Original article

Software PhotoCoasts of Crimea

M. P. Vetsalo *, E. A. Godin, E. A. Isaeva, L. K. Galkovskaya

Marine Hydrophysical Institute of RAS, Sevastopol, Russia
* e-mail: mvetsalo@mhi-ras.ru

Abstract

The article describes the information retrieval system PhotoCoasts of Crimea developed by the staff of the Oceanographic Data Bank group based on the concept of the software PhotoCoasts to systematize and catalogue the collection of digital images of the Crimean Peninsula coasts. The system also ensures effective work with this collection while conducting scientific research. The software system is written in the Python programming language. The application interface is developed using the tkinter package. The system core is a catalogue of meta-information on photosurvey objects. The catalogue is based on faceted classification and includes descriptive facets "Date and Time", "Type of Photosurvey" and specialised facets "Geographic Region", "Coast Genetic Type". The method of extended Boolean retrieval was applied to form the query results in the software system. New images are uploaded and metadata of existing catalogue elements are edited in the metadata editor. Work with the geoinformation part of the metadata base is performed in the geodata editor. The developed software has a significant potential for further evolution and after appropriate adjustment can be used for work with coast images of other regions. It also allows systematisation and classification of image collections in various fields.

Keywords: coasts, images, visualisation, systematisation, cataloguing, faceted classification, information retrieval system

For citation: Vetsalo, M.P., Godin, E.A., Isaeva, E.A. and Galkovskaya, L.K., 2024. Software PhotoCoasts of Crimea. *Ecological Safety of Coastal and Shelf Zones of Sea*, (4), pp. 131–140.

Программный продукт «ФотоБерега Крыма»

М. П. Вецало *, Е. А. Годин, Е. А. Исаева, Л. К. Галковская

Морской гидрофизический институт РАН, Севастополь, Россия
* e-mail: mvetsalo@mhi-ras.ru

Аннотация

Описывается информационно-поисковая система «ФотоБерега Крыма», разработанная сотрудниками группы «Банк океанографических данных» на основе концепции программного продукта «ФотоБерега» для систематизации и каталогизации коллекции цифровых изображений берегов Крымского полуострова, а также обеспечения эффективной работы с этой коллекцией при проведении научных исследований. Программная система написана на языке программирования *Python*. Интерфейс приложения разработан с использованием пакета *tkinter*. Центральной частью системы является каталог метаинформации объектов съемки, который построен на основе фасетной классификации и включает описательные фасеты «Дата и время», «Вид съемки» и специализированные фасеты «Географический регион», «Генетический тип побережья». Для формирования поисковой выдачи в программной системе применен метод расширенного булева поиска. Загрузка новых изображений и редактирование метаданных существующих элементов каталога выполняется в редакторе метаданных. Работа с геоинформационной частью базы метаинформации осуществляется в редакторе геоданных. Созданный программный продукт имеет значительный потенциал для дальнейшего развития и после соответствующей настройки может быть использован для работы с изображениями берегов других регионов, а также для систематизации и классификации коллекций изображений в самых различных областях.

Ключевые слова: берега, изображения, визуализация, систематизация, каталогизация, фасетная классификация, информационно-поисковая система

Introduction

The transition from drawing to photography and film shooting with high resolution has rendered images the most durable medium for capturing the state of natural objects. When image acquisition is undertaken with sufficient accuracy in terms of place and time, a comparison of images facilitates analysis of changes in the state of objects under the influence of natural and anthropogenic factors. However, a significant proportion of images has not yet been digitised. These images are often stored in various collections and lack proper systematisation and cataloguing, which can result in their loss and limited accessibility to researchers. This situation is particularly problematic for images of seashores,

including the Crimean coast. It can be remedied by organising the digitisation of images and creating special software to work with them.

Currently, it is evident that there is a substantial number of software available for the creation of digital image catalogues. However, these products often prove to be inadequate in terms of their efficiency for the purpose of scientific research. Analysis of open source information [1), 2), 3)] and personal user experience have demonstrated that the catalogue in most existing software products was typically constructed on the basis of the user computer file system hierarchy (Adobe Bridge, ACDSee Photo Studio, etc.) or its own hierarchical structure (folders in Adobe Lightroom and Corel AfterShot Pro, collections in darktable, digiKam albums, etc.). This approach is inherently disadvantageous due to the limitations of such classifications. The rigid structure of the hierarchy is a particular problem as it makes it difficult to include new levels of division. The classification process becomes excessively cumbersome and challenging to utilise when the number of levels is substantial, yet it lacks sufficient informative value when the number of levels is minimal. Moreover, such products are primarily oriented towards the management of image metadata (i.e. image parameters, geolocation, etc.), yet their capabilities in determining the characteristics of the photosurveyed object are not sufficiently developed to enable the execution of research tasks. It is also noteworthy that a considerable proportion of existing software is commercial, with closed sources precluding modification to align with user needs. A further impediment to the procurement and utilisation of prevailing commercial digital image cataloguing software pertains to the sanctions policy of multiple states against the Russian Federation.

Taking into account the disadvantages of the existing software described above, the staff of the Shelf Hydrophysics Department and the Oceanographic Data Bank group of Marine Hydrophysical Institute of RAS developed the software PhotoCoasts concept including following general approaches to the creation of software for visualisation, systematisation and cataloguing of digital images for scientific research:

– provision of maximum independence from external factors, use of freely distributed open source components only;

– work with the catalogue (including its extensibility) taking into account the specifics of using images in solving scientific problems;

– geoinformation base support;

– support for working with digital image metadata (time of photosurvey, geopositioning);

– mass import of images.

[1)] Sazhko, D., 2018. [10 Apps to Organize a Photocollection]. *Lifehacker*. [online] Available at: https://lifehacker.ru/kak-organizovat-kollekciyu-fotografij/ [Accessed: 25 November 2021] (in Russian).

[2)] @dnowicki, 2014. [Overview of Windows Applications for Keeping Photo Archives in Order]. *Habr*. [online] Available at: https://habr.com/ru/post/226123/ [Accessed: 25 November 2021] (in Russian).

[3)] Shlyakhtina, S., 2004. [Cataloguing and Storage of Digital Images]. *ComputerPress*. [online] Available at: https://compress.ru/article.aspx?id=12397 [Accessed: 25 November 2021] (in Russian).

The article presents the findings of research conducted on the implementation of this concept in the development of the specialised information retrieval system PhotoCoasts of Crimea. This system has been optimised for solving a practical problem, namely the systematisation and classification of the collection of digital images of the Crimean Peninsula shores. These images have been collected by the MHI Shelf Hydrophysics Department for a period exceeding a century and a half, and the article discusses the effective work that has been carried out with them.

Approaches and methods

The system core is constituted by a catalogue of meta-information on photo-survey objects. The catalogue is based on faceted classification, a method that has been demonstrated to possess greater semantic power than hierarchical classification [3, 4]. At the inception of catalogue formation, a set of concepts (terms) is defined, which are required to describe the catalogue element (image). Then the terms are subjected to semantic or other grouping into so-called facets. The classification of catalogue elements, in turn, is not predetermined, but is constructed by selecting elements from facets and forming a linear chain of them, called a facet formula. The position of each facet in the facet formula is strictly fixed. In this context, the task of information retrieval within the catalogue can be addressed by determining the sequence of transformation of user information needs into a facet formula (retrieval query) and by considering the classification obtained from this formula as a retrieval output.

The method of extended Boolean retrieval was applied to form the query results in the software system [5], when the result is determined by a logical expression formed on the basis of a user request. The retrieval query (facet formula) is converted into a logical expression, which is applied to the metadata of each catalogue element. In order to increase the speed of retrieval formation, reverse indexing of the image metadata database [5, 6] by individual facets is applied. Acceleration of application work when using the index is achieved by reducing significantly (in general) the number of catalogue elements involved in operations and switching to the use of operations on sets instead of performing logical operations (Boolean retrieval) for all images from the program database.

Implementation

To implement the software system, the Python programming language was selected, as it has both a well-developed standard library and sufficient number of freely accessible third-party open source libraries.

The catalogue of the system PhotoCoasts of Crimea includes descriptive facets "Date and Time" (D), "Type of Photosurvey" (T) and specialised ones "Geographic Region" (R), "Coast Genetic Type" (G). The facet "Date and Time", in turn, is divided into three sub-facets: "Date of Period Beginning" (D_B), "Date of Period End" (D_E) and "Seasonality of Photosurvey" (S). The resulting facet formula for the applied classification is as follows:

$$<D_B : D_E : S : R : T : G>.$$

A part of facets contains a finite set of concepts defined during the application development, determined by the semantics of the corresponding feature or the specifics of the system under development. For example, the generally accepted division of the annual cycle into calendar seasons {'winter', 'spring', 'summer', 'autumn'} is used as descriptions for photosurvey seasonality (sub-facet "Seasonality of Photosurvey"), and the content of the facet "Coast Genetic Type" is determined by the geomorphology of the Crimean Peninsula and corresponds to monograph [7].

The content of the facet "Geographic region" and sub-facets "Date of Period Beginning" and "Date of Period End" is not fixed in the application code and can be modified by the user. In particular, the system provides possibility to work with data on photosurvey region, which are the ground for geoinformation base.

To store the catalogue and data of the geoinformation part of the software, the embedded SQL-oriented freely distributed open source DBMS SQLite3 [4] was used. The table Picture contains the basic metadata of a photograph and the identifiers of its relationships to catalogue facet elements. When displaying the catalogue structure, a separate table has been allocated for each facet and sub-facet "Seasonality of Photosurvey" (Fig. 1). Data of sub-facets "Date of Period Beginning" and "Date of Period End" do not exist without a corresponding picture and are implemented as picture attributes in the table Picture. In general, in order to obtain a normalised database, the relationship between a particular photograph and catalogue facets is constructed in one of the following ways:

– for facets with a cardinality of 1 : 1 or 1 : N, the link is organised by including the corresponding identifier in the table Picture;

– for facets with a cardinality of M : N, an additional table is created.

The mapping of the hierarchical facet "Geographic Region" to the relational structure of the database is performed using the list of adjacency from the table GeoRegion [8, 9].
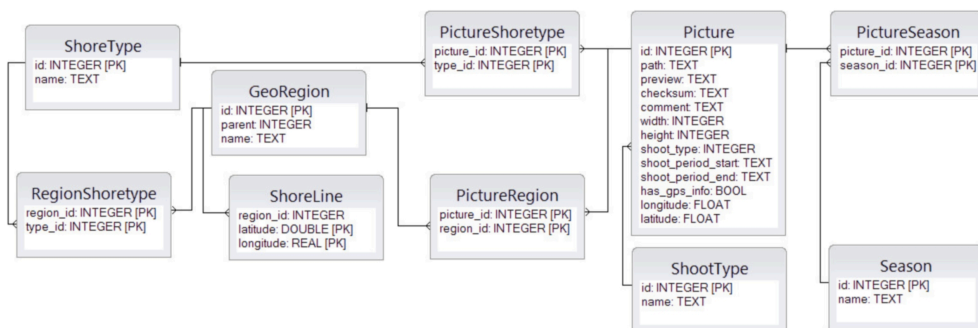

F i g . 1. The scheme of the database for the software system PhotoCoasts of Crimea

---

The software system uses a hierarchy in the hard disc file system to store the original data. Each photograph corresponds to a separate directory containing the original digital image and a thumbnail.

The application interface is developed using the tkinter package [5] and consists of main window, metadata editor and geodata editor (Fig. 2).

The user defines the required set of image characteristics by forming a query with the use of the retrieval panel located on the left side of the main application window. The retrieval panel contains a separate section for each facet of the catalogue and makes it possible to compose a query in an intuitive way without using specialised query languages. The retrieval result is displayed in the working area of the main window and visualised in two main modes. Library mode provides an estimate of the retrieval and an overview of the retrieval results. In its turn, view mode provides an opportunity to examine each image and its associated meta-information in detail. Since digital images obtained using modern photosurvey equipment can have high resolution, which significantly exceeds the resolution of a computer monitor, the view mode has a function of image scaling including automatic scaling to fit the application window size.

The uploading of new images and the editing of the metadata of existing catalogue items are performed in the metadata editor (Fig. 3). It can be used to change the coordinates of the photosurvey location, its geographical region, date, season, photosurvey method, coast genetic type as well as to make adjustments to the photograph
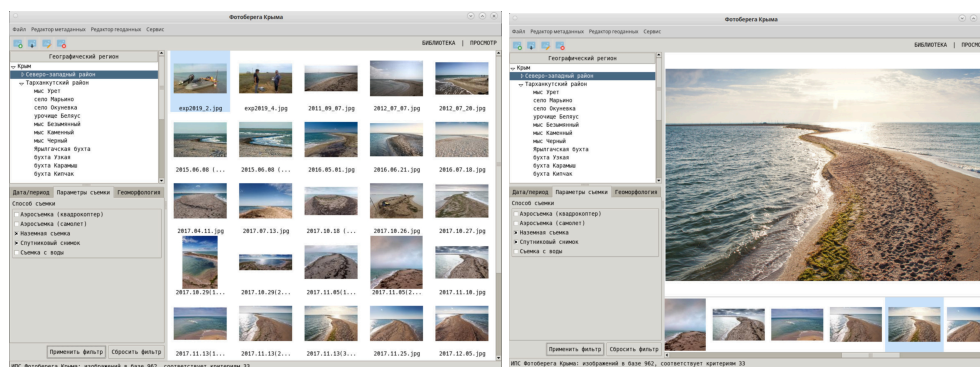
a   *b*

F i g . 2 . The user interface of the application main window: LIBRARY mode (*a*) and VIEW mode (*b*)

---

[5] Python Software Foundation: Graphical User Interfaces with Tk. *Python*. 2001. [online] Available at: https://docs.python.org/3/library/tk.html [Accessed: 2 December 2024].
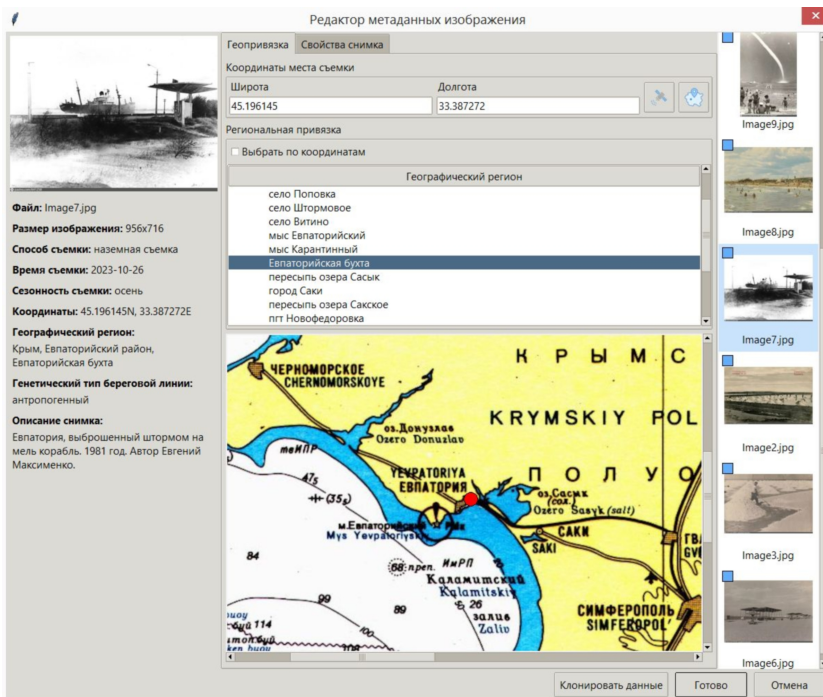
F i g . 3 . The Image metadata editor window

description. The mass upload mode provides the user with the ability to quickly add images with similar metadata to the database, such as the results of an expedition photosurvey at a particular location.

When entering the image metadata, the software system under development implemented the ability to read data from the Exif headers of uploaded files (including the coordinates of the photosurvey location according to GPS data and the date of the photosurvey) and introduced functionality that makes it possible to determine the region and genetic type of the coast by the coordinates of the photosurvey location. The closest region known to the application is chosen as a photosurvey one. Thus, the database of the IRS PhotoCoasts of Crimea contains the information about 140 coastal regions from monograph [7]. To find the region closest to the photosurvey point, the software uses coastline geodata indexing on a uniform grid in the polar coordinate system [10] centred at 45.5° N, 34.0° E.

Work with the geoinformation part of the meta-information base is carried out in the geodata editor (Fig. 4), which can be used to create, delete and correct information about photosurvey regions including coordinates, name and genetic type of the coast.
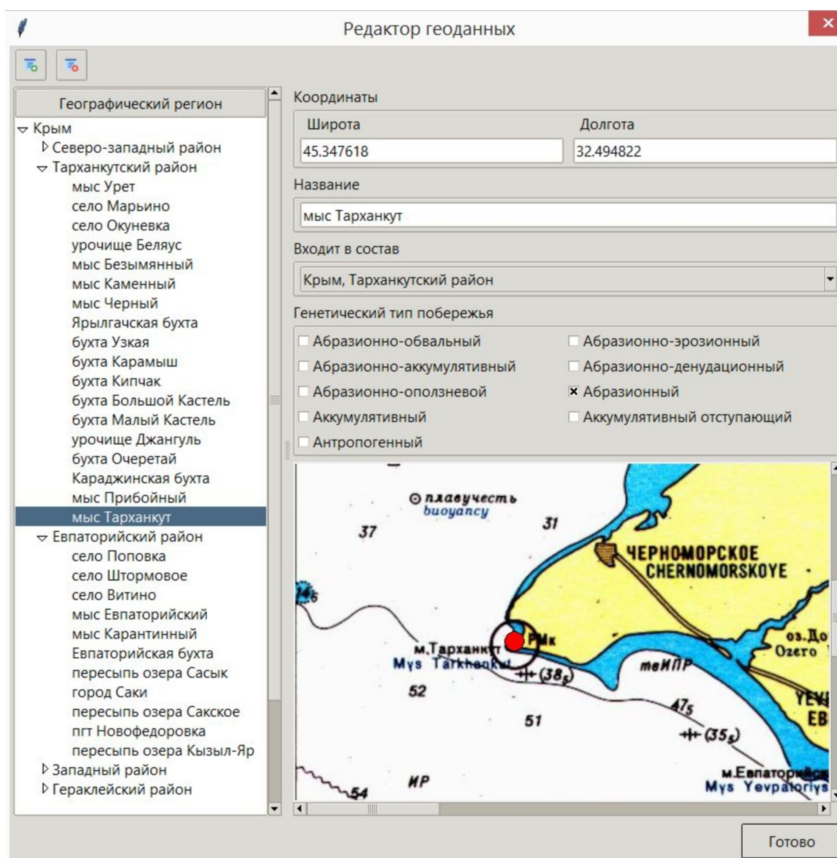
F i g . 4 . The Geodata editor window

Conclusion

The information retrieval system PhotoCoasts of Crimea was developed based on the concept of the software PhotoCoasts. It was optimised for solving the practical task of systematisation, classification and work with the collection of digital images of the Crimean Peninsula coasts in the course of scientific research. The article describes the structure of the image catalogue and the method of storing metadata of the photosurvey subject. The order of work with the system when searching for information and loading new images is described. The key feature of the developed software is the module of formation of the geoinformation base about the Crimean coast. The use of geoinformation together with the ability of reading digital image metadata and mass loading of images significantly facilitates entering information into the system catalogue.

The developed information retrieval system PhotoCoasts of Crimea has a significant potential for further evolution. Its functionality can be extended and adapted

to work with images from other regions. Following appropriate adjustment, this versatile software will allow systematisation, classification and work with digital image collections in a wide variety of scientific fields.

REFERENCES

1. Godin, E.A., Vetsalo, M.P., Galkovskaya, L.K., Goryachkin, Yu.N., Zhuk, E.V., Ingerov, A.V., Isaeva, E.A., Kasyanenko, T.E. and Plastun, T.V., 2022. Information Support of Research in the Black Sea and the Sea of Azov Coastal Zones. In: B. V. Chubarenko, ed., 2022. *All-Russian Conference with International Participation "XXIX Coastal Conference: Field – Based and Theoretical Research in Shore Use Practice".* Kaliningrad: Publishing House of IKBFU, pp. 330–333 (in Russian).

2. Vetsalo, M.P. and Godin, E.A., 2022. Development of a Software System for the Database of Photographic Images of the Crimean Coasts. In: MHI, 2022. *The Seas of Russia: Challenges of the National Science. Proceedings of the All-Russian Scientific Conference. Sevastopol, 26 – 30 September 2022.* Sevastopol: MHI, pp. 287–289 (in Russian).

3. Cherny, A.I., 1975. [*Introduction to Information Retrieval Theory*]. Moscow: Nauka, 238 p. (in Russian).

4. Ranganathan, S.R., 1987. *Colon Classification*. Bangalore: Sarada Ranganathan En-dowment for Library Science.

5. Manning, C.D., Raghavan, P. and Schütze, H., 2008. *Introduction to Information Retrieval.* Cambridge: Cambridge University Press. 2008, 506 p.

6. Witten, I.H., Moffat, A. and Bell, T.C., 1999. *Managing Gigabytes: Compressing and Indexing Documents and Images*. Morgan Kaufmann, 519 p.

7. Goryachkin, Yu.N. and Dolotov, V.V., 2019. *Sea Coasts of Crimea*. Sevastopol: Colorit, 256 p. (in Russian).

8. Bogdanov, D., 2009. Optimal Store and Processing Method for Tree-Structures in Relative Databases. *Software and Systems*, (1), pp. 140–142 (in Russian).

9. Tarassov, S.V. and Burakov, V.V., 2013. Methods of Relational Modeling of Hierarchical Structures of Data. *Information and Control Systems*, (6), pp. 58–66 (in Russian).

10. Bentley, J.L. and Friedman, J.H., 1979. Data Structures for Range Searching. *ACM Computing Surveys*, 11(4), pp. 397–409. https://doi.org/10.1145/356789.356797

About the authors:

Maksim P. Vetsalo, Leading Software Engineer, Marine Hydrophysical Institute of RAS (2 Kapitanskaya St., Sevastopol, 299011, Russian Federation), **ORCID ID: 0000-0002-3543-2124**, **Scopus Author ID: 57222028338**, *mvetsalo@mhi-ras.ru*

Evgeny A. Godin, Research Associate, Marine Hydrophysical Institute of RAS (2 Kapitanskaya St., Sevastopol, 299011, Russian Federation), **ORCID ID: 0000-0002-6469-1379, Scopus Author ID: 56950615200, Researcher ID: AEP-0342-2022,** *godin_ea@mhi-ras.ru*

Elena A. Isaeva, Leading Software Engineer, Marine Hydrophysical Institute of RAS (2 Kapitanskaya St., Sevastopol, 299011, Russian Federation), **ORCID ID: 0000-0002-1860-0026, Scopus Author ID: 57191413519,** *isaeva-ea@mhi-ras.ru*

Lyudmila K. Galkovskaya, Leading Software Engineer, Marine Hydrophysical Institute of RAS (2 Kapitanskaya St., Sevastopol, 299011, Russian Federation), *galkovskaya@gmail.com*