

Ilona Pawełoszek, Cezary Stępnia

Częstochowa University of Technology

e-mail: {ipaweloszek; cestep}@zim.pcz.pl

**SPATIAL VISUALIZATION TOOLS
IN BUSINESS INFORMATION SYSTEMS**

Abstract: The paper presents a prototype application – Spatial Visualization Tool (SVT), that can be used to visualize data from different kinds of Business Information Systems and external sources such as the Internet and document bases in the form of an interactive map. The concept of the paper is based on the assumption that all business data can be visualized in a way analogous to geographic maps. The manner of such a visualization can be very legible and facilitate business decision-making when it comes to analyzing large amounts of data. The first part of the paper introduces the reader to selected aspects of Geographic Information Systems methodology and briefly presents the algorithm of SVT. The second part explains the theoretical algorithm of acquiring data and performing a spatial visualization. The last section of the paper covers the practical example of SVT application to select candidates for a job. The prototype tool was constructed using Microsoft Excel with Visual Basic for Applications. We also present the idea to extend the functionality of SVT using semantic Web technologies for recognizing data sources and types to facilitate the automation of data visualization.

Keywords: data visualisation, spatial information systems, business information systems.

1. Introduction

Information systems of contemporary enterprises are the subject of continuous development. On one hand we can see the changes in technology – the Internet and the related new solutions have changed the face of data and information processing, and on the other hand the scope of information systems changes by the continual addition of functional enhancements.

This paper considers the subject of enterprise information systems development by the application of visualization techniques derived from Spatial Information Systems (SIS) technology. SIS are considered an extension of Geographic Information Systems (GIS), which may also be applied for purposes other than strictly geographic visualizations [Litwin, Myrda 2005].

The aim of this paper is to present an algorithm of using spatial visualization tools (SVT) for visualizing non-geographic data in Business Information Systems (BIS). The paper consists of two parts. In the first part we present the algorithm, and

in the next part the practical example of its application is demonstrated. The practical example of using SVT focuses on the pre-selection of candidates applying for a job or a position in a new project.

2. Theoretical algorithm

2.1. The algorithm environment

The proposed algorithm may be applied to visualizations of objects (O) and processes (P) in organizations of different types. However, it is devoted mainly for visualizing large quantities of uniform phenomena or sets of objects. It facilitates the presentation or comparison of many objects or events, in cases when conventional tables (even with sorting capabilities) are not convenient to use. Usually tables or charts presenting over 50 occurrences become less legible and so it is difficult to evaluate and extract desirable data if the choice must be made based on a few features. Such problems happen very often in many areas of business practice and their solution depends on having the right information technology tools for processing and visualizing data.

Essential data sources of the enterprises are databases of information systems, but data may also be complemented with external resources. The most common types of computer-based information systems used in business organizations are MRP/ERP, in large firms they are often extended with Business Intelligence (BI) modules. Other types of commonly used information systems are Customer Relationship Management (CRM) or Supply Chain Management (SCM). In sectors of manufacturing industry very important applications are Computer Aided Designing/Computer Aided Manufacturing (CAD/CAM) systems. Geographic Information Systems in turn are used, inter alia, by emergency services, public administration and the communication sector.

Enterprises use many types of information systems to support different activities and processes. Data is usually collected in relational databases. The relational model is the most common of all the different types of databases, most often it is also an inherent part of MRP/ERP systems. The data in a relational database is stored in highly structured and formalized data tables. Each table has a key field which is used to connect it to other tables. All the operations like searching, retrieving and manipulating data are realized by sending formalized logical queries formulated in different dialects of SQL. Databases describe predefined types of economic events, which can be grouped in classes of objects K(O) and have their attributes (A). The main and most powerful analytical tools are BI systems, which enable different kinds of data processing and visualizations by incorporating Data Mining techniques, OLAP, and management dashboards.

The aforementioned tools enable performing different kinds of analyses focusing on selected classes of objects K(O) or events K(E). However these analyses are conducted mainly on the basis of data stored in MRP/ERP, although in many other

systems the data may be stored in document bases (DocB) containing loosely structured documents describing objects (O) or events (E). Text documents, like PDFs, HTML pages, spreadsheets or multimedia presentations contain a lot of valuable information which is inaccessible to BI systems in the conventional, standardized way. There have been many studies focusing on mining loosely structured information to extract data, patterns and knowledge [Miner et al. 2012]. Although extracting and integrating data from multiple heterogeneous repositories is a difficult task, it has become possible thanks to harnessing semantic Web technologies. Ontologies and reasoners, along with semantic query languages like SPARQL [Hitzler, Krötzsch, Rudolph 2009], make it possible to find relationships between data from the Internet [Gomez-Perez, Corcho, Fernandez-Lopez 2004], document bases and the data bases of MRP/ERP systems.

The main task of SVT is to enable data processing from all the accessible sources, focusing on selected classes of objects $K(O)$ with predefined criteria in mind, and to visualize all the objects on mathematical maps. The maps are built by using cartographic techniques derived from the GIS concept.

2.2. Chosen aspects of GIS methodology

The GIS concept is based on cartographic methodology enhanced with facilities provided by IT tools for creating maps. In the following considerations we refer mainly to the following aspects of GIS methodology:

- possibility to define any topological spaces for visualization,
- applying cartographic techniques to allocate objects on a grid,
- simultaneous allocation of many objects (O) on maps,
- using cartographic symbolization of objects on maps.

Graphical visualization has been recently considered to be the most legible form of presenting data. Legibility is the key matter, especially when the described entity is composed of many objects (O). Moreover, the cartographic visualization enables the simultaneous presentation of numerous classes of objects $K(O)$ and can be realized according to many different criteria. The application of GIS tools facilitates the suitable choice of the map scale and thematic layers. Objects can be presented as two-dimensional, three-dimensional or using radial coordinates. There is also a possibility of the geometric transformation of the space and objects (O) coordinates.

A cartographic grid enables one to define the space of visualization, for this purpose the spatial attributes must be defined. Depending on the type of space (2D, 3D, etc.) we have to define the corresponding number of spatial attributes. There can be more than three dimensions defined, however only three of them can be used for allocating the given object.

Therefore, it is necessary to define spatial attributes (AS) for all the objects that undergo visualization. The aforementioned attributes should have constant values $V(A_s)$, so the unambiguous identification of each object on the grid is possible. Thanks to SVT, the presented set of objects can be visualized according to different

criteria, which depend on the features (describing attributes A_D) of the objects of the given class $K(O)$. The presentation can be one or multi-criteria. Additionally, it is possible to overlay chosen thematic layers to present the set of objects in different contexts, if needed. While performing the visualization, one can apply the cartographic techniques of symbolizing objects. Depending on the value of the describing attribute $W(A_D)$, every object can be presented as a small/large symbol or by a given color. The rules of visualization are in the hands of the map designer.

2.3. The algorithm procedure

SVT can be applied in numerous areas of the enterprise's business activities. They can also support the processes of decision-making, like solving particular unique tasks. However, regardless of the goal, SVT may be used according to the uniform algorithm, which is structured as follows:

1. Define classes of objects $K(O)$.
2. Identify objects (O) .
3. Set up attributes (A) .
4. Distinguish spatial (A_S) and describing (A_D) attributes.
5. Define space (U) :
 - 5a. random space,
 - 5b. semi-random space,
 - 5c. attribute space.
6. Choose describing attributes (A_D) for visualization
 - 6a. one-attribute visualization,
 - 6b. multi-attribute visualization,
 - 6c. standardized visualization.
7. Visualize the phenomenon.
8. Make a decision.

The described algorithm can be used as a base to build a decision support tool as a part of the Decision Support System (DSS) or become an additional tool enhancing the functionality of Management Information Systems (such as ERP).

3. Description of theoretical algorithm's procedure

3.1. Step one of the theoretical algorithm

In the first step of the algorithm we need to define precisely its aim. The algorithm can describe one or many collections $K(O)$ of objects or events characterized by the same features (attributes (A)), procedures or processes. In the case of describing the processes, the algorithm can describe theoretical process flow or its real-world instances (the only condition to visualize processes is to collect relevant data about process flow – the so-called process logs).

Having in mind the goal of the visualization, one can identify the classes of objects $K(O)$ or events $K(E)$, which will undergo visualization. The given class of object should have identified, accessible data sources. These may be, for example, tables from the relational database of the ERP system, data from text document bases or the Internet (Web pages).

3.2. Step two – the choice of objects

On the basis of the class of objects $K(O)$ defined, one should indicate the objects (O) belonging to the described collection that will be visualized. The visualization can contain the whole class (represented by all the objects belonging to the class: all the records from the source table) or its part (records selected with regard to a given criterion).

3.3. Step three – attributes definition

Objects (O) described in databases (DB) have attributes. In this case the database fields (or table columns) represent object attributes (A). However, the data describing the given object can reside in many tables related to one another in a set of tuples (1:1, 1:n, n:m or 1:0). In the database system, some of the attributes are mandatory and others are not required (their value may be “null”). So, if the attribute of the latter type is chosen for visualization, there will be a need to decide how to present them in the visualization. The choice of attributes is limited to the fields accessible in the given dataset. So if we have data from many disparate sources (DocB or the Internet), there will be a need to extract them and set them in a common table for the purpose of visualization. The issue of extracting and using data from the Internet is a matter of continuous consideration focusing on the aspect of the trustworthiness of the data sources [Golbeck 2006]. When using data from the Internet it is indispensable to assess the quality of data, make the right interpretation and convert it to a common format. The data should be then added to the one source table (ST) with the number of columns with names corresponding to attributes (A). Each column of the source table can constitute a base for creating a map layer (ML) for visualizing objects, features.

The visualization can be enhanced by interactive, dynamic elements – like hyperlinks – redirecting the user to additional data sources, providing descriptions that cannot be directly visualized in a cartographic way.

3.4. Step four – differentiation of attributes

A cartographic visualization usually presents at least two types of data describing each object. On the one hand, it presents the location of object (O) in space (spatial allocation – SA), and on the other, it indicates the weight (value of the attribute

$V(A)$) of the object considering the given criterion that has been chosen as a thematic layer of a map, so the visualized attribute establishes the map layer.

The visualization can be multi-layered, in such cases the individual objects can be described with a few symbols corresponding to different layers of the map. Considering the legibility of the map, it is not recommended to overlay too many attributes at once.

The differentiation of spatial and describing attributes is dependent on the purpose of the visualization. The starting point is a definition of the spatial attributes, because they define the space (U – universe) of visualization. The definition of space may have a twofold character:

- projecting the spatial allocation of objects,
- scattering the objects on a map for improved readability.

When creating a visualization, it is worth remembering that AS should be constant and possible to use in different visualizations. The remaining attributes – which are not spatial – should be treated as describing ones.

3.5. Step five – defining of the universe

In practice defining means describing the space of the visualization. There are three general methods of choosing spatial attributes (A_s).

Method A consists in defining totally random attributes by using a randomizer (an algorithm generating random numbers from a given range). The generated numbers can be used to allocate each object in relation to the X, Y and Z axes. Sometimes there can be more than three spatial attributes, but then it not possible to visualize them.

Method B is used when the randomizer fails in allocating all the objects in a legible way. Method B consists in generating a two-dimensional (or three-dimensional if needed) array of sequential numbers. Every object of the array is then assigned to one record in the source table. Another way to assign spatial attributes is to divide the collection of data into groups and create separate visualizations for the groups, and then combine the visualizations to create one map. Extracting the groups of elements can be done manually or by harnessing a clustering algorithm, for example k-means [Chen 2004].

Method C uses the real attributes of the objects visualized on the map. In this case the spatial attributes can be the geographic coordinates(latitude and longitude) of each of the given objects. Another idea to place the objects on a map is to use some of their attributes to define the X and Y axes, for example a birthdate can be plotted on the x-axis, and the numeric version of the surname on the y-axis.

Regardless of the choice of the method and the spatial attributes (A_s) selected, the universe of the visualization must be defined in a way that lets us unambiguously recognize and locate each object (O) on the map.

3.6. Step six – the choice of describing attributes

The visualization with SVT can be conducted using diverse techniques of SIS. The same data collection may be visualized in many different ways by choosing various spatial and describing attributes. Generally, SVT uses cartographic methods of symbolizing objects while generating maps.

These methods focus on the rules of choosing:

- symbols for objects,
- size of symbols and captions,
- colour, shades and texture.

By applying rules of visualization, every object is properly and unambiguously allocated and its symbol is adequate to its importance on the given map layer.

While creating the map, it is necessary to have in mind the aim of the visualization, because it impacts on the desired composition of the map. While solving decision problems or describing business events, we need to point out the criteria of the visualization. Generally there are three methods of choosing the criteria.

Method A consists in visualizing all the objects (O) according to one criterion, i.e. according to one describing attribute (A_D). Then the symbols for the given series are chosen, and the ranges of values are defended for each symbol. Additional information can be visualized by adding captions of different sizes. The size of the symbol may be directly dependent on the value of the given attribute $V(A_D)$ of the object. The visualization process consists in retrieving the value of spatial attribute (A_S) and describing attribute (A_D) for each object from a source table, and then allocating it on a map in a place resulting from $V(A_S)$ values. Next, the symbol is assigned adequately to the value of $V(A_D)$.

The activities mentioned above are repeated for different criteria, so the layers of the map are created.

The result of the iteration is the presentation of each object every time in the same place on the map, but marked with different symbols reflecting the values of consecutive attributes (A_D).

Method B makes it possible to visualize objects simultaneously according to many attributes (A_D). This method consists in overlying or setting together a few symbols describing each of the given objects, so the multilayer map is created. The rules of choosing the symbols are similar to those described in method A. Applying directly the SIS tools, it is possible to switch on or off particular thematic layers for improving the readability of visualization in a real-time manner. Method B lets us reflect a lot of data about the object, but often it does not provide an optimized view for selecting objects. Too many symbols, colors and shapes, make it difficult for a human to select the most important objects.

Method C provides the optimization of visualization based on statistical techniques. It is a single-criteria visualization that consists in standardizing the description of each object. The standardization can be achieved by proceeding with the following algorithm (OSA – Object Standardisation Algorithm):

1. Choose the attributes for evaluating objects.
2. Standardize the values of the attributes (transform the data to comparable scales, for example from the range $\langle 0,1 \rangle$).
3. Weight attributes.
4. Calculate the standardized value $V(A_p)$ of attributes for each object.
The value $V(A_p)$ will be a base for visualization conducted according to method A.

3.7. Step seven – the visualization

The visualization consists in generating a map on the basis of data describing the chosen class of the object $K(O)$ or the fragment of the class. As stated earlier, the allocation of the object was made on the basis of spatial attributes' values $V(AS)$, and the symbolic representation was made on the basis of describing attributes (A_p) or standardized attributes $V(A_p)$. The principles of cartographic symbolization of objects assume that the most important objects are presented in the most expressive way.

Generating a map does not solve the problem of making a decision, however it sheds some light on the status of all the objects visualized. For enhancing the effect of the visualization of particular objects it is possible to create hyperlinks for the symbols, redirecting the user to the resources describing the object in more detail.

3.8. Step eight – decision-making process

The process of decision-making is always the prerogative of a manager. SVT is just one of the tools facilitating understanding the presented entities. SVT makes it easier to analyze large data collections. Utilizing the map scales, the user can zoom in and out the view to eliminate the less important (smaller) objects on the map in a particular thematic layer view. Using a small scale lets us visualize only the most important objects. Simultaneously with zooming, it is possible to switch on new thematic layers, and visualize them if they are adhering to the same metric spaces. Thanks to the capabilities mentioned above the decision process is multi-criteria.

4. Practical example

The algorithm presented in this paper was applied to staff selection for a new project in a large company. In regard to the large number of candidates, a logical framework for the SVT was developed. In this case, the role of SVT is to visualize the data from the large database of candidates. The tool was developed using the programming language Visual Basic for Applications (VBA) in the Microsoft Excel 2010 environment. The choice of the programming environment was dictated by the following arguments: the prevalence of Microsoft's Excel for analyzing data among entrepreneurs, compatibility with databases, large capabilities of statistical data

processing, availability of many different visualization techniques and the possibility to enhance the visualization by procedures written in VBA [Green et. al 2007].

In this case, the selected visualization technique was the scatter-bubble chart. This type of chart provides the possibility to allocate the points in strictly defined places in the grid, and moreover it offers the rich functionalities of choosing the color, size, symbol, and the caption to visualize the information.

Recruiting staff is a time-consuming task, demanding the analysis of a large amount of data sent by candidates, therefore the tool for the simultaneous visualization of all the candidates, data and facilitating the choice is of great value for decision-makers.

4.1. Step 1

The first step in the process of acquiring staff is data collection. Next, the data undergoes analysis regarding the selection of potential candidates for a position, with whom the job interviews will be conducted, and ultimately making a choice. The candidates may be selected among the employees of the given company or can be persons from outside the organization. Therefore the data used in the recruiting process may originate from different sources.

In the case of recruitment among employees of the given organization the sources of data were:

- databases of the Human Resources Department (containing personal data, skills, qualifications, etc.),
- databases of other departments where the candidates have worked so far (data about achievements such as: sales volume, work efficiency).

The candidates from outside the organization are selected on the basis of:

- documents sent, such as: application, motivation letter, curriculum vitae. The most convenient way to collect data is via an electronic form published on the company's website, because it ensures that the data will be valid and written directly into a database without the need for retyping,
- additional data for candidates verification was acquired from the Internet, for example from social networking sites, branch services etc.

4.2. Step 2

The second step is the collection of data and labelling the candidates. The data should be set in a uniform way to facilitate visualization. The most appropriate way in this case is to define the source table (ST) and the scope of data it must contain. Every record (row) of the table presents the data of one candidate. Each of them should be labeled with a unique identifier O_i . Moreover, the set of features (attributes $A(O_i)$) should be selected to describe each candidate. When defining attributes, one should consider the kind and scope of the collected dataset and the overall aim of the

visualization. In this way the set of spatial attributes is created ($A_s(O_i)$), as well as the set of describing attributes ($A_D(O_i)$).

While arranging the data in the source table (ST), the describing attributes of a particular type should be placed in columns (fields). The number of columns must be limited and uniform for every candidate.

After collecting the data, the source table of 108 records and 20 fields was created in a Microsoft Excel spreadsheet.

4.3. Step 3

The next step is the identification of the most important features of the candidates for the given positions. In the case at hand the selected features are: education (level, profile and the graduate school), the level of knowledge of foreign languages (score in a language test), age, gender, and the score of the test for professional knowledge.

Describing attributes can be of different types, in this case:

- quantitative (measurable) – age, the score in the test, sales level;
- qualitative (immeasurable)
 - dichotomic – can represent one of two possible values – gender,
 - politomic – can represent multiple values -profession, graduated school,
 - ordinal – has values that function as labels rather than as numbers – knowledge of language – basic, intermediate, advanced,
 - categorical (or nominal) – gender, profession, graduated school.

In the case of selecting candidates described here, all of the aforementioned data types are present.

4.4. Step 4

Because we decided the visualization will be performed in two dimensional space, the two space attributes must be defined to identify every object in ST. The values of spatial attributes $V(A_s)$ determine the place of the object's allocation on the map. Spatial attributes in other words, are coordinates, and they can be natural numbers.

The coordinates are created by assigning the random numbers to each candidate, but it is very important for the numbers to be unique and properly differentiated. The differentiation of the numbers has to ensure that the symbols on the map will not superimpose and diminish the legibility of the visualization (see Figure 1). Regarding the situation presented in Figure 1, in the case at hand, we decided to assign consecutive numbers to each object from a source table to gain maximum readability. We decided to place the objects on the map in a grid of twelve columns and 10 rows. These numbers let us visualize all the 108 objects from the source table.

The attributes were assigned by adding two columns to the source table – for the X and Y coordinate. Then the iteration was performed for column X, assigning to each consecutive twelve cells a number ranging from 10 000 to 90 000 (each time

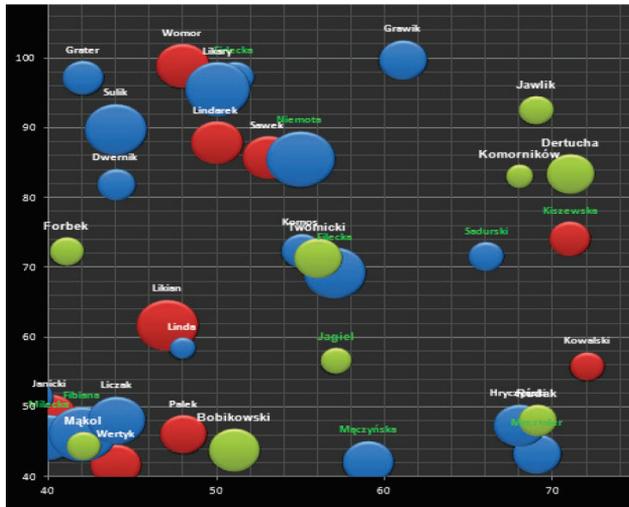


Figure 1. The example of random allocation of objects on the map

Source: own elaboration.



Figure 2. Allocation of objects on the map by means of assigning coordinates as consecutive numbers

Source: own elaboration.

the next iteration value was 10 000 bigger than the previous). The next iteration was performed to fill the Y coordinates, in this case every cell was given a consecutive number ranging from 10 000 to 175 000 (also with the change of 10 000), and such iteration was repeated 12 times – for all the cells in column Y.

The solution consisting in assigning consecutive numbers is a good choice when the allocation of the objects on the map does not reflect any attributes. Figure 2 represents the effect of such a visualization.

4.5. Step 5

The essence of visualization is to place the proper symbols on the map in predefined places. The features can be reflected in various ways depending on the user's preferences. In this case we used the following visual parameters:

- coordination in space,
- size (the diameter of disc),
- colour,
- caption (text and colour).

Certainly apart from those mentioned above, it is possible to use many other parameters. Good data visualization relies on the ability to effectively communicate data. The research [Card 2003] shows the different appropriateness of graphical visualizations (see Table 1).

Table 1. Relative appropriateness of graphical properties in visualising different variable types

Data type		Graphical properties of a symbol						
		Size	Coordinates	Grayscale/ colourscale	Colour	Texture	Shape	Text caption
Qualitat.ve	Ordinal	H	M	H	M	M	L	H
	Dichotomic	H	L	H	H	H	H	H
	Politomic	H	L	H	H	H	M	M
	Categorical	M	L	L	H	M	H	M
	Quantitative	H	H	M	M	M	L	L

Appropriateness: H – high, M – medium, L – low.

Source: own elaboration.

The selection of the parameters should be intuitively adjusted to the type of variable (feature). The type of the attributes and frequency of particular values in the set of data are the basic determinants of choosing the way of visualization.

For example, in the described case, seniority (the number of years of service in the company) is reflected as the size of symbol, and the level of English language skills is mapped by the color of each symbol. Qualitative features, such as the school graduated or surname, can be presented as text captions.

4.6. Step 6

The SVT is being developed by its authors to prompt the user to choose the particular form of visualization. It is realized by algorithms of data analysis and Ontology of Data Types. The Ontology of Data Types helps to recognize and label the type of data in each column of the source table. The ontology and related tools are currently in the phase of development.

The tools of data analysis let us decide whether the data in each column preserves the integrity and validity of type and format. The next step of the analysis is to decide about the character of the data (discrete or continuous, quantitative, qualitative, etc.) and how many sets or clusters can be extracted.

The outcomes of the data analysis are compared with the Ontology of Data Types, to determine the best (optimal) way of visualization.

In the SVT presented here, it is possible to switch thematic layers. Figure 2 presents an interface with buttons allowing the user to change the properties of the map. In this case we can see the following attributes visualized:

- seniority – reflected by the diameter of the disc,
- language skills level – presented by color,
- additionally every data point is annotated with a text label displaying the surname of the candidate, the color of the label is green for women and white for men.

There is also the possibility to change the content of the labels to display alternatively the shortened names of the schools graduated by the candidates, their profession or any other attribute values from the source table. The visualization is not limited to a few selected attributes, it is possible to define and switch on new layers. However, for constructing each single view, it is best to choose from two to five most important features. With more than from features the visualization will become hardly readable. The only limitation is the necessity to designate a uniform way of visualization for each attribute. For example, the color on a given thematic layer can be used only to visualize one feature (in Figure 2 it is the language skills attribute). If the size of the symbol expresses seniority, it is not possible to simultaneously present the age as the size of the symbol.

Some of the data types require pre-processing before they are visualized on the map. The pre-processing can consists in:

- weighting,
- ordering.

Weighting and ordering were applied to preprocess data about the level of languages skills. Ordering was performed by assigning each of the candidates the attribute of language knowledge: English, German and Russian on a basis of a language test. The result of the test was a score on a scale 0–100. For the needs of the recruiting process, three levels were selected: basic, intermediate and advanced.

Weighting was applied to the choice of the most important features and their desired values. The standardized attribute A_i was constructed. The knowledge of

English was of the highest priority and that feature was assigned with the weight of 0.5. Russian was assigned with the weight value of 0.3, and German 0.2. Then, for the uniform visualization of the three considered languages, the scores from the tests were multiplied by the corresponding weights and divided by the sum of the weights. The results of these calculations were the basis of creating the layer of weighted language skills.

4.7. Step 7

The choice of candidates for a job interview is supported by displaying the corresponding thematic layers reflecting the desired features. Then the symbols on the map with the appropriate size (the largest ones) and color are selected. The user responsible for selecting and inviting the candidates for the interview can easily generate a list of candidates by clicking the selected symbols on the map.

The list is actually a new table containing the whole records describing only selected candidates copied from the source table. The list of candidates is automatically created and appended in a new worksheet, which can be used later as a source for serial correspondence for printing or e-mailing invitations for a job interview.

4.8. Step 8

The final decision concerning employing the selected persons should be supported with detailed data. As was stated earlier, the verification of the candidate can be conducted on the basis of information retrieved from the Internet. This is a complicated task and difficult to automate. A particular problem is the verification of the truthfulness and relevance of the information. However, it is possible to automate the process of collecting such information. It is relatively easy to automate the data collection for the following tasks:

- determining whether there is any information about the given person in the publicly accessible World Wide Web,
- providing statistics about the documents containing the person's data, or information about him/her,
- determining the kind of web pages containing information about the person (subject of the web page, affiliation to an organization, character: formal or private, etc.).

For facilitating the searching activities in the web, the Custom Contextual Search (CCS) tool was prototyped. The tool is being developed by exploiting the Google Custom Search API [Google Developers 2012]. The aim of the CCS tool is to retrieve the data about a person on the basis of name, surname or e-mail address. The results are aggregated in a table containing hyperlinks to the web pages related to the given person, the description and designation of the page character.

The character of a page is determined in a twofold way: by extracting information from meta-tags, and by comparing the URL of a page and some of its elements with WebSite Types Ontology (WSTO) and a corresponding database describing the types of Internet resources.

The WSTO is currently developed by the authors for the needs of analyzing data from the Internet. The base classes of the WSTO are describing: discussion forums, social networks, websites of institutions and companies. The database contains already classified instances of web pages. If there is no information about the given website in the database, an attempt is made to identify it and describe on the basis of WSTO.

The data from the Internet is one of the components of each candidate's description, although it is treated as supplementary information. In future the semantic CCS tool is going to be integrated with VST so it will be possible to visualize and create a hyperlink to the data from the Internet directly on the maps.

5. Conclusions

In this article the prototype SVT was briefly presented along with the example of its exploitation in the recruiting process. In practice, the introduced tool constitutes the enhancement of contemporary business information systems and it is also possible to use it in settings different than business applications (such as public administration, healthcare, science etc.).

Moreover, the work recently conducted by the authors leads to an integration of SVT with other classes of information systems, such as: ERP/BI, Business Process Modeling Systems, Customer Relationship Management applications, and Internet search engines. The next step in developing the SVT is to enhance the visualization capabilities by integration with SIS technology, which is meant to enrich visualizations by new elements and make it more automated and user-friendly compared to the Excel spreadsheets that were used in the recent prototype.

It seems that the proposed solutions can be the source of well-arranged managerial information extensively facilitating the process of decision-making at every level of a business organization.

References

- Card S.K., Information visualization, [in:] J. Jacko, A. Sears (Eds.), *The Human-Computer Interaction Handbook*, Lawrence Erlbaum Associates, 2003.
- Chen C., *Information Visualization: Beyond the Horizon*, 2nd edition, Springer, 2004.
- Golbeck J., Trust on the World Wide Web: A survey, *Foundations and Trends in Web Science* 2006, Vol. 1, No. 2, pp. 131-197.

- Gomez-Perez A., Corcho O., Fernandez-Lopez M., *Ontological Engineering with Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web*, Springer-Verlag, London 2004.
- Google Developers, Custom Search, 2012, <https://developers.google.com/custom-search/v1/overview> (date of access: February 2012)
- Green J., Bullen S., Bovey R., Alexander M., *Excel 2007 VBA Programmer's Reference*, Wiley Publishing, Indianapolis 2007.
- Hitzler P., Krötzsch M., Rudolph S., *Foundations of Semantic Web Technologies*, Chapman & Hall CRC, 2009.
- Litwin L., Myrda G., *Systemy informacji geograficznej. Zarządzanie danymi przestrzennymi w GIS, SIP, SIT, LIS*, Helion, Gliwice 2005.
- Miner G.D., Delen D., Elder J., Fast A., Hill T., Nisbet R.A. (Eds.), *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*, Elsevier, New York 2012.

NARZĘDZIA PRZESTRZENNEJ WIZUALIZACJI W SYSTEMACH INFORMATYCZNYCH BIZNESU

Streszczenie: Artykuł prezentuje prototypową aplikację o nazwie Spatial Visualization Tool (SVT), która może być użyta do wizualizowania w postaci interaktywnej mapy danych pochodzących z różnych rodzajów systemów informatycznych i zewnętrznych źródeł, takich jak Internet i bazy dokumentów. Koncepcja artykułu jest oparta na założeniu, że dane w systemach informatycznych mogą być wizualizowane w sposób analogiczny do map geograficznych. Taka wizualizacja może być bardzo czytelna i służyć jako wsparcie procesów decyzyjnych opierających się na dużych ilościach danych. Pierwsza część artykułu prezentuje wybrane aspekty metodologii geograficznych systemów informacyjnych (GIS). Druga część przedstawia teoretyczny algorytm pozyskiwania danych i dokonywania ich wizualizacji przestrzennej. Ostatnia część omawia praktyczny przykład działania aplikacji SVT w rekrutacji kandydatów do pracy. Prototypowe narzędzie zostało skonstruowane przy użyciu pakietu MS Excel i języka Visual Basic for Applications (VBA). Zaprezentowano także koncepcję rozszerzenia funkcjonalności SVT przy użyciu technologii sieci semantycznej w celu rozpoznawania źródeł i typów danych w celu automatyzacji konstruowania wizualizacji.

Słowa kluczowe: wizualizacja danych, przestrzenne systemy informacyjne, systemy informacyjne biznesu.