IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4839

Visual Analysis of Human Health Data

P. Schulze-Wollgast, H. Schumann, and C. Tominski
University of Rostock, Computer Science Department
Albert-Einstein-Str.21, D-18057 Rostock
Tel.: +49 381 498 3418, Fax: +49 381 498 3426
Email: {psw, schumann, ct}@informatik.uni-rostock.de

ABSTRACT

To ensure a high level of human health data protection, the huge amount of data gathered day by day by the physicians of the health service has to be analyzed in an efficient way. The current statistical and numerical methods allow only a limited analysis of these data. In this paper we discuss visualization as an efficient approach for analysis and exploration of time and space-dependent human health data. Furthermore, we describe a visualization system which has been developed for visualizing health data of the German state Mecklenburg-Vorpommern. This system is part of a tele-consultation-system for monitoring of health situation.

1. INTRODUCTION

Day by day the physicians of health service collect a huge amount of medical data describing the diseases of patients. An effective analysis of these data is necessary to ensure a high level of human health protection. However, the current information flow in Germany is too slow. Therefore, a better monitoring and a faster data analysis are necessary. One approach to reach this aim is the application of visualization.

During the last years the visualization of data became a commonly accepted and widely used approach for the extraction of relevant information from huge abstract data sets. In many cases, a better insight into complex processes and phenomena can be gained by means of graphical representation. The human brain allows the simultaneous perception of a large amount of optical information. Thus, a lot of information can be encoded in one single picture. Therefore, visual analysis offers an effective way for the extraction of information from huge data sets. For doing so, abstract data sets have to transformed into graphical representations to allow a better analysis, understanding and communication of data, models, and concepts in science and engineering.

A visual representation of abstract data has to be expressive, effective, and appropriate [SM00]. The expressiveness relates to the requirement, that all relevant attributes and only those must be expressed by the visualization. The effectiveness depends on the ability of the observer to interpret the presented facts correctly and efficiently. Here, the visual abilities of the observer play a major role. A visualization is appropriate if benefit and expense are balanced.

This paper focuses on the visualization of human health data collected in the northeast of Germany. This visualization is embedded in a tele-consultation-system. The paper is organized as follows: First we give a short overview of the architecture of the tele-consultation-system called Tecomed (see section 2). Then, we describe the visualization of spatial and temporal dependencies of the health data and some special problems in the field, like icon positioning and color-coding (see section 3). Finally, we present the visualization system as part of the tele-consultation-system Tecomed (see section 4) and conclude with some remarks on future works.

Finally we present the visualization system as part of the tele-consultation-system TeCoMed (see section 4) and conclude with some remarks on future works.

2. TECOMED

Tecomed is a tele-consultation system for visual analyzing health data. Figure 1 shows the architecture of this system. The *Data Base Management System* (DBMS) mainly contains patient data recorded by the health insurance AOK (one of the compulsory sickness insurance in Germany). These data cover about 20% of the population of Mecklenburg-Vorpommern and describe every medical case in detail with specific information, e.g. age, sex, duration of illness, and kind of treatment. Every day about 1.500 data records are saved. These data have to be anonymized, e.g. for Tecomed personal data must be disconnected from health data to protect the privacy of the patients. Furthermore, data access has to be controlled and must only be granted to authorized users. This requires storing of special data about an user in a database to ensure data security.

Tecomed uses a client-server architecture to allow access via the Internet. The *client* is operated by a user to select data and to get a visual representation of these data. The *server* provides the *clients* with a lot of data-analyzing and data-service processes. The server also performs access control and session management.

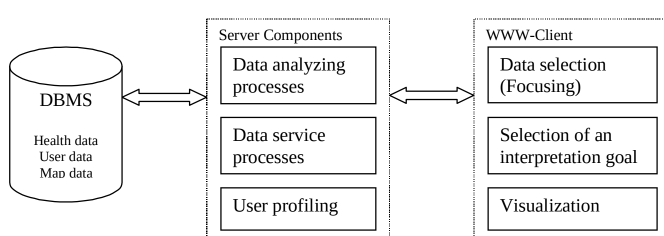
The *visualization system* of Tecomed represents the data within their spatial dependencies for different levels of granularity. Moreover, temporal dependencies are also visualized in an expressive manner. This approach allows an easy assignment of relevant data features to their spatial and temporal context.

Often it is impossible to visualize the whole data set in one single picture due to the large amount of data. In many cases it is even not necessary to display the entire data set because most of the users are interested in some specific parts of the data only. Therefore, the amount of data to be displayed can be reduced by *focusing*. This is either done by *projection* or by *selection*. A *projection* reduces the amount of information by decreasing the number of parameters. A *selection* reduces the amount of information by decreasing the number of data records. In Tecomed *focusing* can be done with regard to three different main criteria:

- *Focusing* based on health data: In this context *projection* reduces the number of parameters describing the diseases or the patients. Furthermore, a *selection* of diseases with regard to their parameter values is possible.
- *Focusing* based on geographical space: Here, levels of granularity are chosen or geographical areas are selected.
- *Focusing* based on time steps: This means the selection of a number of times steps or time intervals.

Focusing has to be done according to the aims of a user to ensure that all desired information are obtained. Whether to start focusing with regard to health data, geographical space, or time steps depends mainly on the application domain and the goals of a user.

Figure 1. General architecture of the tele-consultation system Tecomed for visualizing and monitoring of health data.



Tecomed is designed for a lot of different kinds of users, e.g. physicians, apothecaries, health insurance funds, and tourists. They differ from their medical knowledge, and a lot of them have only little experience in visualization. They are all able to specify in what facts they are interested in, but they are not able to specify and select an appropriate visualization technique. Therefore, we specified several user goals for different user groups as verbal given questions. Selecting a question is easy and intuitive. This increases the usability of the visualization system.

3. VISUALIZATION OF HEALTH DATA

Health data are multi-parameter data with temporal and spatial dependencies. The success of the visualization mainly depends on the fact whether the user can detect these relationships between the various parameters from the graphical representation and relate them to the underlying context. Therefore, an effective visual representation has to encode the data values in this way.

Moreover the following demands have to be satisfied:

- Intuitive perception, i.e. graphical representation has to be intuitive interpretable for different user groups with little experience in computer graphics
- No artifacts, i.e. incorrect interpretations of the graphical representation must be avoided. False conclusions can be very dangerous especially in the context of medical applications. Thus, falsification and interpolation of the data are not allowed.

3.1 Visualization of geographical dependencies

A representation of the spatial frame of reference is necessary to assign important data features to their geographical context. Therefore, a main requirement is the adequate display of the different spatial granularities. Maps are very suitable for the display of geographic spatial dependence because they have a low level of abstraction, which allows an intuitive perception of spatial relations.

Because of the huge amount of information an effective visual representation of multi-dimensional data sets with spatial dependencies is difficult to find. Although a lot of techniques have been developed in the last few years, the display of complex spatial relationships still remains problematic, especially if we have to consider different granularities.

In Germany states are divided into smaller parts like countries for administrative reasons. Countries can be divided into municipalities. This builds up a hierarchy. To visualize health data in their spatial context we developed a map file format for such hierarchical structures.

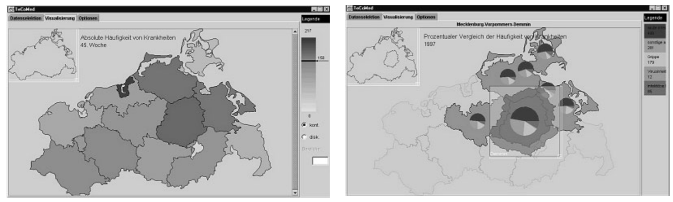
The format fulfills the following demands:

- Saving a hierarchical structure.
- Enforcing the consistency of the map in every hierarchical level.
- Enabling the calculation of neighborhoods.

To meet these criteria we used a topological structure (see [KO96]). This enables us to build up the map file by first defining topological entities and using them later by referencing them. The result is a set of areas. These areas have to be related to each other according to the hierarchical structure. This is done by defining parent-child-relations. Using references avoids the occurrence of redundancy. This leads to a consistent map for each hierarchical level. Furthermore, the topological decomposition makes the required calculation of neighborhoods possible. The map file for visualizing our health data consists of the German country Mecklenburg-Vorpommern divided into smaller districts, which again are divided into zip code regions.

With the defined map file we can visualize maps in different levels of granularity. However, when the amount of health data is large it is not possible to present all information on the map. In this context we have to deal with two demands. For easy analysis we first need a detailed view of special parts of the data a user is interested in. Secondly, we need a view of the whole data set to avoid that a user looses general overview. Two concepts are known meeting both demands:

Figure 2. Left: Color-coding of the number of cases of one single disease in the counties of Mecklenburg-Vorpommern.



Right: Visualizing the number of cases of several (five) diseases with pie chart icons for only some counties of the German state Mecklenburg-Vorpommern.

- Overview and detail
  A detailed view and an overview of the data are provided in different pictures (see [CMS99]). By switching the views the user is able to analyze the data in detail without loosing overview, but he has to combine both views mentally. Furthermore, overlapping between detailed view and overview can occur and information can be lost. By using transparency this problem can be solved.
- Focus and context
  Focus and context techniques provide a detailed view and an overview in one picture (see [Kea98]). This often is realized by enlarging a focus region, while the rest of the picture (the context) is scaled down. Scaling down the context means loosing some information in that region, but a general overview is always possible. Examples for *focus and context* techniques are *Graphical FishEye views* [SB94] or the *Cartographic Lens* [Ras97].

In our visualization system Tecomed (see section 4) we included both concepts. By doing so we join flexibility and expressiveness of *focus and context* with user acceptance of *overview and detail*.

3.2 Visualization of health data with traditional techniques

Several visualization techniques are suitable for the graphical representation of health data. Since Tecomed is designed for a great number of different users some traditional and easy to understand visualization techniques (e.g. color coding, pie chart icons) are included in our system. For visualizing only one disease simple color-coding of the areas on the map is used. Pie chart icons are applied for visualizing different parameters for several diseases at a fixed time. Furthermore, animation of color-coded maps is used to visualize changes over time. The animation can be controlled interactively.

All visualization techniques are connected with questions a possible user may be interested in (see section 2). Moreover, the parameterization of these techniques is fixed to avoid unexpressive visualizations.

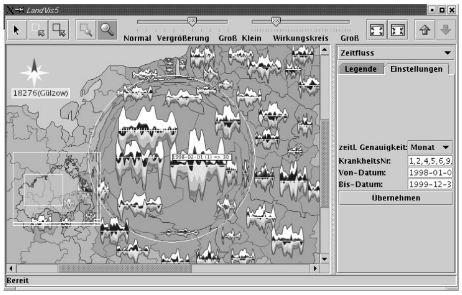
3.3 Visualization of the temporal dependencies

For visualization of health data we need techniques making it possible for a user to identify significant temporal changes. Furthermore, revealing periodicity of a disease or detecting significant changes in the number of cases of a disease are important goals for visualization.

Many visualization techniques for representing time series are known. However, they do not consider spatial dependencies as well. For the visualization of health data we need a combination of both. Therefore, we introduce a new concept: *iconification and positioning*. We use known visualization techniques to produce special types of icons and call this process *iconification*. It is important to consider the expressiveness of the created icons. In optimal case the icons keep the expressiveness of the original technique. As a minimal demand we require a general overview of the encoded data. For icons of the latter case we need special visualization concepts like *overview and detail* or *focus and context* to make a detailed analysis of the data possible.

For visualizing temporal dependencies of our health data we examined several known time visualization techniques. Among others we

Figure 3. ThemeRiver *icons positioned on a map for visualizing temporal changes with a* focus and context *display.*



found *ThemeRiver* [HHN00] and *spiral displays* [WAM01] suitable for our approach. For the visualization system Tecomed we implemented *ThemeRiver* icons (see figure 3). Spiral icons were implemented too but not yet integrated into Tecomed. For visualizing temporal changes on a map we create an icon for each area. Then, each icon is positioned on the map according to a corresponding point of reference (see section 3.4).

3.4 Special problems

In this section we want to discuss two problems which often occur when visualizing multidimensional data sets on maps: Arrangement of icons and color coding.

Positioning of icons

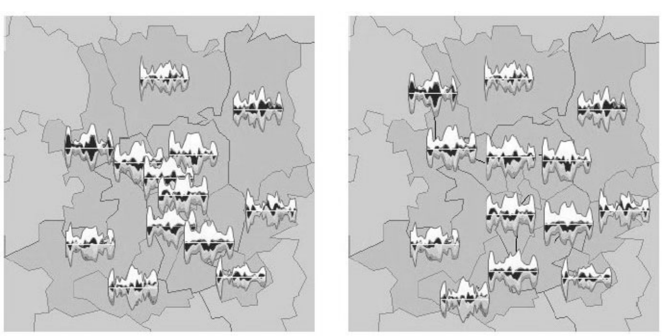
Drawing icons on maps is a suitable way to visualize data. Icons exactly can be positioned on maps and doing so spatial relations can be presented. However, this approach is problematic, especially when visualized data are large or the related regions are very small. When placing a large number of icons on a map using a simple arrangement, overlapping of icons can hardly be avoided. This leads to an unacceptable loss of information.

The problem of placing icons on a map is strongly related to the cartography problem of label placement. For this reason we use concepts of computer cartography to find more expressive points of reference for the icons. Starting with a given point of reference for each region on which icons have to be placed, we move this point until some requirements are fulfilled. To model the requirements we use a cost function, which has to be minimized (Figure 4).

Automatic color coding

If different levels of granularity of the maps are given, usually color-coding is used by automatic scaling the colors according to the ranges of data values to be presented. This leads to different color scales,

Figure 4. A simple arrangement of ThemeRiver *icons over regions of a map (left picture), and the positioning of these icons after a replacement (right picture)*



and therefore to a conflict if comparing the values of different geographical regions.

Ensuring the expressiveness of visual representations for comparing purposes requires strictly applying the same color scales for all granularity degrees. This results in uniform color coded pictures for low levels of granularity.

To solve this problem we propose the following approaches:

- The absolute values have to be normalized over the amount of insured persons.
- The values can be classified according to the specified questions (see section 2), e.g. if a question requires the comparison of special situations (e.g. high number of influenza, normal or little occurrence) we need 3 colors for all levels only. In doing so, the used color scales are comparable. In this case a uniform color-coding contains the relevant information.

4. THE VISUALIZATION SYSTEM OF TECOMED

The tele-consultation system Tecomed was developed for visualizing human health data for the German state Mecklenburg-Vorpommern via the Internet. Concepts for visualization of huge data sets according to their spatial and temporal dependencies within different levels of granularity were implemented.

Tecomed provides several interaction functions to support focusing. The diseases and their parameters can be interactively selected from simple lists or buttons. For the specification of time steps and intervals a user can choose between different temporal resolutions, e.g. day, week, month, quarter and year from a hierarchical list. Geographical areas can also be selected from a hierarchical list or by clicking on the map.

Several visualization techniques have been implemented for the graphical representation of the health data. We integrated some traditional visualization techniques. Furthermore, we used iconification to create new icons from known techniques for visualizing temporal dependencies.

Furthermore, we have realized several interaction functions to explore the data. One of them is a *semantic zoom*. The hierarchy contained in our map is used to provide information of different levels of granularity. The user can switch between the levels interactively. By doing so one can change the density of information displayed on the map. Switching from lower to higher granularity increases the density of information. This kind of interaction is called *semantic zoom*. In contrast to the normal graphical zoom *semantic zoom* does not change the graphical data but the density of information. *Semantic zoom* makes it easy to get a first general expression of the health data and when interesting parts are found to zoom in.

The visualization system of Tecomed has been designed in a modular way so that more visualization techniques can be integrated after analyzing more user groups with their specific visualization goals.

5. CONCLUSIONS

Visualization of data is now a commonly accepted and frequently used approach for extracting relevant information from a huge amount of data and thus gaining a better understanding and insight into the underlying context.

We proposed a system for visualizing health data in their spatial and temporal context within different levels of granularity. We integrated *overview and detail* and *focus and context* to make visualizing of the large amount of health data on the map possible. *Iconification and positioning* allows us to use known techniques for visualizing of temporal changes on a map.

The possibility of specifying a users goal by selecting from a list of *verbal questions* increases usability and avoids unexpressive visualizations.

Further work will concentrate on specifying more qualitative questions corresponding to the addressed user groups. Examining and integrating more suitable visualization techniques is planned for the future as well.

ACKNOWLEDGMENT

The authors wish to thank Georg Fuchs for implementation work and providing some of the pictures.

LITERATURE

[CMS99] S. K. Card, J. D. Mackinlay, and B. Shneiderman (eds.): *Readings in Information Visualization – Using Vision to Think*. Morgan Kaufman, San Francisco, 1999.

[HHN00] Susan Havre, Beth Hetzler, and Lucy Nowell: *ThemeRiver: Visualizing Theme Changes over Time*. In: Proceedings of IEEE Symposium on Information Visualization 2000, p. 115-123, IEEE Computer Society, Los Alamitos, 2000.

[Kea98] T. A. Keahey: *The Generalized Detail-In-Context Problem*. In: Proceedings of Vis'98, IEEE Visualization, Research Triangle Park NC, 1998.

[KO96] Menno-Jan Kraak, and Ferjan Ormeling: *Cartography: Visualization of Spatial Data*. Longman Singapore Publishers, Singapore, 1996

[Ras97] Wolf-Dieter Rase: *Fischauge-Projektionen als kartografische Lupen*. In: Salzburger Geographische Materialien IX = Angewandte Geographische Informationsverarbeitung (Heft 26), Selbstverlag des Instituts für Geographie der Universität Salzburg, 1997.

[SB94] M. Sarkar, and M. Brown: *Graphical fisheye views*. Communications of the ACM, 37(12):73-84, 1994.

[SM00] Heidrun Schuman, and Wolfgang Müller: *Visualisierung – Grundlagen und allgemeine Methoden*. Springer-Verlag, Berlin, 2000.

[WAM01] Marc Weber, Marc Alexa, and Wolfgang Müller: *Visualizing Time-Series on Spirals*. In: Proceedings of the IEEE Symposium on Information Visualization 2001, p. 7-14, IEEE Computer Society, Los Alamitos, 2001.