GIS Odyssey Journal

2023• Volume 3 •Number 1

https://doi.org/10.57599/gisoj.2023.3.1.73

Marzenna Miłek[1], Jerzy Stanik[2], Maciej Kiedrowicz[3], Jarosław Napiórkowski[4]

COMPARATIVE GIS ANALYSIS USING TAXONOMY AND CLASSIFICATION TECHNIQUES

Abstract: The article discusses the methodology of comparative analysis of GIS class computer systems using the current elements of taxonomy and classification theory. Eighteen selected GIS class systems that meet the criterion of completeness of all data required in the conducted research were fully analyzed. The proper comparative characteristics were preceded by the recognition of the market situation in terms of the availability of GIS systems. Eight thematic groups of criteria were used in the research, based on which the selection of GIS solutions for comparison was carried out. The adopted system selection criteria carry out the selection of objects in a binary manner. The chosen rules of classifying the system into a set of objects for comparison covered the issue of the availability of the required information. Due to the characteristics carried out, this information was obligatory, because a full set of data is required, which will allow for a comprehensive and factual comparison, based on which a given product (GIS system) can be indicated to the consumer with full responsibility. The set of features and comparative criteria was created based on own experience and numerous consultations with specialists and field experts. The selected criteria are the most widely used and most accepted in the environments that systems of this class use daily. Both the functional scope (features, functions, properties, advantages and disadvantages) as well as the degree of fulfillment of subsequent criteria by the considered systems were determined and described.

Keywords: GIS, classification, estimators, hierarchical methods

Received: 16 May 2023; accepted: 18 June 2023

[1] Military University of Technology, Faculty of Cybernetics, Institute of Computer and Information Systems, Warsaw, Poland, ORCID ID: 0009-0002-9806-6755, email: marzenna.milek@wat.edu.pl
[2] Military University of Technology, Faculty of Cybernetics, Institute of Computer and Information Systems, Warsaw, Poland, ORCID ID: 0000-0002-0162-2579, email: jerzy.stanik@wat.edu.pl
[3] Military University of Technology, Faculty of Cybernetics, Institute of Computer and Information Systems, Warsaw, Poland, ORCID ID: 0000-0002-4389-0774, email: maciej.kiedrowicz@wat.edu.pl
[4] Military University of Technology, Faculty of Cybernetics, Institute of Computer and Information Systems, Warsaw, Poland, ORCID ID: 0000-0003-3286-3751, email: jaroslaw.napiorkowski@wat.edu.pl

Introduction

The development of geoinformation technologies and methods of geodata analysis in the early dz2000s ąbrought classification and taxonomy to a qualitatively new level. Many sciences developed independently of each other. Therefore, in publications on classification issues, one can find special cases of methods that were already known and systematized in the statistics used at that time. The article discusses the current issues of the use of classification and taxonomy in relation to geoinformation systems in the field of comparative analyses. The problems of GIS classification for the purpose of conducting comparative characteristics have been studied in the works of many authors (Miłek et al., 2023a; Miłek et al., 2023b; Xuan, 2015). There are also materials on the subject on the Internet, as well as in foreign sources (Zhang et al., 2917; Lubis et al., 2017; Neema et al., 2020; Risky, 2018).

In this article, the overarching goal of GIS classification is to determine the similarity of individual instances and their clusters (i.e. classes). Generally, the classification can be defined as a division of the set of considered objects – here GIS systems into classes, containing such objects that will be similar in terms of the observed values of their features. It is very important here that the objects allocated in different classes have the greatest possible differences. They should show the greatest correlation within the cluster – class. Hence, a well-conducted classification guarantees the fulfillment of the two postulates: internal coherence and external isolation.

The purpose of this article is to present the methodology of selection, ranking and classification of GIS enabling the selection of a properly tailored and effective solution for a given enterprise. The problem is not trivial, but difficult. Therefore, this paper presents its own way of thinking and several techniques that can be used by a decision maker in making everyday decisions. It is important and repeatedly emphasized that all these techniques are universal and can be applied to almost any problem that can be encountered today.

The results obtained both by means of multi-criteria optimization in the sense of PARETO or classification can and very often constitute a kind of help, guidance and advice in decision-making, however, one should not base one's choice solely on them. As indicated at the research stage, these methods are characterized by high sensitivity (sensitivity) to data, and their proper implementation is associated with the meticulousness, knowledge and precision of the researcher.

Assumptions and literature review

The GIS software industry includes a wide range of commercial and open source products that provide some or all of these capabilities under various information technology architectures (Fu & Sun, 2010).

With the transition to networking and cloud computing, and integration with real-time information via the Internet of Things, GIS has become a vital platform for almost every human endeavor – enterprise GIS. Table 1 lists notable types – Enterprise GIS.

Table 1. Types of GIS – Enterprise GIS

GIS type	Characteristic (reference to literature)
Enterprise GIS	Enterprise GIS refers to a geographic information system that integrates geographic data across multiple departments and supports the entire organization (Chang & Kang-tsung, 2016). The basic idea of enterprise GIS is to deal with the needs of departments collectively, not individually. When organizations began using GIS in the 1960s and 1970s, the focus shifted to individual projects where individual users created and maintained datasets on their own desktop computers. Due to the extensive interaction and workflow between departments, many organizations in recent years have moved from independent, standalone GIS systems to more integrated approaches that share resources and applications (Goodchild, 2010).
Corporate GIS	A corporate geographic information system is similar to enterprise GIS and meets the needs of the organization as a whole in terms of spatial information in an integrated manner. Enterprise GIS consists of four technological components, which are data, standards, information technology, and expertise personnel. It is a coordinated approach that moves away from fragmented desktop GIS. The corporate GIS project involves the construction of a centralized corporate database, which is to be the main resource for the entire organization. The enterprise database is specifically designed to efficiently and effectively meet the requirements of the organization. Essential to enterprise GIS is the effective management of the enterprise database and the establishment of standards such as the OGC for mapping and database technologies. The benefits include that all users in the organization have access to shared, complete, accurate, high-quality, and up-to-date data. All users in the organization also have access to common technology and people with knowledge. This improves the efficiency and effectiveness of the organization as a whole. An effectively managed corporate database reduces unnecessary collection and storage of information throughout the organization. By centralizing resources and efforts, it reduces overall costs (Somers, 1996; Vastag et al., 1994; Weil & Broadbent, 1994).
Internet GIS	Internet GIS, or online geographic information systems, is a term that refers to a broad set of technologies and applications that use the Internet to access, analyse, visualize and distribute spatial data (Broome & Meixler, 1990; Fitzgerald, 2007; MacHarg, 1971; Tobler, 2009). Web GIS (also known as Web-Based GIS) or Web Geographic Information Systems is a GIS that uses the World Wide Web to facilitate the storage, visualization, analysis and distribution of spatial information on the Internet (www.esri.com; www.wiki.osgeo.org; Xuan, 2015).
Mobile GIS	With ~80% of all data considered to have a spatial component, modern Mobile GIS is a powerful geocentric business process integration platform enabling Spatial Enterprise (Fu & Sun, 2010). The number of mobile devices in circulation surpassed the world's population (2013) thanks to the rapid acceleration of the use of iOS, Android and Windows 8 tablets. Tablets are rapidly becoming popular in field applications. Affordable MIL-STD-810 certified cases transform consumer tablets into fully ruggedized yet lightweight devices for field use that cost 10% of older ruggedized laptops. While not all mobile GIS applications are device-limited, many are. These restrictions apply more to smaller devices such as mobile phones and PDAs. Such devices have: small screens with poor resolution, limited memory and computing power, weak (or no) keyboard and short battery life. Additional limitations can be found in web client-based tablet applications: poor graphical web interface and device integration, on-line reliance, and very limited off-line network client cache (www.wiki.osgeo.org; Weill & Broadbent, 1994).

Source: own study

Some of the potential benefits that an enterprise GIS can provide include greatly reduced system-wide data redundancy, greater accuracy and integrity of geographic information, and more efficient use and sharing of data (Maliene et al., 2011). Since data is one of the most important investments in any GIS program, any approach that reduces acquisition costs while maintaining data quality is important. Deploying enterprise GIS can also reduce overall GIS maintenance and support costs, ensuring more efficient use of departmental GIS resources. Data can be integrated and used in decision-making processes throughout the organization (Maliene et al., 2011).

GIS today gives people the ability to create their own digital map layers to help solve real-world problems. GIS has also evolved into a means of data exchange and collaboration, inspiring a vision that is now rapidly becoming a reality – a continuous, overlapping and interoperable GIS database around the world covering virtually all topics. Today, hundreds of thousands of organizations share their work and create billions of maps every day to tell stories and reveal patterns, trends, and relationships about everything. GIS is all about discovering meaning and insights from within data. It is rapidly evolving and provides a whole new framework and understanding process.

As our world faces issues of increasing population, loss of nature and pollution, GIS will play an increasingly important role in how we understand and solve these problems and provide the means to communicate solutions using a common mapping language.

Review of the studied GIS systems

There are plenty of GIS systems on the market, designed for a wide range of applications. Many of the offered solutions come from various domestic and foreign companies. To illustrate the number of foreign players and market tycoons, a summary in the form of the Gartner magic quadrant was presented (Fig. 1).

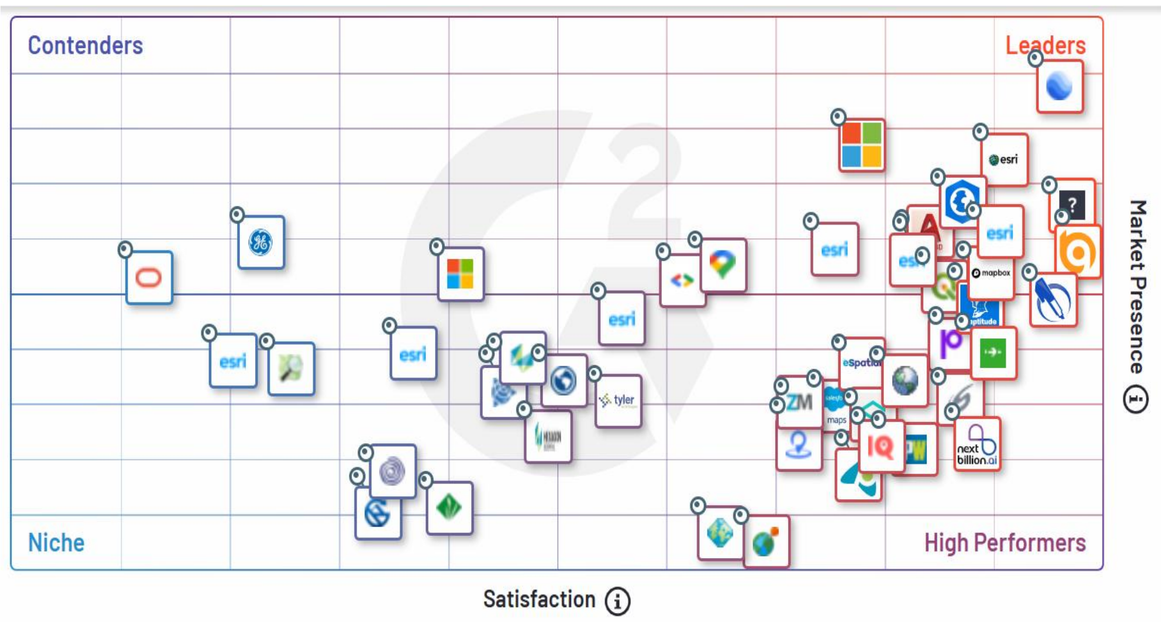
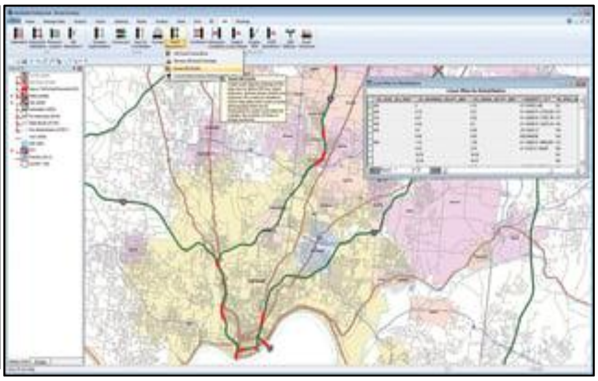


Fig. 1. G2 Grid® for GIS
Source: own study

Bearing in mind the limitations of the article, only six systems will be characterized in this subchapter: Hexagon GoeMedia, MapInfo Pro, Surfer, GIS Cloud Track, ArcGis Enterprise and GE Smallword. Other GIS can be found in studies (Miłek et al., 2023a; Miłek et al., 2023b; Mironova, 2018; Mironova, 2020).

Hexagon GeoMedia – GIS software with a powerful set of remote sensing tools. It is a powerful, flexible GIS management platform that allows you to aggregate data from various sources and analyze them in unison to extract clear, actionable information. It provides simultaneous access to geospatial data in almost any form and displays it in one unified map view for efficient processing, analysis, presentation and sharing. It is a comprehensive and dynamic GIS that extracts intelligence from geospatial data. This app can access geospatial data directly; There is no need to import, convert or use proprietary technologies to connect to spatial databases. The most popular versions of Hexagon GeoMedia are Desktop 16.0 and 15.0. The actual developer of the software is Hexagon Geospatial. Official screenshots:

Basic functions:
GeoMedia's functionality makes it ideal for extracting information from a range of dynamically changing data to support informed, smarter decision making. functional scope:
Data management:
- Data capture.
- Data storage.
- Data manipulation.
- Data visualization.

Creating a map:
- Buffer zone query.
- Overlaying.

Analysis:
- Spatial analysis.
- Reporting.

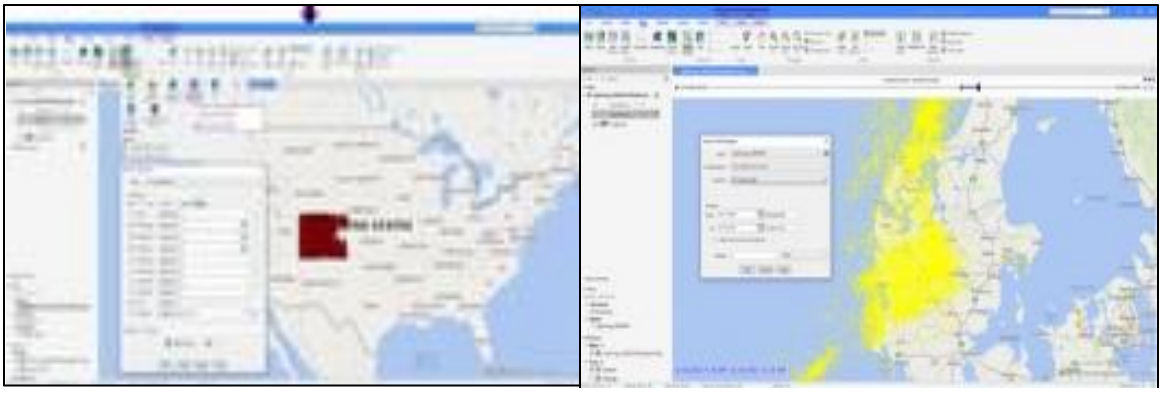
Pros:

- Fast query and analysis.
- Robust cartography with smart labelling.
- Remote sensing with ERDAS Imagine.
- Universal mapping with multiple layouts.
- Perfect editing with smart snapping.
- Mature software with 40+ years of history.

Minuses:

- Confusing license levels.
- A small community of users to troubleshoot.
- You cannot drag and drop files into GeoMedia.
- Poor interoperability with other GIS formats.
- Database connectivity can be slow.

MapInfo Professional. GIS software with a focus on business and location intelligence. MapInfo Professional is a well-rounded GIS software suite with a greater focus on business decision making. It is based on location intelligence like GeoMedia . MapInfo Pro works with a variety of spatial data formats and many types of databases. Data imported to the program are displayed in the form of layers consisting of a map and a table of attributes. This enables the integration of data from e.g. Microsoft Excel, Access or database servers (Oracle, PostgreSQL , etc.) with the map, and then their visualization. Official screenshots:



Basic functions:

MapInfo Pro is a GIS class program that allows you to create, manage and update spatial data. With its help, you can visualize and analyze data on maps, as well as successfully build advanced spatial information systems. functional scope:

Data management:

- Data capture.
- Data storage.
- Data manipulation.
- Data visualization.

Creating a map:

- Geocoding.
- Data vectorization.
- Generating spatial queries.
- Editing objects on the map and in tabular data.
- Visualization of data on a map from external data sets.
- Overlaying.
- Publishing.

Analysis:

- Spatial analysis.
- Reporting.
- Real-time streaming.
- Spatial analysis of customer locations.
- Analysis of the potential of the regional market.
- Analysis of the scope of competition.
- Distribution network analysis.

Cartography:

- Map design.
- Data visualization.
- Overlaying.

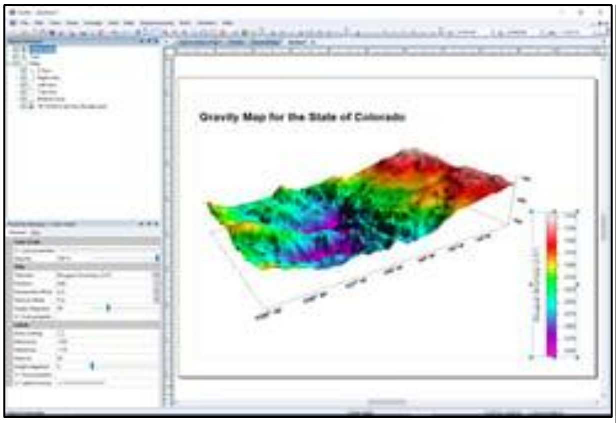
Pros:

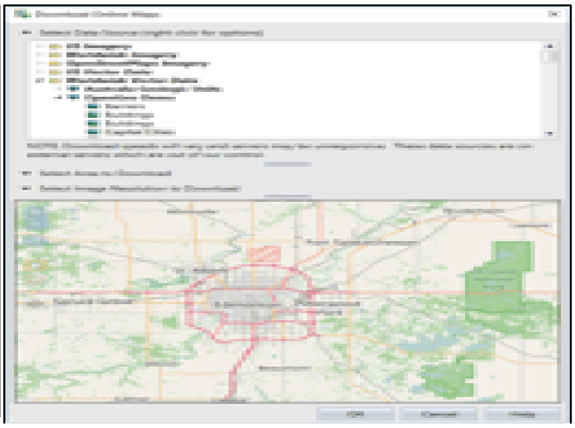
- Ease of use and 64-bit processing.
- Query execution and improved table management.
- Advanced addressing and geocoding.
- Side-by-side mapping.
- Improved visualization integration.
- Smart ribbon-based navigation.

Minuses:

- Data processing and analysis will take a long time.
- Interoperability and poor format support.
- High license cost.
- No cloud-based platform.
- Low functionality of online web maps.
- Weak support for remote sensing analysis.

SURF. Product by Golden Software, Inc. Surfer ® is a full-featured 2D and 3D mapping, modeling and analysis software package for scientists and engineers. Surfer's advanced interpolation engine quickly converts XYZ and XYZC data into publication-quality maps. Virtually every aspect of the map can be customized. Enhance your maps with profiles, legends, titles and labels, faults and breaklines, or external maps from any online map service. The Surfer is widely used by geologists, geophysicists, hydrologists, archaeologists, oceanographers, biologists, consultants, engineers and many others around the world. Thanks to its basic features, Surfer has been used primarily for creating maps. It not only allows you to create contour maps and their spatial images, but also allows you to make calculations based on a regular grid of values. Specialized procedures allow you to generate map sections along any selected polyline. The program allows you to calculate the area of curvatures, the area of projections on the XY plane and volumes. Official screenshots:





Basic functions:

Surfer ® 25 is a program designed for comprehensive visualization of XYZ data, thanks to which it is most often used for mapping and modeling terrain surfaces, but this is not the only application of this program. The built-in wide set of interpolation methods for generating a regular grid of values allows you to choose the optimal algorithm for the nature of the input data.

Functional scope:

Data management:

- Data capture.
- Data storage.
- Data manipulation.
- Data visualization.

Create a map:

- Geocoding.
- Buffer zone query.
- Overlaying.
- Publishing.

Analysis:

- Spatial analysis.
- Reporting.
- Real-time streaming.

Advantages:

- Is the easiest to use contouring, meshing and surface mapping software.
- It is very accurate and helps you to design the best rough map with grids.
- Has a wide selection of mapping methods and tools, including all known and tested grid algorithms.

- It has more map types including vector, wireframe and base unlike other apps.
- Does not require high-capacity machines.

Defects:

- It is difficult to manage.
- No direct integration with ESRI.
- Map resolutions are not at the best level.
- Sometimes the interpolation was not very accurate.
- No in-depth analytics like other comparable platforms.

GIS Cloud Track – Track your fleet, people and assets. GIS Cloud Track – is a new generation platform for applications that manage location information. GIS Cloud Track allows you to dynamically visualize your assets in real time, track their location, and get detailed reports, charts, and historical data based on multiple parameters. Official screenshots:

Basic functions:
GIS Clouds Track supports multiple vector and raster formats, rich GIS symbology, and has built-in collaboration features for real-time editing and sharing. functional scope:
Data management"

- Data storage.
- Data manipulation.
- Data visualization.

Analysis:

- Spatial analysis.
- Reporting.
- Real-time streaming.

Pros:

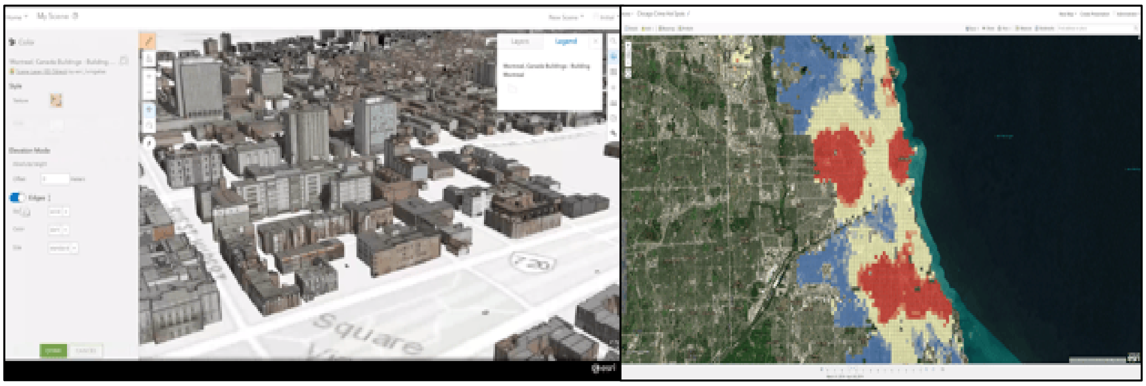
- Extensible with ArcGIS , QGIS and WhiteBox tools.
- Reads many formats.
- Publishing web maps via MangoMap.

Cons:

- Poor symbols and print layouts.
- Editing tools are not reliable.
- No new technologies.

- Limited cartography, labeling and symbolization.
- There is no data management directory.
- Metadata cannot be saved or edited.

ArcGIS Enterprise. ArcGIS Enterprise is a full-featured mapping and analysis platform that includes a powerful GIS server and a dedicated web-based GIS infrastructure to organize and share your work. ArcGIS Server Enterprise consists of four major software components, such as ArcGIS Server, Portal for ArcGIS , ArcGIS Data Store , and ArcGIS Web Adapter . These components facilitate the creation of web-based spatial data management and sharing systems that can be run on both local and cloud platforms. The online system consists of options for publishing, managing, organizing, customizing and downloading spatial data. This allows users to work in an integrated environment and helps create efficient geospatial workflows . In addition, it can support advanced image processing, real-time analytics, and big data integration with additional license purchases. Official screenshots:



Basic functions:
ArcGIS Enterprise provides a secure, browser-based mapping and analytics platform that leverages Esri 's advanced technology to extend big data, imagery, and more in real time. functional scope:
Data management:

- Data capture.
- Data storage.
- Data manipulation.
- Data visualization.

Create a map:

- Geocoding.
- Buffer zone query.
- Overlaying.
- Publishing.

Analysis:

- Spatial analysis.

- Reporting.
- Streaming in time of things.

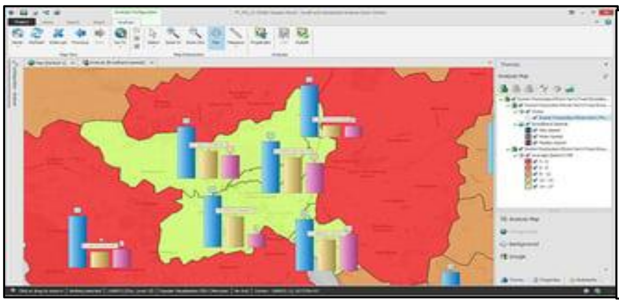
Advantages:

- Very easy to use and to create a quick web map.
- Best of all, geography and track logging.
- Can create apps without coding; easily share data within your organization and publicly.
- ArcGIS is very user – friendly; This mapping app allows for quick and organized data collection. You can run the program on a computer on a portable device such as an iPhone.
- The centrality of the tool is extremely helpful and allows you to collaborate and build solutions to all your geospatial problems as well as assist with hosting solutions.

Defects:

- You cannot migrate directly from bigdata or teradata.
- There are some bugs in the software that cause delays and problems randomly and there is no pattern to solve them, and they require extreme patience and can be annoying at times.
- Receiving data online is not good; data download is quite slow.

GE Smallworld. GE's market-proven comprehensive suite of integrated tools enables customers to reduce cost of ownership, simplify critical network infrastructure, and provide the flexibility required to meet dynamic network requirements. This revolutionary, object-oriented, database-driven product provides an efficient, consistent architecture at the heart of many applications, such as those used to plan electricity, gas and water distribution systems, design telecommunications networks and evaluate strategic market opportunities. The software integrates with other products that require spatial information, including systems for customer relationship management, market analysis, network and work management. The architecture is built on open technology standards and supports scaling to the largest customer needs. Deployment options using industry-leading technology platform solutions. Official screenshots:



Basic functions:
GE's innovative networking software is mainly based on Smallworld ™ Core technology Spatial Technology. Smallworld Core provides a comprehensive portfolio of solutions that support critical processes in planning, designing, building, operating and maintaining the lifecycle of network-intensive industries. functional scope:
Data management:
- Data storage.
- Data manipulation.
- Data visualization.

Creating a map:
- Geocoding.
- Overlaying.
- Spatial analysis – GIS.
- Reporting – GIS.
- Data visualization – GIS.

Advantages:
- The ability to capture detailed network inventory and connectivity between network elements.
- Stability and speed compared to other platforms.
- Smallworld was very detailed and allowed the company to tailor the product to what it felt was important.
- It has many positives for the GIS Utilities market.

Defects:
- The mapping capabilities and some spatial queries from a user perspective are quite limited.
- A bit more complex than other platforms; harder to learn.
- Bulk data loading is the biggest problem and makes it not very useful for the initial customer.

Methodology

The following concept of conduct was adopted in the study. Comparison criteria and rules for the selection of GIS systems for comparison were defined. Then, factors were specified, which, based on the research and experience, are considered critical in the process of selection and implementation of appropriate GIS class solutions. The most important selection criteria – selection of systems were indicated and on their basis a list was prepared and a comparative analysis of GIS systems was carried out. The comprehensive characterization was carried out based on a number of experiences, interviews and information obtained from many different sources, ranging from periodicals and industry literature with a high eigenfactor, to websites and consultations with system manufacturers. However, environmental conditions made companies reluctant to share any information about their products. Most of the data was very general.

Due to the large number of GIS systems on the market, it is impossible to obtain information about all products. The main difficulties were related to the acquisition of the necessary materials. Therefore, for the purpose of comparison, the following rules have been defined, according to which the solutions operating on the market will be classified into the set of compared systems:

- the system selection criteria were tailored to the needs of the small and medium-sized enterprise (SME) sector,
- the selection of systems for comparison was based on the binary technique, in which the presence of any information in the aspect of eight thematic groups of criteria was detected,
- the size of the considered set of studies consists of GIS systems selected from all available on the world market today, source " systems, but to be a true GIS, the system must contain a significant group of components (Fig. 2.)

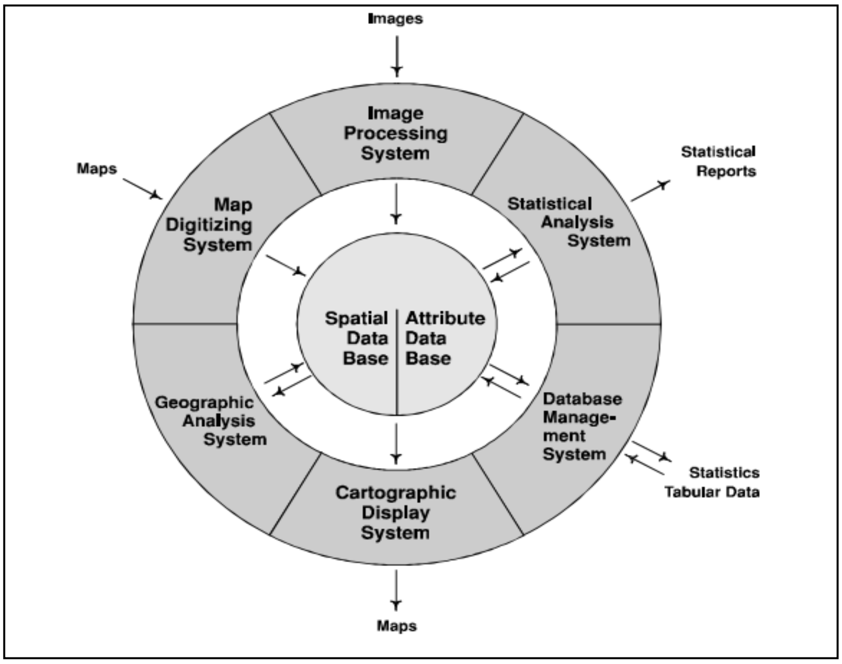


Fig. 2. Basic components of the GIS system
Source: own study

It is also mentioned that the list was based on a number of own experiences, as well as the experience of people who are closely related to the GIS class under consideration.

Since the implementation of the work objective requires a mutual comparison of the examined systems, in the second step, the criteria for comparing selected GIS solutions were selected. Previously, a selection of the market was carried out, during which a classification (depending on the adopted rules and selection criteria) of subsequent systems dedicated to small and medium-sized enterprises (SMEs) was carried out. When comparing the systems, decomposed component criteria (sub-criteria, sub-criteria) of eight, thematically aggregated groups of main features were used, on the basis of which the systems for the study were classified. Hence, the current criteria are a subset of the features that were used to select GIS systems for comparison.

The adopted comparative criteria have been presented in a hierarchical form, grouped into appropriate thematic classes using a mindmap scheme (Fig. 3):

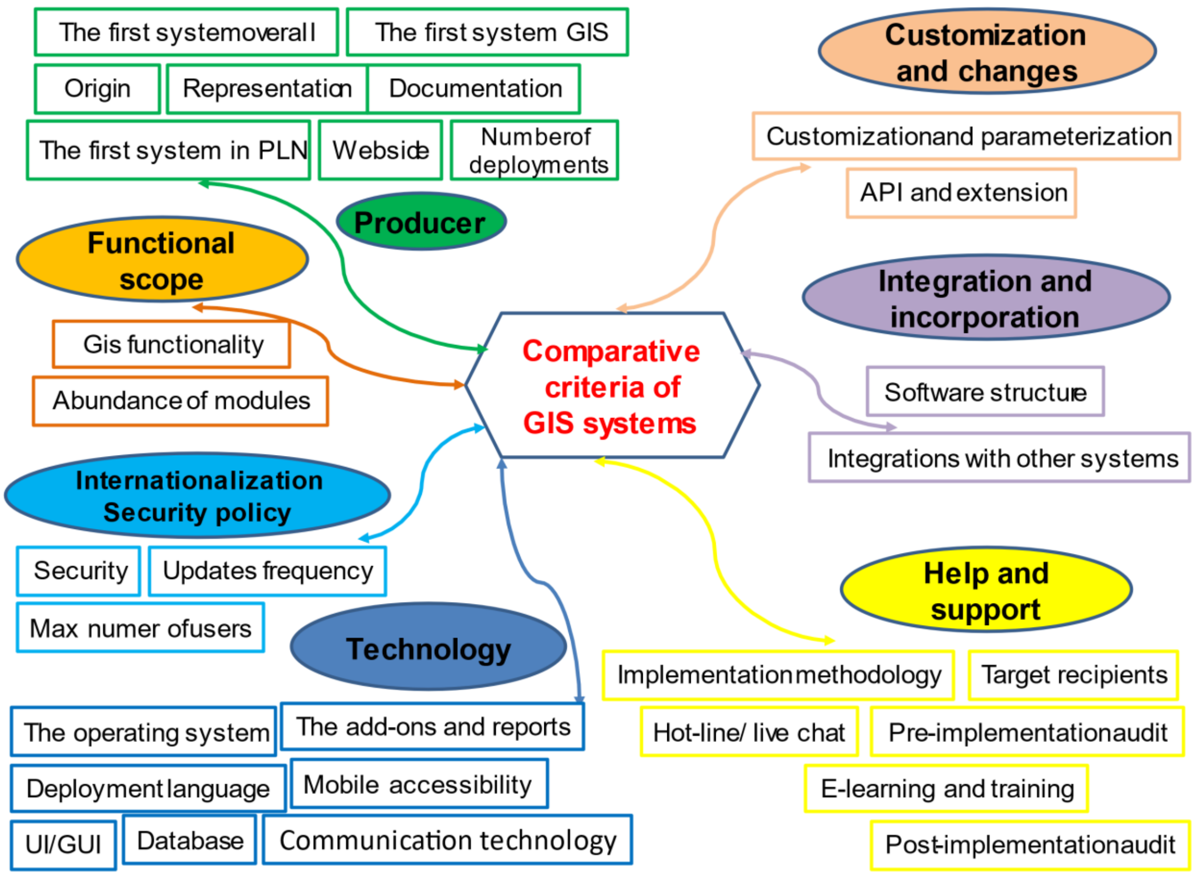


Fig. 3. Groups of comparative criteria of the considered GIS class systems
Source: own study

Eight thematic groups of criteria were used in the research, based on which the selection of GIS solutions for comparison was carried out. The adopted criteria for the selection of systems carry out the selection of GIS in binary way . If a given criterion is not met (required information not obtained), then the GIS is excluded from further consideration and analysis . The chosen rules for the classification of the GIS system for the set of objects to be compared covered the issue of the availability of the required information. Due to the characteristics carried out, this information was obligatory, because a full set of data is required, which will allow for a comprehensive and factual comparison, on the basis of which z a given product (GIS system) can be indicated to the consumer with full responsibility. In further research, thematic groups of features were decomposed into sub-criteria (component sub-criteria), incl based on which a proper comparison of the systems under consideration was carried out .

When considering the problem of selecting the appropriate GIS class system, classification methods were used. For the normalized form of data, classification was carried out using agglomeration combinatorial methods, for which distances were estimated using the Euclidean metric (i.e. for p=2). The performed classification grouped and allocated the considered objects to particular classes. The direct result of the classification of objects is the resulting matrix, while its visual representation is presented later, using the created dendrogram of objects. It presents (dendrogram) in a very legible way individual categories to which each of the eighteen considered systems was allocated.

The presented dendrogram is a graphical result of the classification of the eighteen considered systems. In order to be able to more accurately interpret its structure, the idea of classifying using hierarchical and non-hierarchical methods was outlined.

When making a proper analysis of the obtained results, the presented dendrogram should be carefully analyzed. It presents in a graphic way the division of the set of examined GIS systems, obtained in the process of its classification. When grouping objects, the mathematical apparatus turned out to be very helpful, which showed the degree of correctness and quality of the model through a set of appropriate indicators (indicators).

Result and discussion

In relation to GIS-class systems, classification and taxonomy are methods of organizing and categorizing geoinformation solutions in a form that people are able to understand. These are tools that allow us to maintain GIS computer systems so that they can be easily compared and contrasted. For the purposes of this article, the classification is defined as a set of clusters (in other words, clusters or classes) of the form:

$$K = \{K_1, K_2, \dots, K_n\},$$

which are appropriately highlighted in the set O defined as the following set O of objects:

$$O = \{O_1, O_2, \dots, O_n\}.$$

Additionally, objects included in the O set must have the following triduum of features:

$$\vee_{t=1,2,\dots u}[(K_t \in O) \wedge (K_t \notin \emptyset)] \vee_{t\neq v; t,v=1,2,\dots,u}(K_t \cap K_v = \emptyset); \ \cup_{t=1}^{u} K_t = O.$$

The primary purpose of classification is to determine the similarity of individual instances and their clusters (i.e. classes). Generally, the classification can be defined as a division of the set of considered objects - here GIS systems into classes, containing such objects that will be similar in terms of the observed values of their features. It is very important here that the objects allocated in different classes have the greatest possible differences. They should show the greatest correlation within the cluster - class. Hence, a well-conducted classification guarantees the fulfillment of the two postulates: internal coherence and external isolation.

The most important stages of classification and typical actions taken in this type of research are listed below:

1. selection of objects to be classified and classification features:
   a) all objects of the population or its selected samples may be qualified. Very often, the appropriate sample size of the entire population is determined mathematically;

   b) the choice of object features is very important – the quality of classification and the accuracy of the decisions made depend mainly on them;
   c) it is extremely important to include in the set only those objects and features that bring the greatest added value to the purpose of classification;
   d) discriminant assessment of features is made possible by, among others, asymmetry (skewness) and kurtosis (flattening, Donoghue's measure) metrics;
   e) aggregates, i.e. artificial features that are substitutes for subsets of the original features, are often introduced; these are usually linear combinations of the original features;
2. selection of the formula for normalizing the values of object features;
3. choice of distance measure between objects:
   a) the classification is based on the distance matrix;
   b) the distance matrix is always created on a standardized, i.e. normalized, data matrix;
   c) most often, the distances separating individual objects are determined by Euclidean and Mahalanobis metrics;
4. choice of classification method:
   a) single-linkage, as the distance of the nearest neighbor in terms of a single link. The distance between two classes is the shortest distance among all distances between objects belonging to the considered classes;
   b) complete linkage, as the furthest neighbor metric in terms of complete linkage. It represents the distance between two classes, which is the greatest distance among all distances between objects belonging to the considered classes;
   c) median as the distance between two classes, which is the median of all distances between objects belonging to the studied classes;
   d) group average-link, as a group average, in which the distance between two classes is the arithmetic average of all distances between objects belonging to the considered classes;
   e) centroid, i.e. the center of gravity – the distance between the two classes, which is estimated as the distance between the centers (centers) of the considered classes;
   f) ward – the grouping criterion is based on minimizing the variance (square of the difference) of the distance between objects belonging to the considered classes;
5. determination of the number of classes;
6. classification evaluation.

It is also worth emphasizing that the classification of the same set of objects in the same feature space, which (classification) will be carried out using different methods, very often can lead to different results. Hence, there is a need to assess the quality of the classification carried out in this way. The available metrics allow for an unambiguous assessment of the relationship that characterizes two objects classified into different classes. Therefore, their heterogeneity is investigated. Among many measures, the following estimators are widely used:

- individual – they are used to assess individual classes from the point of view of their homogeneity and heterogeneity;
- homogeneity – they examine the compactness of all created classes; lower values of these measures indicate a better classification, which means that the created classes are more compact;
- heterogeneity – they examine the distinctiveness of all created classes; higher values of these measures indicate a better classification, i.e. greater differentiation of the created classes (then they differ more from each other);
- correctness of clusters – they are used to assess the quality of the entire classification and are usually constructed as quotients of homogeneity and heterogeneity measures; they are based on the maximum and average distance between the objects of the classified community, while their lower values mean a more correct classification;

It should be noted that each of the above-mentioned metrics uses the estimated distance matrix of the examined objects as input data. As you can see, classification methods are very widely supported by the statistical and mathematical apparatus. Currently, there are many methods of classifying a set of objects, among which the following can be mentioned:

- agglomeration combinatorial methods (eleven methods);
- agglomeration non-combinatorial methods (two strategies);
- method – FANNY algorithm (Fuzzy Analysis);

Considering the problem of choosing the appropriate GIS class system in the current aspect, in addition to the previously presented ranking strategies, classification methods were used. For the normalized form of data, classification was carried out using agglomeration combinatorial methods, for which distances were estimated using the Euclidean metric (i.e. for $p=2$). The performed classification grouped and allocated the considered objects to particular classes. The direct result of the classification of objects is the resulting matrix, while its visual representation is presented later, using the created dendrogram of objects. It presents (dendrogram) in a very legible way individual categories to which each of the eighteen considered systems was allocated.

The dendrogram presented on Fig. 4 is a graphical result of the classification of the eighteen considered GIS systems. In order to be able to interpret its structure more precisely, the idea of classifying using hierarchical and non-hierarchical methods will be outlined. The second of the mentioned types is based on the arbitrary determination of a certain size of the similarity measure of objects, which is a threshold value for their division. Area methods are mainly used here, in which the fragmentation of a set of systems (elements) into appropriate classes is carried out by assigning (classifying) them to previously distinguished areas in the feature space of objects. As part of non-hierarchical methods, e.g. the hyperlink technique is very often used. Hierarchical methods constitute a contrasting group, which, from the point of view of the study, are very important. Hierarchical methods lead to the extraction of a complete hierarchy.

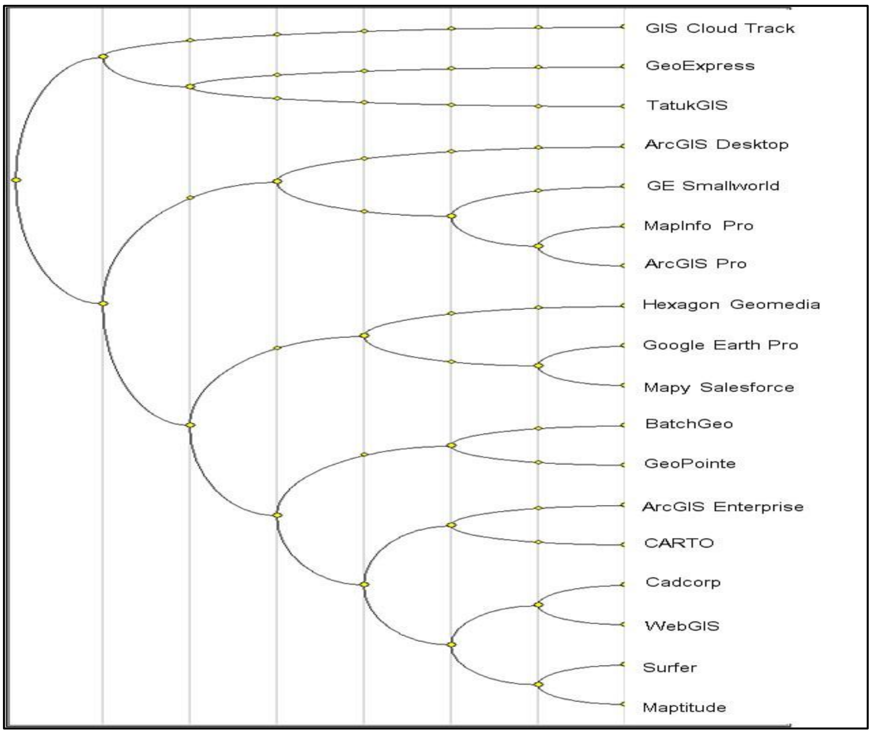


Fig. 4. Dendrogram as a graphical representation of the classification results of the tested systems
Source: own study

These methods create cluster structures that can be represented by a dendrogram. Dendritic techniques (another name for hierarchical methods) allow to obtain such a hierarchy of classes, in which a class belonging to a higher hierarchical level (level) is a multiplicity sum of disjoint classes belonging to the immediately lower level. The considered non-hierarchical methods in a direct way, omitting all intermediate sets, allow to obtain classes at the lowest level of the hierarchical structure.

These hierarchical structures are subject to further decomposition into agglomeration and division methods, deglomeration. In the first of them, each object is considered as an autonomous class, containing a single instance – one element, object (similar to the Singleton pattern). Then the classification process is based on combining these classes until an interesting – desired division is obtained. In the opposite, deglomeration approach, a complex set of objects is treated as a single class, which is then subject to gradual decomposition (division).

When making a proper analysis of the obtained results, the presented dendrogram should be carefully analyzed. It presents in a graphic way the division of the set of examined GIS systems, obtained in the process of its classification. When grouping objects, the mathematical apparatus turned out to be very helpful, which showed the degree of correctness and quality of the model through a set of appropriate indicators (indicators). The discussed group of measures that were used to validate the model structure included:

1. an h-type meter that tests the heterogeneity of the indicated class; it has the following three degrees showing heterogeneity:
   a) h** – very good heterogeneity;
   b) h* – good heterogeneity;
   c) h – weak but acceptable heterogeneity;
   d) h'– the weakest degree of isolated point heterogeneity; occurs when a class containing only one object;
2. type H meter, which tests the quality of the entire classification carried out; it also has three indicators that show differentiation:
   a) H** – very good classification quality; shows that all classes have degree h** heterogeneity;
   b) H* – good classification, in which all classes have heterogeneity to a minimum degree of h;
   c) H – weak but acceptable classification; for this indicator to take the value of H, all classes must have a differentiation of at least degree h;
3. estimator – SC (Silhoute Coefficient) algorithm shaping its value in the range closed on both sides (sharp) from zero to one. This coefficient is used when the evaluation has to be done with the model itself. When the Silhouette Coefficient takes a higher value, it reflects a model with better quality and number of clusters. The SC metric is determined for each sample and is expressed by the following relationship:

$$s = \frac{b-a}{\max(a,b)}$$

where:

a – average distance between the sample and the remaining points within the class;

b – average distance between the sample and all others points in the next closest cluster.

From the presented group of measures, it can be deduced how much value added is brought by the mathematical apparatus, which is fully and 100% applicable in the problem under consideration.

Returning to the resulting dendrite and its final analysis, it is worth mentioning that through a series of studies it was noticed that this method is characterized by high sensitivity to the data. After a slight change in the configuration of the values of selected attributes of the input matrix, both the construction of the dendrogram itself and its individual measures h, H, SC change. It was also noticed that for certain values the main cut (constituting the classification of nodes in the first order counting from the root - the root of the dendrogram) contains an error which is not propagated in the subsequent classifications. Examining the successive values of the coefficients for each row and object of the dendrogram in more detail, it can be seen that the quality of grouping is significantly improving, reaching a very good level of both the degree of differentiation (heterogeneity) and the classification itself. The obtained dendrogram was tagged with appropriate indexes for subsequent cuts of the set of examined objects – GIS systems (Fig. 5):

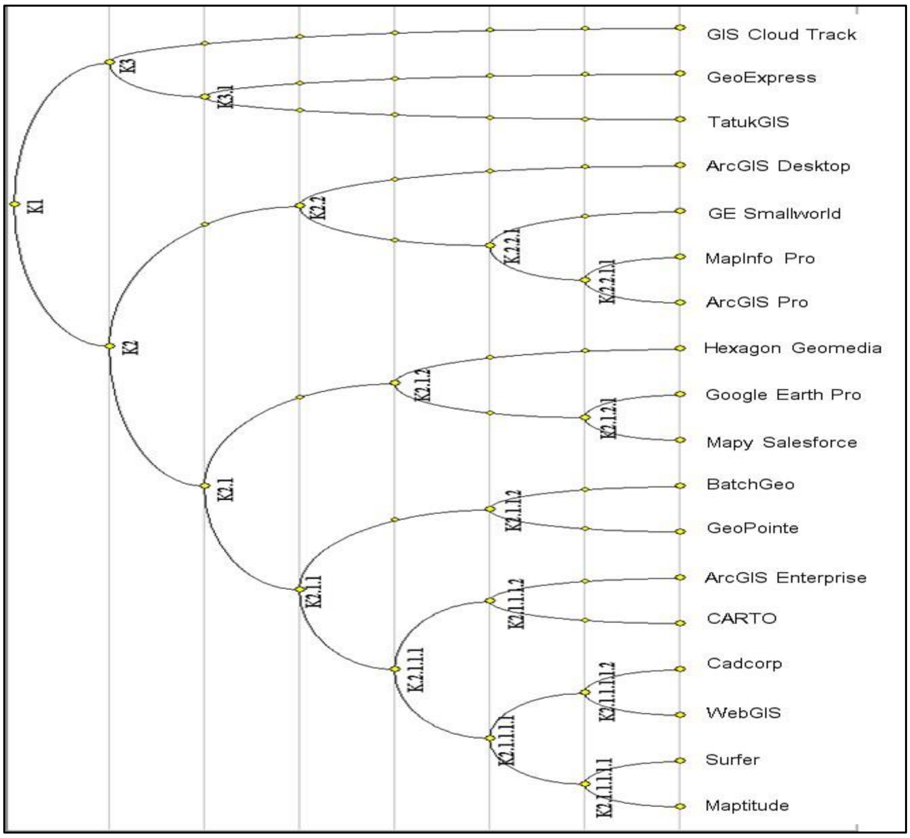


Fig. 5. The resulting dendrogram with appropriately tagged cuts of individual sets
Source: own study

The dendrogram shows the obtained distribution of classes and the association of individual systems to the appropriate sets. The operation of grouping objects consists in dividing the dendrite into appropriate units (parts) that contain homogeneous sets of objects. The first main, abstract class is superior, while the next ones are further decomposed. The final division of classes is presented in the table 2.

For the division into subsequent classes, each cut was assigned an appropriate value, indicating the quality of the heterogeneity of the resulting sets. Subsequent numerical ratings reflect the degree of class differentiation. The value of one means the basic, sufficient and poor quality of the division, while the score of "five" identifies the desired result, i.e. very good, which proves the greatest dissimilarity of the newly created sets. It is obvious here that the best division is the one that guarantees obtaining two most different subsets, where the objects in each of them are as homogeneous as possible in relation to each other. In the initial phases (stages) of creating a tree, this quality is improved, which in turn allows to obtain the most heterogeneous classes in relation to each other. This dependence has been shown in the form of the Pareto diagram presented below (Fig. 6).

Table 2. Sets of objects and their features created as a result of classification

Class/category	The number of instances – objects contained in the class	Variation index – heterogeneity of sets
K1	18	1
K2	15	2
K2.1	11	2.5
K2.2	4	2.5
K2.1.1	8	3
K2.1.1.1	6	3.5
K2.1.1.2	2	3.5
K3	3	4
K2.1.2	3	4
K2.1.1.1.1	4	4.5
K2.2.1	3	4.5
K3.1	2	5
K2.1.1.1.2	2	5
K2.1.1.1.1.1	2	5
K2.1.1.1.1.2	2	5
K.2.1.2.1	2	5
K2.2.1.1	2	5

Source: own study

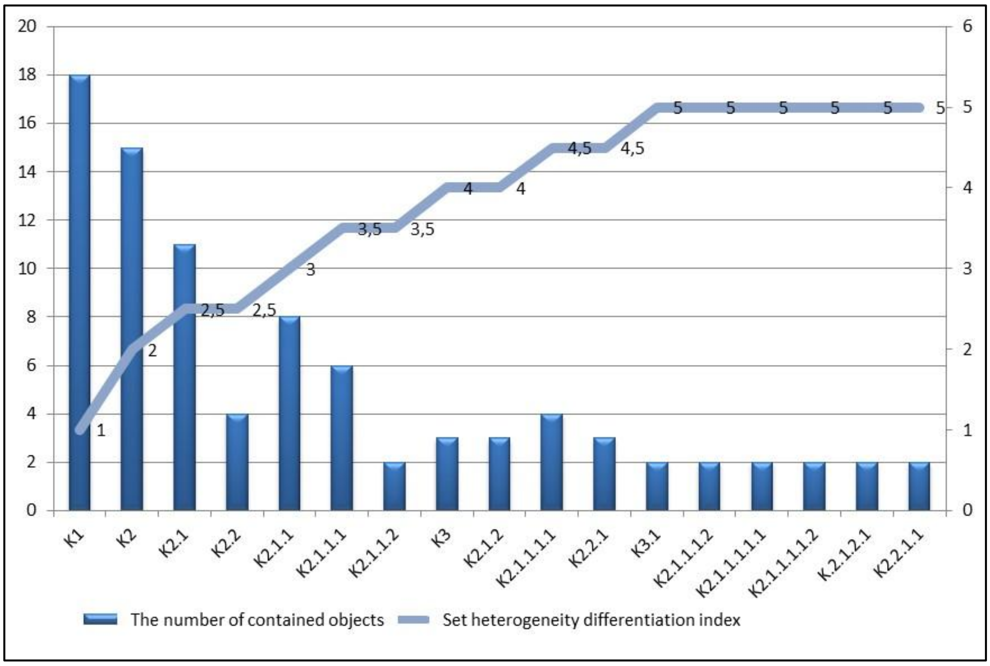


Fig. 6. Pareto diagram showing the diversity of heterogeneity of sets
Source: own study

The above considerations directly lead to the conclusion that the greatest separation of objects was noticed at the lowest level of the tree structure, which indicates the correct construction of the dendrogram and the appropriately high quality of the classification, which is confirmed by the SC, h, H indicators. An effort was made to present the qualitative distribution of the standard of individual classification cuts, carried out on a set of tested systems, using the author's magic quadran'ut form (Fig. 7):

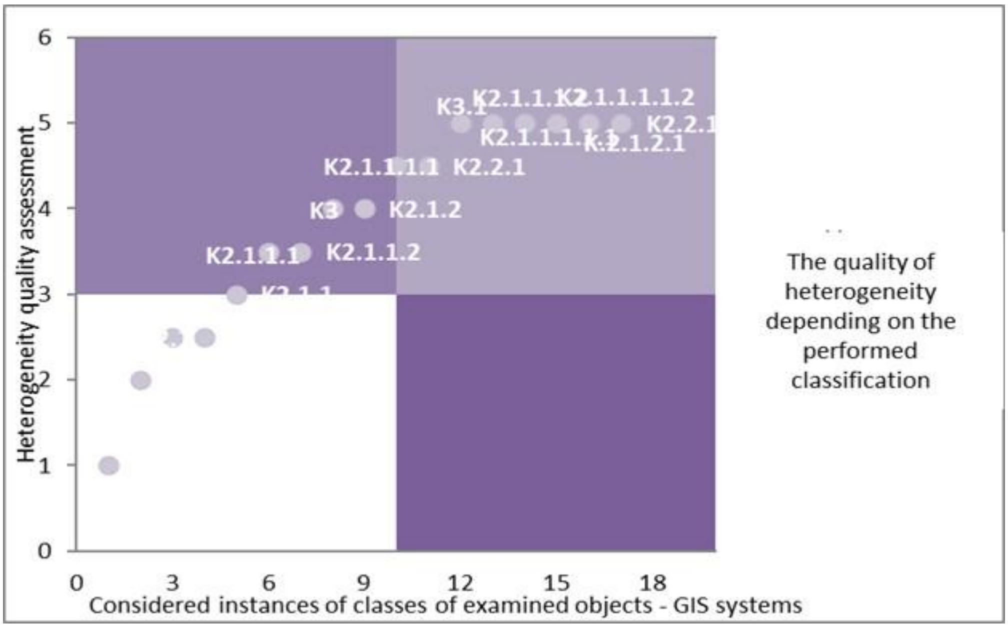


Fig. 7. Quality of collection heterogeneity depending on classification
Source: own study

Conclusion

Through the use of a developed mathematical and statistical apparatus, it was shown how the applied techniques support decision-makers, providing them with credible and reliable support. The results obtained in the work fully agree with the current state, which characterizes the considered set of GIS systems. At the same time, it is mentioned that the obtained results should be treated as advice and guidance in the decision-making process. Elements of taxonomy and classification, which determine the set of compared systems, were integrated into the analysis. Objects characterized by appropriate values of selected coefficients (e.g. heterogeneity and homogeneity) are allocated in appropriate collections (classes, containers), grouping units with similar properties. It was also noted that the differences in the obtained results are very often implied by the specificity of the methodology used - the algorithm. Therefore, it was decided to use a taxonomic approach, evaluating the studied GIS systems not in terms of best-worst, but in the context of classification, allocating them to appropriate systems comparison sets. The applied approach brings an advantage to the analysis of systems and leaves the user the opportunity to make the final choice between the set of objects that he is interested in. Advantage is in other words profit, benefit. It is about the intangible advantage that distinguishes this work, among others, by evaluating the considered set of objects through many different techniques, strategies and methods, which makes the obtained results credible and representative. This aspect is particularly important due to the fact that the adopted criteria do not take into account all the outlines in which a given GIS will be installed, embedded and used. Therefore, due to the knowledge of the specifics of the target GIS workplace, the final decision is left to the user. The taxonomy used complements and confirms all dependencies that have been detected during the evaluation and comparison of systems. Another element of innovation is the visualization of all obtained results by using the structures of the Voronoi diagram and dendrogram, which are a clear and accessible form of presentation and inference of all existing dependencies.

Acknowledgments

This work was financed/co-financed by Military University of Technology under research project UGB 22-810.

References

Bolstad P. (2019). GIS Fundamentals: The First Text on Geographic Information Systems (6th ed.). XanEdu. ISBN 978-1-59399-552-2.

Broome F.R., Meixler D.B. (1990). Structure of the TIGER Database. Cartography and geographic information systems. 17(1):39–47. doi:10.1559/152304090784005859. ISSN 1050-9844.

Chang, Kang-tsung (2016). Introduction to Geographic Information Systems (9th ed.). McGraw-Hill. p. 1. ISBN 978-1-259-92964-9.

DeMers M. (2009). Fundamentals of geographic information systems (4th ed.). John Wiley & Sons, Inc. ISBN 978-0-470-12906-7.

Fitzgerald J.H. (2007). Map printing methods. Archived from the original on 4 June 2007. Retrieved 9 June 2007.

Fu, P. and Sun. J. (2010). Web GIS: principles and applications. ESRI Press. Redlands, California. ISBN 1-58948-245-X.

Goodchild M.F. (2010). Twenty Years of Progress: GIScience in 2010. Journal of Spatial Information Science (1). doi:10.5311/JOSIS.2010.1.2.

http://www.esri.com. "GIS History | The Early History and Future of GIS – Esri". [accessed January 12, 2023].

http://www.wiki.osgeo.org/wiki. Open-Source GIS History – OSGeo Wiki Editors. [accessed March 21, 2022].

Lubis M.Z., Taki H.M., Anurogo W., Pamungkas D.S., Wicaksono P., Aprilliyanti T. (2017). Mapping the distribution of potential land drought in Batam Island using the integration of remote sensing and geographic information systems (GIS) IOP Conf. Series: Earth and Environmental Science 98 012012 DOI:10.1088/1755-1315/98/1/012012

MacHarg I.L. (1971). Design with nature. Natural History Press. OCLC 902596436.

Maliene V., Grigonis V., Palevičius V., Griffiths S. (2011). "Geographic Information System: Old Rules with New Possibilities". Urban Design International. 16(1):1–6. doi:10.1057/udi.2010.25. S2CID 110827951.

Miłek M., Kiedrowicz M., Stanik J., Napiórkowski J. (2023a). Multi-criteria comparative analysis of GIS class systems, GIS Odyssey Journal, vol. 3, no. 1, pp. 97–122.

Miłek M., Kiedrowicz M., Stanik J., Napiórkowski J. (2023b). Comparative characteristics of GIS using the AHP method, GIS Odyssey Journal, vol. 3, no. 1, pp. 41–72.

Mironova Y.N. (2018). Some aspects of Geoinformatics. Monograph (Moscow: Pero Publishing House) p. 98.

Mironova Y.N. (2020). Mathematical Methods in Geoinformatics 2020 International MultiConference on Industrial Engineering and Modern Technologies (FarEastCon) 6-9 Oct. 2020, (Vladivostok, Russia) URL: https://ieeexplore.ieee.org/document/9271594 DOI: 10.1109/FarEastCon50210.2020.9271594

Neema N.L., Zhenfeng S., Mgelwa A., Nana Y., Danquah T., Orhan A. and Sanga C.A. (2020). A Fuzzy Logic-Based Approach for Modeling Uncertainty in Open Geospatial Data on Landfill Suitability Analysis International Journal of Geo-Information. ISPRS Int. J. Geo-Inf. 9,737 URL: www.mdpi.com/journal/ijgi doi:10.3390/ijgi9120737

Risky Y.S., Wahyu N., Raden R., Yudha A., Brada I.T., Artanti P.Z., Rosyita A., Wisudawan P.D. and Sudaryatno (2018). Using remote sensing and geographic information system (GIS) for peak discharge estimating in catchment of way ratai, pesawaran district, lampung province IOP conf. Series: Earth and Environmental Science 165 012032 DOI:10.1088/1755-1315/165/1/012032

Somers R., 1996, Managing GIS: What do we really know? In Proceedings of the Conference on 'Managing Geographic Information Systems for Success' (Melbourne: Center for Geographic Information Systems and Modelling, The University of Melbourne), pp. 1–10.

Tobler W. (2009). Automation and cartography. Geographic overview. 49(4): 526–534. doi:10.2307/212211. JSTOR 212211. Retrieved March 10, 2022.

Vastag P.H., Thum P.G., Niemann Jr. B.J., 1994, Project LOCALIS: implementing LIS/GIS in local government. URISA Journal, 6, pp. 78–83.

Weill P., Broadbent M., 1994, Infrastructure mix and match. Management Information Systems (October), pp. 2–55.

Xuan, Zhu (2016). GIS for environmental applications A practical approach. ISBN 9780415829069. OCLC 1020670155.

Zhang J., Zhang J., Du X., Hou K., Qiao M. (2017). An overview of ecological montirng based on geographic information system (GIS) and remote sensing (RS) technology in Chna IOP Conf. Series: Earth and Environmental Science 94 012056 DOI:10.1088/1755-1315/94/1/012056