Developing an object-oriented model database of the smart library of the works of medieval scientists of Uzbekistan

Ilkhom Bakaev*[a], Tursun Shafiev[a], Xamza Eshankulov[a], Gulsina Ataeva[a]
[a]Bukhara State University, 11, M. Iqbol, Bukhara City, 200100, Uzbekistan

ABSTRACT

The article considers the issue of developing an object-oriented database model of the electronic library of the works of medieval scientists of Uzbekistan. Based on the formulated requirements for the software platform and its information content, the main objects, their attributes and permissible connections between objects were determined. The ER model and the corresponding logical and physical database models were developed taking into account common approaches to the formalization of bibliographic information and the general trend in the evolutionary development of the architecture and functionality of electronic libraries. The developed object-oriented model and database structure made it possible to optimize the composition and structure of the source information for the tasks of the electronic library system. Subsequent analysis of the developed model from the point of view of how successfully it allows solving the tasks of readers showed that the designed database, with minimal redundancy, has good data consistency and operability, has the properties of flexibility and extensibility to adapt to more complex tasks that may arise with further development of the system.

Keywords: data model, object-oriented model, smart library, library, entity relationship model, conceptual model, bibliographic, metadata models

1. INTRODUCTION

In the 21st century, society has entered a qualitatively new era of life which is accompanied by the digitalization of all life of mankind. In this era society has faced with the task of preserving the historical, cultural and spiritual heritage for the coming generation.

The preservation of the culture of peoples occurs not only through national traditions and cultural monuments of antiquity, but also through the preservation of literary works that reflect the culture and life of peoples. One of the centers of the national cultural heritage is the treasuries of literary works. Uzbekistan has the richest cultural and literary heritage preserved in the museums and libraries of the Republic. Everyone knows the works of thinkers of the East, a huge part of which is the cultural heritage of the Republic of Uzbekistan, stored in museums, libraries and archives of the country. They contain the works of medieval authors who lived from the 9th to the 19th century.

Using historical manuscripts for the reliability of research, scientists recreate historical pictures from the past of the Muslim East and Central Asia. These primary sources contain a lot of information used in historical, philosophical, sociological and cultural studies. The libraries of Uzbekistan store the manuscripts of such educators as Imam at-Termizi, Imam al-Bukhari, Bahauddin Naqshband, Burkhaniddin Margilani, Najmiddin Kubro, Ahmad Yassavi, Abu Raykhan al-Biruni, Muhammad al-Khwarizmi, Mirzo Ulugbek, Abu Ali Ibn Sina and others which to this day are of interest to scientists and researchers.

The Government pays special attention to the preservation of cultural values encouraging the active use of ancient manuscripts in scientific and cultural research. In this regard, in 2017, the Decree of the President of the Republic of Uzbekistan “On measures to further improve the system of storage, research and propaganda of ancient written sources” was adopted.

*tursun@buxdu.uz; phone 998997098020

Third International Conference on Digital Technologies, Optics, and Materials Science (DTIEE 2024),
edited by Khamza Eshankulov, Arthur Gibadullin, Proc. of SPIE Vol. 13217, 1321710
 · 0277-786X · doi: 10.1117/12.3036772

The Decree of the President of the Republic of Uzbekistan dated 07.06.2019 No. PP-4354 "On the further improvement of information and library services for the population of the Republic of Uzbekistan" states that "The goals and objectives of the development of information and library activities in modern conditions of reforming the socio-economic sphere of the republic should correspond to the changes taking place in the country and international practice". Readers of modern libraries are advanced due to the possibilities of modern information technologies that allow them to provide remote access to resources. The means of digitizing valuable scientific sources allow you to increase the possibilities of their use for in-depth study without limiting readers in terms and quantity. Digitization of resources is a great achievement of modern information technologies.

Today the basis for providing information resources online is electronic libraries and they are based on a clear organization of electronic resources. Electronic libraries provide structured access to electronic resources equipped with modern search mechanisms based on artificial intelligence. Today, a newfangled trend in information technology is the use of neuro systems. Electronic libraries, in comparison with traditional libraries, provide faster and better service.

2. LITERATURE REVIEW

The creation of digital libraries began in the early 1970s when M. Hart (USA) set out to provide the whole world with electronic literature. Then the Gutenberg project was launched the motto of which was to "destroy the barriers of ignorance and illiteracy".[1] Called the Guttenberg project which is a means of free distribution of electronic copies of literary works around the world, the oldest part of the modern free culture movement. At the same time, the process of automating the library catalog was begun which facilitated the search for the necessary literature.

In the modern world, in almost all parts of the world, work is underway to create digital libraries[2]. Almost all countries are digitizing information resources and providing open access to these resources[3]. Projects for the creation of digital libraries may have their own specification, for example, some create libraries for specific subject areas[4], while others are built according to the type of publications (books, newspapers or magazines)[5]. There are electronic libraries that ensure the safety of intellectual property with copyright protection[6].

One of the largest catalogs is the international copyright registry maintained by the US Copyright Office. The Bureau is part of the Library of Congress, the world's largest library with over 160 million items.

The Copyright Registry of the United States Copyright Office is archival storage of all registered works - objects copyright since 1870.

One of the international projects that ensure the development and informatization of society is the World Digital Library (World Digital Library). This project was implemented with the support of the international UNESCO organizations. The World Digital Library provides access to digital resources such as rare books, manuscripts, photographs, maps, architectural drawings, snapshots, musical scores and films. The library contains the most valuable digitized materials on the history and culture of different countries. At World Digital Library there is free access to many materials on many world languages that represent world culture.

This library provides opportunities to learn interactive maps and timelines on the history of the whole world, ancient engravings and manuscripts, and commented photographic documents. The documents are available in 133 languages of the world.

Among the many e-Library projects, one should also note the European project of a digital library that reflects various aspects of European culture – Europeana. This project provides access to European information resources held in museums, libraries and archives in Europe.

Modern Uzbekistan also has many digital resources including textual and multimedia data organized at a high level. However, there are no projects in Uzbekistan that would have specialized information resources, united by themes on the Uzbek and Central Asian historical cultural heritage[7].

This paper considers an attempt to implement an electronic library project that contains information resources in the form of works of medieval thinkers and theologians of Uzbekistan[8], consisting of digitized copies of historical manuscripts, modern publications with translations of medieval books, thematic scientific works of contemporary

scientists, bibliographic information, encyclopedic data and reference information on various objects of cultural heritage of the country in general.

3. METHODOLOGY

The process of developing an e-Library system requires the preliminary formation of concepts about books, manuscripts, events, facts and places that the system being developed will actually operate on. In order to bring these concepts to a specific data model, it is necessary to replace them with the corresponding ones.

The most convenient tool for such data representation, regardless of the software that implements it, is the entity-relationship model or the ER model which is based on semantic information about real-world objects.

From the analysis of thematic publications, it can be seen that the authors offer a variety of approaches and methods for the implementation of conceptual models of e-Library databases. In particular, one popular approach is to use the FRBR (Functional Requirements for Bibliographic Records) family of models or their derivatives.

For example[9], the authors recommendations present a conceptual transition from document-centric metadata to content-oriented metadata. In technical terms, this implies a transition from MARC (machine-readable cataloging) standards to standards for describing information resources, in particular RDA (Resource Description Access).

The authors analyzed the scientific literature on the use of modern metadata models such as FRBR (Functional Requirements for Bibliographic Records) and RDA (Resource Description Access) and linked data technologies in libraries. The main conclusions of the study are that the currently used library data description standards have inherited significant limitations that are inherent in the traditional cataloging system. That is, the resulting metadata is human-readable and poorly suited for efficient machine processing such as while implementing smart search. The authors propose new ways to solve the problems of formalizing bibliographic data with an emphasis on the application of modern principles of linked data, metadata exchange, data mining, multifaceted navigation and RDF (Resource Description Framework)/XML (eXtensible Markup Language) and SPARQL (Protocol and RDF Query Language) technologies in library practice.

The authors[10] present a data and metadata model developed for use in a music digital library to support the search and navigation of music content in various formats.

To implement the conceptual model of the library database, the authors used a modification of the FRBR (Functional Requirements for Bibliographic Records) data model. It is noted that the proposed multidimensional model provides a number of advantages over the traditional flat database structure based on MARC (Machine-Readable Cataloging). As a result, the digital library of musical works is characterized by increased completeness and accuracy of search results.

More traditional approaches to the design of e-Library databases are still quite popular, especially if the automation of library processes is at least to some extent affected. Thus, for the development of the digital library system of the Henan Polytechnic University, the authors[11] designed a relational database data focused on the storage of scientific and technical publications. The authors have made efforts to minimize the redundancy of data (entities, their attributes and relationships) while maintaining the effective performance and security of the system.

To achieve the goal of the project to create a "smart library", the ER (Entity-Relationship) model as well as the corresponding logical and physical database models were developed taking into account the approaches noted above and the general trend in the evolutionary development of the e-Library architecture and functionality.

The requirements for the e-Library system being developed in the form of the system core (backend), web interface (frontend) and API (Application Programming Interface) for mobile clients, formulated at the preliminary design stage imply the following main functions:

- Input of authoritative data including an extended biography of personalities.
- Input of related encyclopedic information (existing monuments, mausoleums, places of interest, geotags, etc.).
- Input of bibliographic data of printed works, manuscripts, other documents.

- Thematic classification of works in general and their content in particular (for example, the classification of hadiths in collections).
- Presentation of e-Library resources through the interface of a web application, a mobile application. Support for the presentation of information using augmented reality technology.
- Checking the knowledge of users in the form of a system of test tasks.
- User authentication and authorization.
- Adding comments, evaluating the relevance and usefulness of the content.
- Automatic formation of thesauri with indication of semantic relations between terms.
- Automatic formation of galleries of files, images, audio and video recordings.
- Intellectual search in e-Library content (textual as well as based on optical recognition of objects, for example, book covers, geolocation, buildings, etc.).
- Objects, for example, book covers, geolocation, buildings, etc.).

Thus, the conceptual model of e-Library has the form shown in Figure 1.

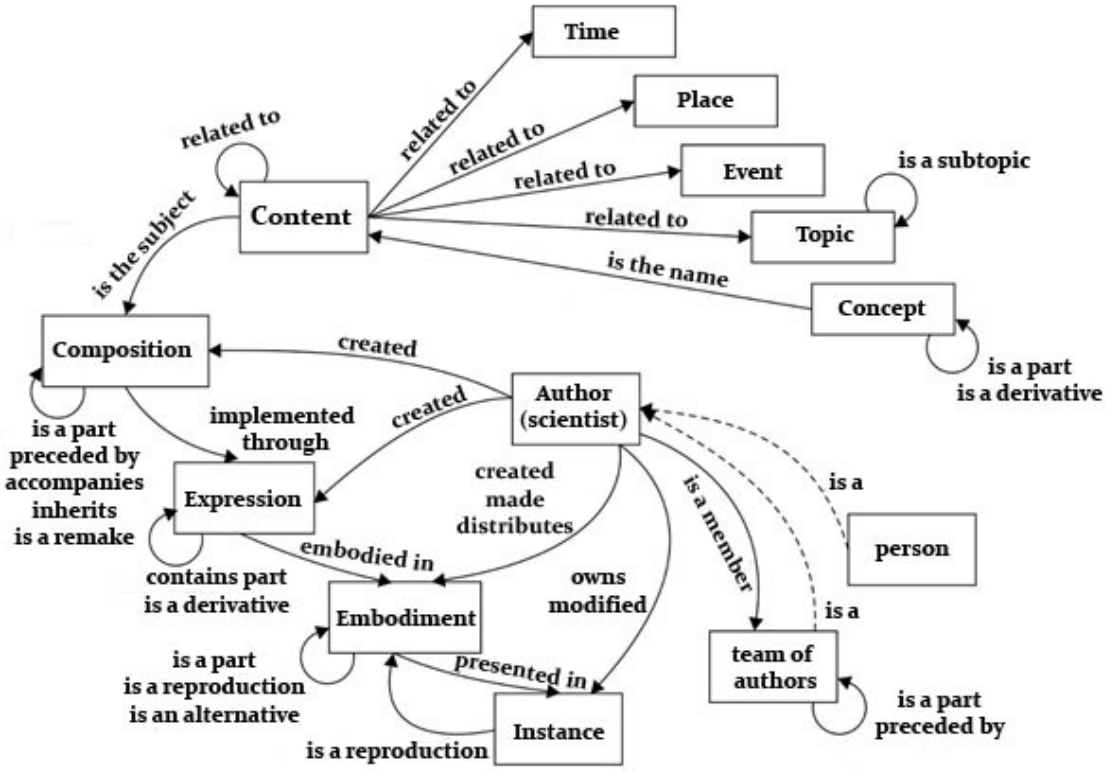

Figure 1. ER-model of the digital library.

Diagram of the logical model of the e-Library database with a list of entities and relationships in Figure 2.

The physical model of the EL database contains the following main tables (Table 1).

Table 1. The main entities of the ER model of the digital library.

Entity name	Definition
scientists	Information about authors
scidirections	Scientific direction of the author's works
scientists_scidirections	Table-decoupling for “many-to-many” relationship transformation for "Scientist" and "Direction" tables
contents	Information resources or separate units of DL content (a work, an independent part of a work, related data, etc.).
themes	Thematic section
themes_contents	Table-decoupling for converting the many-to-many relationship between information resources and topics.
scientists_contents	A table-decoupling that establishes a link between authors and information resources.
Books	Bibliographic data about editions, manuscripts, other documents.
books_contents	Table-decoupling for “many-to-many” relationship transformation for Book and Contents tables.
objects	Digital objects (files) in DL are text files, images, audio, video, animation, 3D models, etc.).
objtypes	Types of digital objects.
contents_objects	Table-decoupling for information resources and related digital objects.
Comments	Commentary on information resources
users	Users
Roles	User roles
terms	Terms found in all content items.
terms_contents	Connection of terms with information resources.
searches	Records of search queries and related search results.
tests	Test tasks
userhistories	History of user actions
Publishers	Publishers

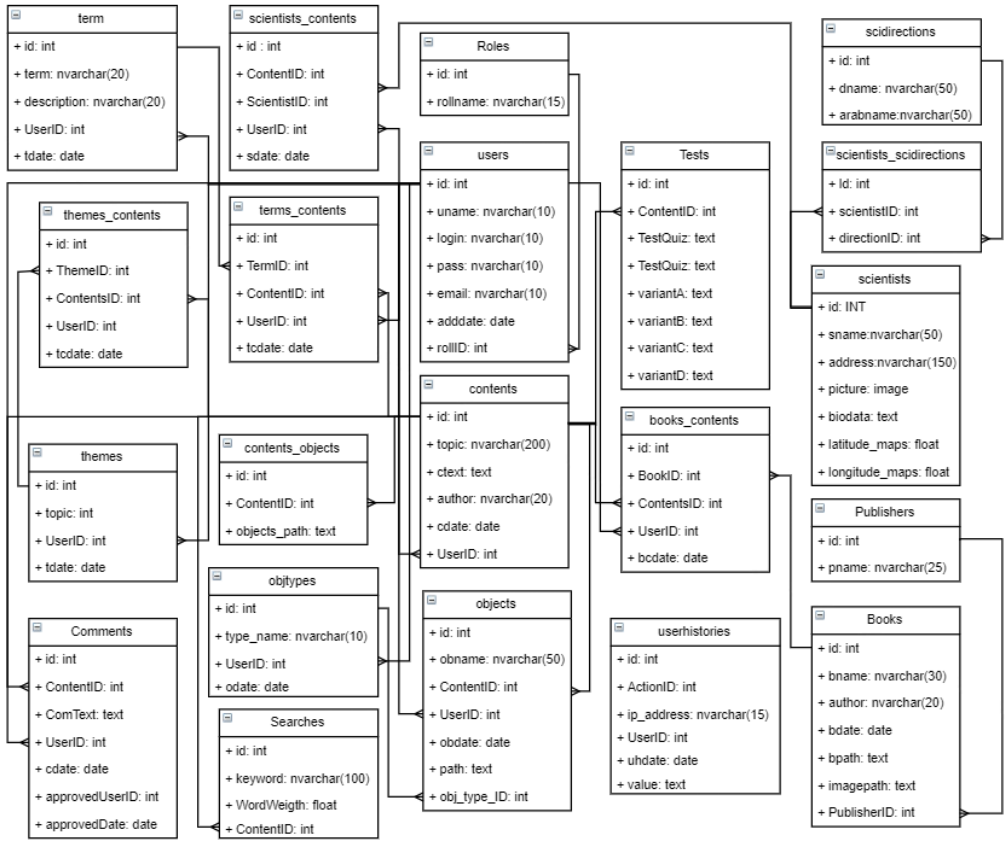

Fig.2. E-library database schema.

Description of relationships. Bilateral description of links between the main entities that have informational value for EL users. The concept of content. In order to optimize the search and increase the relevance of the results in response to search queries, each individual document (book, manuscript, video, etc.) within the e-Library is divided into constituent "atomic" parts - an information resource or content. The set of contents can be classified according to such features as: concept, topic, event, time and place. Each content is defined by a vocabulary list (thesaurus) that make up this content. In the future, to implement the search mechanism, each element of the set of words is assigned characteristics: frequency of occurrence, inverse content frequency, word weight, ranking which increase the relevance of answers to search queries.

Authorship identification. When forming data about scientists, it is necessary, for example, to take into account the fact that there is a fairly large volume of publications devoted to medieval theologians. Works related to the life and work of Imam al-Bukhari can be written by many authors. Accordingly, the authorship of a single content may belong to one or several authors, so the relationship between the entities "scientists" and "Contents" in the database is organized as follows (Figure 3).

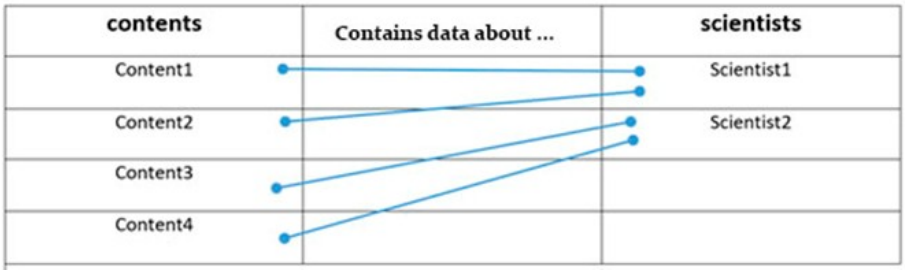

Fig.3. Relationship between Scientist and Contents entities.

Content of books (incarnations). To search and display the contents of books, the general text is preliminarily divided into separate parts (for example, sections, chapters, titled paragraphs). In the case of a collection of hadiths, the contents may be divided into separate hadiths. Each such atomic fragment of the text of the book is the content. Such a division allows you to classify content by topic (Figure 4) or any other attribute indicated above (Figure1). In addition, a dictionary is generated for each piece of text in the book mainly to determine the weight of words in relation to content.

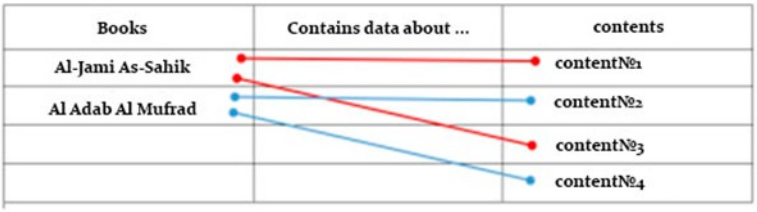

Fig.4. Relationship between Books and Contents entities.

Direction of activity of authors (scientists). Implementation mechanism filtering authors in the search, implies the possibility of selecting them from the database on the basis of specialization or scientific direction. An example of the connection "scientist - scientific direction" is shown in Figure 5.

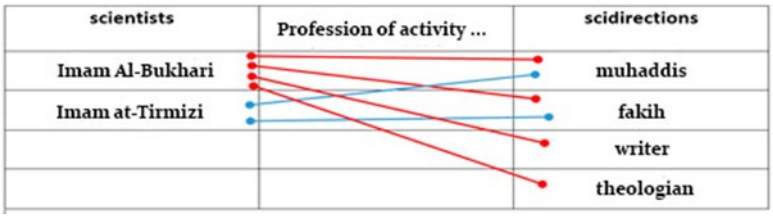

Fig.5. Relationship between Scientist and Direction entities.

Content subject. Each individual content, in a similar way, can be correlated with a specific thematic section or with several at once, so the relationship between entities also has a many-to-many type (Figure 6).

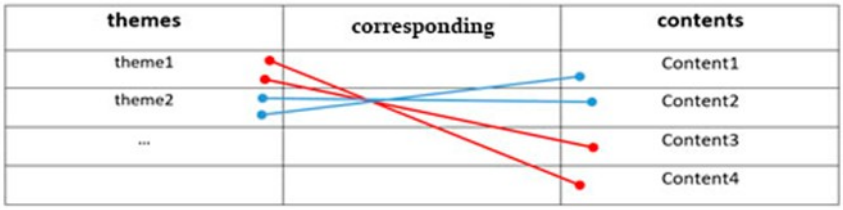

Fig.6. Relationship between Theme and Content entities.

Dictionary (thesaurus) by content. The peculiarity of historical manuscripts or highly specialized publications (for example, hadis studies) is such that their texts contain a large number of ancient words and specific terms. The language and terminology used in them are usually unfamiliar to a reader. The extraction and ordering of terms are carried out to provide content search for special terms as well as to organize an explanatory dictionary (Figure 7).

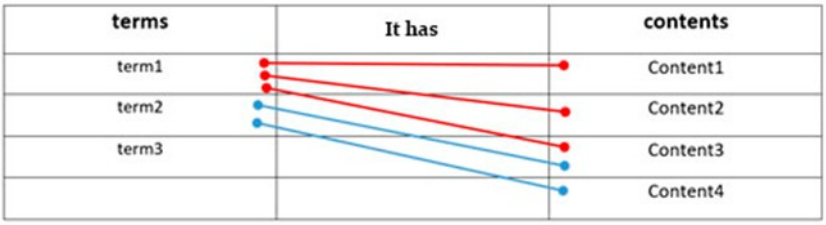

Fig.7. Relationship between the Terms and Content entities.

User comments on e-Library content. As part of the e-Library, authorized users are provided with the opportunity to write comments on the content being viewed (opinion, questions, etc.). In this case, the relationship between entities is of the one-to-many type (Figure 8).

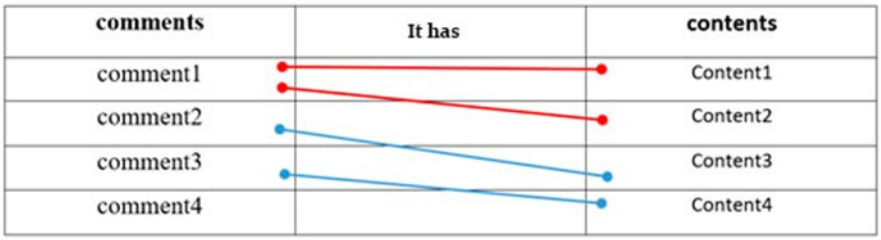

Fig.8. Relationship between the Comments and Contents entities.

Keywords. As with specific terms, keywords also serve to optimize searches. However, in this case, all words are extracted from the text of each content to create a corpus of texts, calculate n-grams and determine the weight of words and phrases in relation to the content. A certain group of words can belong to the same content. Therefore, the connection is organized as "one to many" (Figure 9).

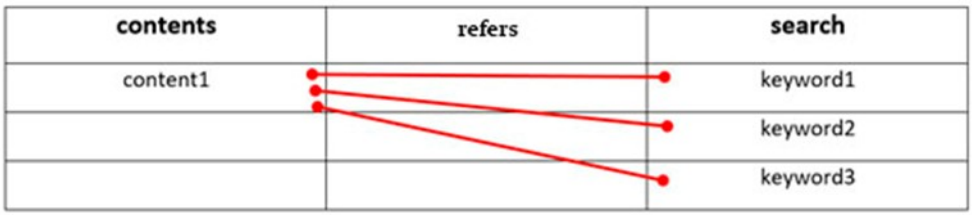

Fig.9. Relationship between the Contents and Searches entities.

Content objects. In addition to textual information, content may contain another type of data such as an image, video, animation and the list could go on. The functionality of the developed digital library assumes the presence of searching and viewing digital objects outside the presentation of content, that is, in the form of independent collections. In addition, digital objects themselves can act as a search feature, in particular, when they are used as AR tags (Augmented Reality Tags).

To organize the search for content on digital objects, the following link is used (Figure 10).

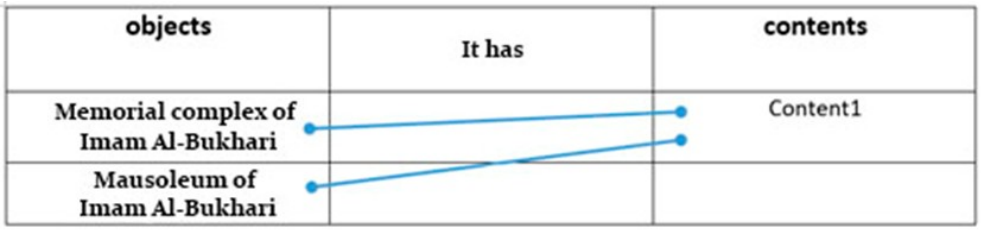

Fig.10. Relationship between Objects and Contents entities.

Output data of publications. Entities "Instance" from fig. 1 corresponds to a specific edition or document if the "Incarnation" is unique. That is, it can be a copy of a book, CD, photograph, archival document, reproduction or electronic file. For example, in table 1 you can see one of the main such entities - "Book", that is, a book edition. A book publication has its own set of bibliographic imprints which include the publisher. The Publisher entity is used as a search criteria and data sorting in the user interface. Each publishing house may own one or many publications (Figure 11).

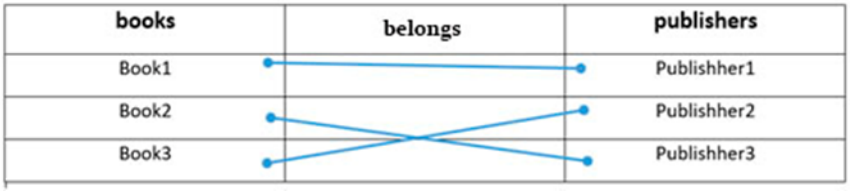

Fig.11. Relationship between entities Books and Publishers.

Interrelations between other objects were organized in a similar way, including those links that relate to the functions of service subsystems: "Users - System logs", "Users - Comments", etc.

4. CONCLUSION

Thus, on the basis of the formulated requirements for basic (minimum necessary) information for solving the tasks of e-Library users, the main objects, their attributes and admissible links between objects were determined. The developed object-oriented model and the structure of the "smart library" database made it possible to optimize the composition and structure of the initial information for the given e-Library tasks:

- Search and collection of information about content that match the search criteria.
- A clear understanding of the nature of the found content and the distinction between similar content.
- Selection of content that meets the needs of the user.
- Gaining access to the content of the selected content.
- The study of contents, using the links between them in order to correctly place them in the user's context.

An analysis of the resulting model in terms of how successfully it allows solving user problems using a preliminary version of the e-Library (or digital library) system showed that the designed database with minimal redundancy has good data consistency and performance. Also, the developed model is sufficiently flexible and extensible to be adapted to solve more complex data processing tasks that are possible with the further development of the system.

REFERENCES

[1] Hancoc, T., "Impossible Thing2: Comprehensive Free Knowledge Repositories like Wikipedia and Project Gutenberg," Free Softw. Mag. (2008).

[2] Aithal, P. S., "Smart Library Model for Future Generations," Int. J. Eng. Res. Mod. Educ. 1(1), 693–703 (2016).

[3] Cao, G., Liang, M. and Li, X., "How to make the library smart? : The conceptualization of the smart library," Electron. Libr. 36(5), 811–825 (2018).

[4] Baryshev, R. A., Verkhovets, S. V. and Babina, O. I., "The smart library project: Development of information and library services for educational and scientific activity," Electron. Libr. 36(3), 535–549 (2017).

[5] Yang, X., He, D., Huang, W., Zhou, Z., Ororbia, A., Kifer, D. and Giles, C. L., "Smart Library: Identifying Books in a Library using Richly Supervised Deep Scene Text Reading" (2016).

[6] Simovic, A., "A Big Data smart library recommender system for an educational institution," Libr. Hi Tech 36, 498–523 (2018).

[7] Eshankulov, K. and Bakayev, I., "Information exchange in multi-agent systems," Proc. Vol. 13065, Third Int. Conf. Opt. Comput. Appl. Mater. Sci. (CMSD-III 2023), 1–7 (2024).

[8] Ravshanov, N. and Bakaev, I., "Development of an object-oriented model of 'smart library' database," Probl. Comput. Appl. Math 5(23), 117–129.

[9] Alemu, G., Stevens, B., Ross, P. and Chandler, J., "Linked Data for libraries," New Libr. World 113(11/12), 549–570 (2012).

[10] Minibayeva, N. and Dunn, J., "A Digital Library Data Model for Music," Proc. Second ACM/IEEE-CS Jt. Conf. Digit. Libr., 154–155 (2002).

[11] Zhai, H., Shi, H. and Zhai, R., "The Design and Implementation of Database on Library Management Information System," Proc. 2nd Int. Symp. Comput. Commun. Control Autom. (ISCCCA 2013), 834–836, Atlantis Press.