Data exploration of historical and cultural content in an interactive web based system with geolocation

Alexander Petkov

View Online   Export Citation
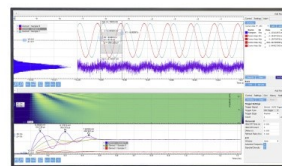
Data Exploration of Historical and Cultural Content in an Interactive Web Based System with Geolocation

Alexander Petkov

Faculty of Applied Mathematics and Informatics, Technical University of Sofia, Bulgaria
Research and Development Sector, Technical University of Sofia, Bulgaria

alex@acstre.com

Abstract. This paper presents an interactive web based system with geolocation for the cultural and historic heritage in Bulgaria. The system uses a three-tier architecture with a presentation layer executed in a browser and the application is accessible over the Internet and allows the users to find cultural and historical content using three search strategies: by category, by keywords and by spatial search over an interactive map. We perform a case study on a real-world dataset from original research and provide some guidelines for future authors of such datasets.

INTRODUCTION

The cultural and historical heritage is an important part of every nation's history. On August 17, 2018, the Council of Ministers of The Republic of Bulgaria approved the "Cultural Heritage, National Memory and Social Development National Research Program" (Council of Ministers of The Republic of Bulgaria, 2018). The program has the following specific goals:
1. Development of digital instruments for research, presentation and popularization of the Bulgarian cultural and historic heritage (CHH)
2. Development of information systems and platforms with geolocation of the CHH for research, preservation and popularization
3. Creation of R&D and educational programs related to the Bulgarian CHH.

The expected results of the research program are as follows:
- New ways to present the scientific results of the humanitarian studies, including the use of "Linked Open Data"
- Products that improve the awareness of the general public regarding the CHH and the national identity (University of Sofia, 2019)

The Technical University of Sofia participates in the research program by developing digital instruments for research, presentation and popularization of the CHH and an information system with geolocation of the CHH. During the first year of the research program, the focus was on the development of the core of the software needed for the project. For the second year, integrating existing datasets created by the other partners and improving the end-user experience becomes the priority.

This paper presents an interactive web-based system with geolocation for data exploration of historical and cultural content. We also delve into a real-world dataset from the research program in a case study where we provide some guidelines for the creation of future datasets.

AN INTERACTIVE WEB-BASED SYSTEM WITH GEOLOCATION

The interactive web-based system with geolocation is a public interactive map of the cultural and historic heritage that provides the general audience with an easy to use, yet powerful tool for spatial, keyword-based and

category-based search. The main purpose of the system is to make the CHH **discoverable and accessible** – living in the information era, we have access to huge amounts of information at our fingertips. Although we sometimes take the process for granted, not all information becomes available as soon as it is created. Actually making the information discoverable requires structure – something that many original sources lack. We approach that problem with three search strategies:

1. We organize our data in categories which helps narrow down the search space
2. We provide keyword-based search functionality which, depending on the keywords, gives very specific results
3. We allow spatial searches on an interactive map – the spatial component of the search is orthogonal to the other search strategies, which also narrows down the results, but most importantly it enables a whole new class of searches, for example – by proximity.

Making the information accessible means using a medium that is suited to the way the general audience consumes information – for example using content management systems that display HTML and inline images instead of Word or PDF documents. We will cover the data formats in greater detail in the case study.

Conceptual Model of the System

The interactive web-based system uses the Acstre GIS platform, developed by the Acstre Software Group, a part of the R&D Sector of the Technical University of Sofia (Acstre, 2020). A conceptual model of the system is shown in Fig. 1.
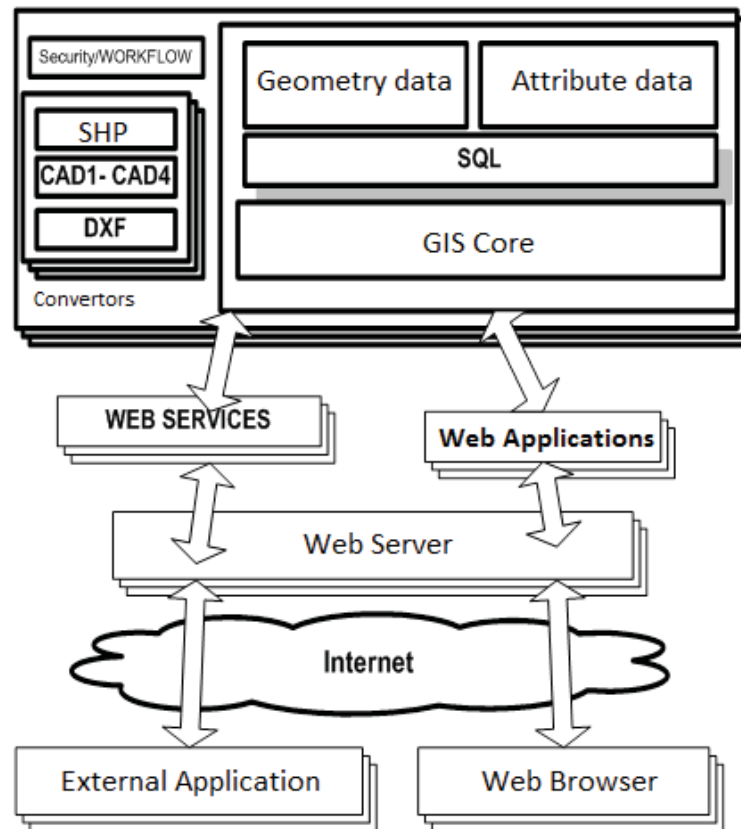

FIGURE 1. Conceptual model of the system.

The client-side module of the interactive web-based system is executed inside a user's web browser. From there, the requests are routed through the Internet to a designated web server. The web server runs a web application that fetches data from Acstre GIS. The main components inside Acstre GIS are the GIS Core, an SQL database that stores the geometry and attribute data, and various other modules, most notably the import module that can read popular GIS formats and convert the data into Acstre GIS's internal representation.

Features of the Interactive Web-based System with Geolocation

The web-based system provides an interactive map using the WMS standard. The web application consists of two parts:

- A server side application which conforms to the WMS standard
- A client side application which uses the information fetched from the server to create an interactive map

The server-side application uses the interfaces provided by the GIS core. It can publish different layers of the digital model on the web. To do that it fetches the data from the GIS database and applies the necessary transformations to create raster images of the digital model. The application can query the attributes of the objects near a point on the map or by using an attribute-based search.

The client-side application renders the raster images in the browser. The application can display the attributes of the objects in the raster images on the map. The interactive map has tools for measuring distance and area and can search for objects on the map using the name of the CHH, cadastre identifier and other attributes. The client application can display the map on a base layer, containing satellite images, a road map or a terrain map to provide some extra details. The base layer is positioned under the raster images coming from the server.

The interactive map can also query the location of the client device and display the map centered on the user's location. This is especially useful when the application is used by tourists to find the location of cultural and historic heritage around them, helping them find what they're looking for or helping them find new object they didn't know existed.

Data Sources

For the system to be useful, it has to provide access to some meaningful, real-world data. There are two main sources of data for the project: public data and original research from partners. During the first year of the research program, we focused on public data, as original research from the partners wasn't yet available.

Some sources of public data for the CHH are UNESCO, The Ministry of Culture of Bulgaria and The National Institute for the Immovable Cultural Heritage. We selected the Ministry of Culture's registry of all museums as the first dataset to import, as it provides a large list of well-dispersed points of interest for the entire country. A view of the interactive map showing all museums in Bulgaria is shown in Fig. 2.
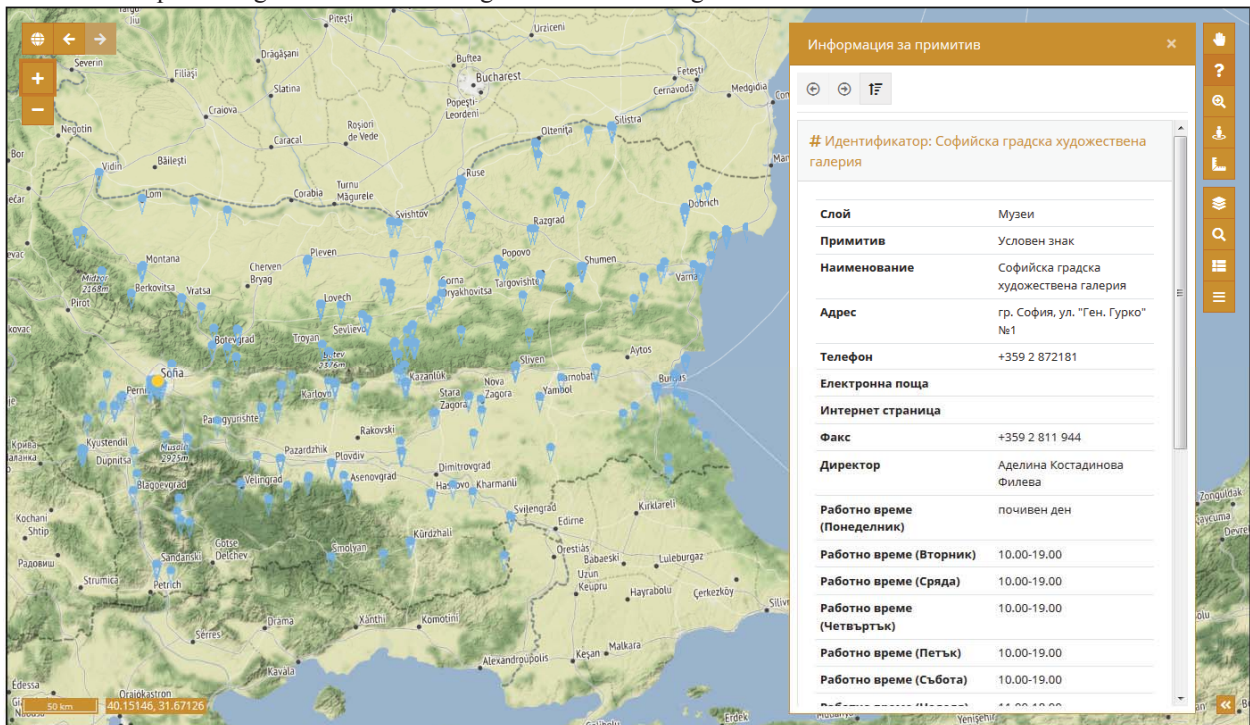

FIGURE 2. A view of the interactive map that shows all museums in Bulgaria.

The data was available as an MS Excel spreadsheet. Coordinates for geolocation weren't available, so we used the administrative addresses, Google's Geocoding API and a process of manual verification for the import.

The first piece of original research that found its place in the system were the "Megalithic Monuments in Southwestern Bulgaria" – a research project carried out by the South-West University "Neofit Rilski" – Blagoevgrad. A case study of the provided data and the import process follows.

CASE STUDY: MEGALITHIC MONUMENTS IN SOUTHWESTERN BULGARIA

The Data

The "Megalithic Monuments in Southwestern Bulgaria" was the first dataset we received from the partners in the research program and needless to say we were eager to see what it looked like. What we saw was a simple structure that would work great for people – and that was to be expected from authors in the humanitarian studies. The dataset consisted of the following elements:

- A single, 78 MB Microsoft Word document describing the findings
- A folder with 108 photos of the monuments
- A GIF slideshow with all 108 photos as separate frames with reduced resolution

A description of each element follows.

The .docx File

The one thing that stands out the most is that **everything** is in that file. Our task is to break it down into structured components, so we examined the structure of the document. It was, in fact, reasonably well structured – for each monument, we have this recurring pattern:

- A roman numeral, followed by the name of the monument
- GPS coordinates on a separate line, some with altitude
- Several paragraphs of text description
- A section for photos, starting with "Снимки:" ("Photos:")
- For each photo, a number, a link to the photo in the folder and a short description

The Photos

The photos followed a sound naming convention: the Roman numeral from the .docx file, underscore and an Arabic numeral for the consecutive number of the photo for that monument.

Another thing to note is that the photos had an excellent resolution for web purposes. All photos were made with 20MP digital cameras.

The Slideshow

The slideshow GIF had all the images from the "photos" folder but with significantly reduced quality. Since it did not add any new data, it was not used in the import process.

Import Procedure

We used a simple two-step approach to the import procedure:
1. Identify the features in the dataset and their attributes
2. Select an import format, if one is available, or perform data extraction to a usable format

Feature Identification

The main feature in the dataset was clearly a "megalithic monument". We identified the following attributes:
- Monument number
- Monument name

- Latitude and longitude
- Description
- Altitude

For each monument, we also had a list of photos with their descriptions – a one-to-many relationship between the monuments and the photos.

Import fFormat and Data Extraction

No directly importable format was present. However, a Microsoft Excel spreadsheet can provide sufficient structure. The main task is how to map the text from the MS Word document to the columns of the spreadsheet, which represent the attributes of the feature. We considered two approaches: parsing the .docx file programmatically or populating the Excel file manually. We chose the manual approach for the following reasons:

1. The total number of features was relatively small (there were 26 monuments in the file)
2. The GPS coordinates were in decimal degrees for some monuments and in degrees, minutes and seconds for others
3. Only some monuments had an altitude field
4. Some monument descriptions were intertwined with the photos

The irregularities in the file (points 2, 3 and 4) would make programmatic parsing more difficult, whereas the relatively small number of monuments (point 1) makes the manual approach feasible.
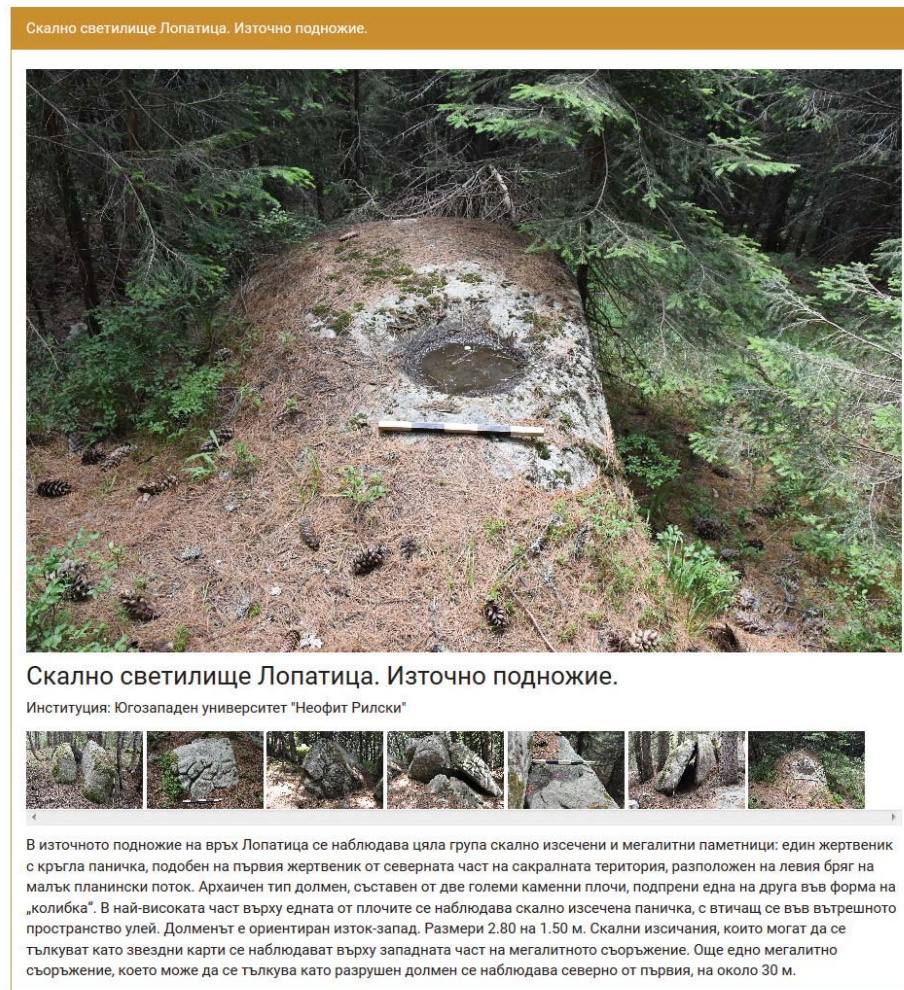

FIGURE 3. A structured view of a single megalithic monument.

Case Study: The Conclusion

Overall, we can consider this dataset a good start. The lack of structure and consistency could be a major concern for larger datasets. Should the monuments have been much larger in numbers, the manual approach would have become infeasible and a complex program that handles all the inconsistencies in the unstructured data would have been needed. To prevent this in the future we had two meetings with the partners where we presented our guidelines for structuring the data using Excel files. Another point to make is that the dataset had no categorization. It is clear that a unified list of categories for the data would be beneficial. Also, additional metadata, such as keywords or tags, could make the data easier to search for. A structured view of a single megalithic monument after the import is shown in Fig. 3.

RESULTS

In this paper, we presented an interactive web-based system with geolocation that provides spatial, keyword and category based search. Real-world data was successfully imported in the system and guidelines on structuring the data were developed, which should provide more structure to future datasets.

ACKNOWLEDGMENTS

REFERENCES

1. Acstre. (2020). *За Акстър ГИС*. Retrieved from Софтуерна група Акстър: https://acstre.com/subsection-61-za_akstyr_gis.html
2. Council of Ministers of The Republic of Bulgaria. (2018, 08 17). DCM No 577. Sofia, Bulgaria.
3. University of Sofia. (2019). Национална научна програма „Културноисторическо наследство, национална памет и обществено развитие". Retrieved from КИННПОР: https://kinnpor.uni-sofia.bg/programme