civilization.

Similar cases can be often observed when documenting historic facts and events in the middle ages. There are situations with substantial number of versions for the name and title of historic figures like the medieval Bulgarian ruler Asparuh, named as Asparuh/Asparukh (Bulg. *Аспарух*), Isperih (Bulg. *Исперих*), Esperih (Bulg. *Есперих*), Ispor (Bulg. *Испор*), Aspar-hruk, Batiy, etc. with several versions of title "han", "khan", "knyaz", and "tsar".

Then are performed analysis for each multi-valued field in any column that does not have the "keep all" designator. The net result is a collection of possible rules and a count of the number of times each occurs.

Finally, the rules are sorted into frequency order and present the first one to a human along with a sample of the clusters to which it applies. The human is asked to respond "yes", "no" or "maybe". The rule is automatically applied or discarded in the first two cases. In the third case, it asks a human to start tagging values as "correct" or "not correct". Based on this training data, is formed a decision tree model for the collection of clusters. This process of examining the most frequent possible rules continues until a human decides that the point of diminishing returns has occurred (Stonebraker, 2017).

5 Conclusions

A modern digital content management system has to cover a complex set of functionalities in order to operate as complete solution. The management of digital objects in a virtual museum for cultural heritage requires a well-designed architecture that embeds services for content presentation, content management, administration of user data and analysis. This set of services is interdependent and demands a high level of data integrity, which is hard to achieve when digital objects originate from disparate data sources with specific and non-standardized data formats and elements. In such a scenario, intelligent data curation that leverages machine learning to clean and unify data, is a sound approach that increases efficiency and eliminates errors of duplicate and inaccurate data. In this process are employed logistic regression, decision trees and ad-hoc models that use training data to fit a model, often assisted by human feedback and active

learning. These models undergo continuous evolution and get improved by additional techniques with each new use case.

Acknowledgements.

This work was partially supported by the Bulgarian Ministry of Education and Science under Cultural Heritage, National Memory and Social Development National Research Program, approved by DCM No 577 of 17 August 2018.

References

Blank, G. (2012). *Studyguide for the sage handbook of Internet and online research methods.* Cram101.

Borgman, C. (2015). *Big data, little data, no data: scholarship in the networked world.* Cambridge, MA: MIT Press.

Bosworth, A. B., & Baynham, E. J. (2000). *Alexander the Great in fact and fiction.* New York: Oxford University Press.

Codd, E., Codd, S., & Salley, C. (1993). *Providing OLAP to user-analysts:*. Retrieved January 27, 2019, from An IT Mandate: http://www.minet.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/Cod93.pdf

Furht, B., & Escalante, A. (2011). *Handbook of data intensive computing* (1 ed.). New York: Springer-Verlag. doi:10.1007/978-1-4614-1415-5

Goynov, M., Paneva-Marinova, D., & Dimitrova, M. (2011). Online access to the Encyclopaedia Slavica Sanctorum. In R. Pavlov, & P. Stanchev (Ed.), *Digital Preservation and Presentation of Cultural and Scientific Heritage. 1*, pp. 99 – 110. Sofia, Bulgaria: Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences. Retrieved March 10, 2019, from http://dipp.math.bas.bg

Green, P. (1991). *Alexander of Macedon, 356—323 B.C.: A historical biography.* Los Angeles: University of California Press.

Ilyas, I. (2018). Data unification at scale: data tamer. In M. Brodie (Ed.), *Making Databases Work* (pp. 269-277). New York, NY, USA: Association for Computing Machinery and Morgan & Claypool. doi:10.1145/3226595.3226619

Luchev, D., Paneva-Marinova, D., Pavlova-Draganova, L., & Pavlov, R. (2013). New digital fashion world. In B. Rachev, & A. Smrikarov (Ed.), *ACM International Conference Proceeding Series (14th International Conference on Computer Systems and Technologies, CompSysTech'13). 767*, pp. 270-275. NY, USA: The Association for Computing Machinery.

Miller, R. (2014). Big data curation. *20th International Conference on Management of Data (COMAD).* Hyderabad, India: Computer Society of India (CSI).

O' Brien, J. (2005). *Alexander the Great: The Invisible Enemy: A Biography.* London ; New York: Routledge.

Paneva-Marinova, D., Goynov, M., & Luchev, D. (2017). *Multimedia digital library: Constructive block in ecosystems for digital cultural assets. Basic functionality and services.* Berlin, Germany: LAP LAMBERT Academic Publishing.

Paneva-Marinova, D., Pavlov, R., & Rangochev, K. (2010). Digital Library for Bulgarian Traditional Culture and Folklore. *The Proceedings of the 3-rd International Conference dedicated on Digital Heritage (EuroMed 2010)* (pp. 167-172). Lymassol, Cyprus: ARCHAEOLINGUA.

Pavlov, R., Paneva-Marinova, D., Goynov, M., & Pavlova-Draganova, L. (2010). Services for content creation and presentation in an iconographical digital library. (P. L. Stanchev, A. Eskenazi, & R. Pavlov, Eds.) *International Journal "Serdica Journal of Computing", 4*, 279-292.

Pavlova-Draganova, L., Paneva-Marinova, D., Pavlov, R., & Goynov, G. (2010). On the wider accessibility of the valuable phenomena of Orthodox iconography through digital library. In M. Ioannides, D. Fellner, A. Georgopoulos, & D. Hadjimitsis (Ed.), *Proceedings of the 3rd International Conference dedicated on Digital Heritage (EuroMed 2010)* (pp. 173-178). Lymassol, Cyprus: ARCHAEOLINGUA.

Splendiani, A. (2017, September 28). *AI for data curation. Yes, can we?* Retrieved from https://www.slideshare.net/sergentpepper/artificial-intelligence-in-data-curation

tamr. (2019, January 14). *Agile data mastering raising expectations for master data management (MDM).* Retrieved from http://www.tamr.com: http://www.tamr.com/wp-content/uploads/2019/01/Tamr_WP_Agile-Data-Mastering-_01-14-19.pdf

Walbank, F. (2019, February 08). *Alexander the Great : King of Macedonia*. Retrieved March 18, 2019, from Encyclopædia Britannica: https://www.britannica.com/biography/Alexander-the-Great

Received: June 01, 2019
Reviewed: June 25, 2019
Finally Accepted: July 05, 2019