

# Meta Data for Visual Data Mining

Thomas Nocke  
Institute of Computer Graphics  
University of Rostock  
Germany  
nocke@informatik.uni-rostock.de

Heidrun Schumann  
Institute of Computer Graphics  
University of Rostock  
Germany  
schumann@informatik.uni-rostock.de

## ABSTRACT

Increasing amounts of large data and information sets require new analysis techniques. The domain of data mining investigates new paradigms and methods adapted for scalability, flexibility and problem abstraction for large data sets. In particular the field of visual data mining offers valuable methods for analyzing large amount of data intuitively. Visual data mining combines several visual and non-visual methods to reveal patterns, coherences and other features of data sets. The application of meta data can support the selection of suitable mining methods as well as of appropriate parameter values to control these methods. This paper defines a variety of meta data for visual data mining purposes, for instance by specifying ranges of values, cluster structures and regions of interest. These meta data can be applied for a more efficient visualization of the data set. We introduce a framework for effective extraction of these meta data.

## KEY WORDS

Visual Data Mining, Meta Data, Visualization

## 1 Introduction and motivation

The analysis of complex heterogenous data requires sophisticated exploration methods. Complex data mining processes which apply many different analysis techniques can benefit especially from visual data processing and new visualization paradigms. Additionally, visualization provides a natural method of integrating multiple data sets and has been proven to be reliable and effective across a number of application domains. Visual methods cannot replace analytic non-visual mining algorithms. Rather it is useful to combine multiple methods during data exploration processes [1], [2] [3].

Visual data mining focuses on this combination of visual and non-visual techniques as well as on integrating the user in the exploration process. Ankerst [4] classifies current visual data mining approaches into three categories. Methods in the first group apply visualization techniques independent of data mining algorithms. The second group uses visualization in order to represent patterns and results from mining algorithms graphically. The third category tightly integrates mining and visualization algorithms in such a way that intermediate steps of the mining algorithms

can be visualized. Furthermore, this tight integration allows users to control the mining process directly based on the given visual feedback.

Special features and structures which have to be calculated depend on the properties of the data set. These properties can be described by meta data. Meta data can support the selection of suitable mining methods as well as of appropriate parameter values to control these methods. For instance the applicability of a specific similarity measure depends on the basic scale types of the variables of a data set.

Applying meta data for visualization purposes is not a new concept and has been utilized for a long time (cf. e.g. [5], [6] and [7]). In general meta data can be classified into *DESCRIPTIVE*, *DERIVED* and *HISTORICAL* meta data (cf. [8]). *DESCRIPTIVE* meta data specify underlying properties of a data set which define the access and the storage structure of the data. Meta data – which have been extracted from a data set (in any form) – are called *DERIVED* meta data. *HISTORICAL* meta data provide important information about the origin of a data set. This might be errors and uncertainties introduced due to operations which have been applied to the data previously. Right until now mainly *DESCRIPTIVE* meta data have been applied in visualization context. Nevertheless, some methods for gathering important properties of a data set for a compact description and visualization are known from the field of feature visualization (e.g. [9], [10]). The intention of this is to visualize relevant features instead of the the original data set. These relevant features can be considered as *DERIVED* meta data.

Most of the known meta data specifications are described in a formal way. An important question for a practical system is how to handle and to store the meta data. Prevalent formats used in current visualization systems (e.g. NetCDF or Data Explorer file format) store both data and meta data in the same file. These formats predominantly store *DESCRIPTIVE* meta data, and some of them can be extended. Moreover, practical problems arise due to the fact that most of the stored data do not include meta data descriptions for visualization purposes. That's why new methods for gathering and storing meta data are needed for flexible and practicable solutions. The aim of this paper is not to define a new meta data file format, but to define a set

of relevant meta data for visual data mining problems and to describe ways to get them.

Let's describe a special problem to make this clear. A data set as output of a climate model should be explored to get a better understanding of climate phenomena [11]. Such data sets consist of different 1D, 2D and 3D data on structured grids that describe dynamics for oceans, atmosphere, ice and others. The data sets are defined in time series with differing granularities. They consist of 3 to 40 attributes for a data record. The problem is to link partial models in respect to different scales in space and time. This can be supported by using meta data describing the scales and definition areas for each partial model according to a global coordinate system. In doing so, heterogeneous data can be combined and displayed to gain an overview of the whole data set as well as to find out coherences between attributes from different partial models.

This example demonstrates, that a more comprehensive and systematic treatment of meta data would be desirable. In this paper we describe our general framework, which provides a variety of meta data for exploring and visualizing large data sets. In our framework the process of exploring data can be refined interactively by applying a sequence of steps. In a first representation certain properties of a data set can be presented to the user. Those properties and the according knowledge can be refined in further steps by adaptive gathering of meta data and by adapting exploration tasks. In this way gradually more knowledge about the data will be generated.

The paper is organized as follow: First we describe a specification of meta data, which includes a great variety of aspects (2). Furthermore, we introduce our concepts for gathering meta data (3), followed by a short description of our framework (4). After that we show some methods for the visualization of meta data (5) and conclude with a short discussion for future work (6).

## 2 Specification of meta data

In many application fields a variety of different structured data sets and a variety of exploration tasks can be applied to support data analysis. A tight linkage of visualization modules is often too inflexible to support data and/or user specific exploration. Our goal is to gather as much as possible additional information about the data to support flexible visual data mining.

For specification of meta data three major classes of questions have been investigated. These are:

1. **questions referring to description of kind, storage and content of the data** (e.g. What storage structure does the data have and what can be derived from this structure? What quality does the data have and how reliable and concise can analysis results be?),
2. **questions referring to the 3 main steps of the visualization pipeline** to support the visualization process

(Which meta data can support which filtering, mapping and rendering processes?) and

3. **questions referring to the application of different non-visual mining techniques for gathering meta data** (e.g. What techniques produce results relevant for visualization? Which techniques require meta data as input to produce useful results? Do the techniques produce errors and how can the reliability of their results be evaluated?).

In respect of these questions, a variety of meta data can be derived. To manage this variety we use the following concepts for systematic specification of meta data:

- **Separation:** separating meta data which are usable for all kinds of data from meta data which are only relevant for specific data classes
- **Integration:** integrating all kinds of meta data (DESCRIPTIVE, DERIVED and HISTORICAL meta data)
- **Structuring:** applying meta data for the description of the internal structure of the data set
- **Validity:** integrating concepts for handling errors, uncertainties and quality
  - of the data
  - of the analysis techniques and the DERIVED meta data
- **Extensibility:** offering the opportunity for extension of further meta data, for instance:
  - meta data for coupled data sets
  - meta data for new data classes
- **User orientation:** integrating meta data for different kinds of user interaction
  - integrating meta data both dependent and independent of special user profiles and special exploration tasks
  - using user knowledge

For the specification of DESCRIPTIVE meta data we follow the principles of the NetCDF format [12]. In addition we specified a great variety of DERIVED meta data. Furthermore, we introduced a hierarchy of dependencies of meta data to support effective gathering thereof. Let us explain this in more detail:

A data set consists of values for different variables. Thus we need meta data to describe these variables, to describe the properties of the whole data set as well as to describe special characteristics of special data classes. Consequently, we subdivided meta data into meta data for the variables, meta data for the whole data set and meta data for data classes (see fig. 1).

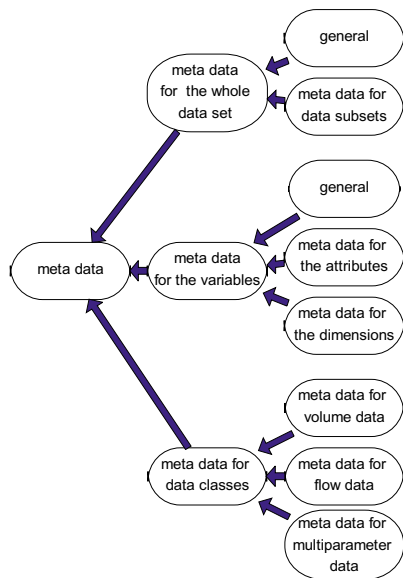


Figure 1. Hierarchy of meta data

**Meta data for the variables** specify the characteristics of the variables in general as well as the special characteristics of the specific variables. Because of some differences between dependent and independent variables we want to distinguish between meta data for dimensions (independent variables) and meta data for attributes (dependent variables).

The **meta data for the whole data set** specify the properties of the entire data set. We distinguish between general meta data, which refer to general properties of the whole data set (e.g. the common relationships of the variables), and meta data for relevant subsets, which specify subsets of interest and their related properties.

Furthermore, we introduce **meta data for data classes**, which describe specific characteristics of specific classes of data. We specified meta data for volume data, for flow data and for multi-parameter data.

In the following some of the meta data that are realized in our framework (cf. [13]) will be presented in a compact form.

## Meta data for the whole data set

### General meta data for the data set:

- the number of variables<sup>◦</sup>
- the number of data records<sup>◦</sup>
- structural information of the variables<sup>\*</sup> containing
  - tuple and key information
  - correlations and joint information content
  - hierarchies of variables<sup>\*</sup> to specify dependencies (e.g. month and year) or to model grouping of similar variables by using factor analysis methods
- the average information content of the data set<sup>\*</sup>

- the quality of the whole data set<sup>\*▷</sup>, for instance missing values

### Meta data for relevant subsets:

- the number and kind of subsets of interest<sup>◦\*</sup>
- meta data defining subsets containing
  - number of dimensions<sup>◦</sup>
  - number of related attributes<sup>◦</sup>
  - data records<sup>◦</sup>
  - specific properties of subsets<sup>\*</sup>

## Meta data for the variables

### General meta data for the variables:

- the name of the variable<sup>◦</sup>
- the scale type<sup>◦</sup>
- classification of dependent (attributes) and independent (dimensions) variables<sup>◦</sup>
- further semantic information<sup>◦</sup>
- the information content<sup>\*</sup>, e.g. based on Shannon's entropy
- the quality of variables<sup>\*▷</sup> (number of missing values and the mean quality information on the variables values)

### Meta data for the attributes:

- the data type<sup>◦</sup>
- histogram properties<sup>\*</sup>, e.g. distribution of values, minimum and maximum, variance, mean values

### Meta data for the dimensions:

- the kind of the space defined by the dimensions<sup>◦</sup> (e.g. cartesian space with space and/or time dimensions or an abstract space combining dimensions of any kind).
- the dimensionality of the space<sup>◦</sup>
- description of the spatial and/or time dimensions<sup>◦</sup>, including the specification of:
  - domain dimensionality and shape of underlying grid<sup>◦</sup>
  - nature of range<sup>◦</sup> (whether the data are defined point-by-point or over regions of the domain)
- segmentations of the space<sup>\*</sup> according to special features:
  - distribution of data points and attribute values
  - heterogeneity of space areas
  - quality of values<sup>\*▷</sup>
  - relevance, interest (according to the properties specified above)

Legend: ◦ DESCRIPTIVE \* DERIVED ▷ HISTORICAL meta data

## Meta data for data classes

### Meta data for flow data:

- time dependency<sup>◦</sup>
- dimensionality of the flow vectors<sup>◦</sup>
- topological properties of the flow field<sup>\*</sup> (kind and location of critical points and separating regions)
- further general field properties<sup>\*</sup> (e.g. curvature, divergence, circulations, wave fronts, ...)

#### Meta data for volume data:

- volume data specific segmentations\* of the space according to
  - the functional behavior of the underlying scalar function
  - the topology, shape and number of inner bodies according to relevant thresholds
- properties of the field of gradients\*

#### Meta data for multi-parameter data:

- special data records\* such as
  - outliers
  - data records with equal attribute values
  - typical data records
- the structure of the data records\* according to
  - the distribution and size of clusters
  - the properties of clusters according to the attribute values

Legend: ◊ DESCRIPTIVE \* DERIVED ▷ HISTORICAL meta data

The **flexibility** of the specification allows extensions at any level, for instance the inclusion of special characteristic of GIS<sup>1</sup>- or of scattered data. We implicitly included some of their characteristics by specifying meta data for the dimensions and for the whole data set.

Up to now, many of the introduced meta data have been considered separately. Our aim is to combine several approaches from different fields (e.g. visualization, data base technology, statistics).

For example we use a **key analysis** technique for variables of a data set with unknown types of these variables (e.g. this can be an ascii table without further knowledge about it). The result is a set of minimal keys<sup>2</sup>. By taking the shortest key(s) the classification of dependent and independent variables can be achieved. For instance we applied this method to a medical data set describing patient and illness data. The algorithm identified inconsistencies in this data set. No key could be found, forcing us to investigate more deeply and leading us to discover that some patients had been entered twice into the data set. After removing this inconsistency the key analysis algorithm detected a single variable as minimal key. This variable was defined as dimension and could be used for suitable parameterization of visualization techniques.

Another interesting point is **space segmentation** on the basis of properties of space areas and the extraction of **regions of interest** in conjunction with these properties. This relates to several works in the domain of feature visualization [9], [10], where special data features are extracted, focused and displayed. For instance this can be

<sup>1</sup>Geographic information systems

<sup>2</sup>Keys are combinations of variables which tuples allow an unequivocal mapping to each data record.

done for **flow data** by calculating critical points and separating planes between these critical points. These topological characteristics are of major importance for the analysis of flow data and can be used to separate interesting and uninteresting regions. For example we investigated the electrostatic field of a water molecule. A conventional streamline visualization shows an untypical region in the middle of the data set with a supposed accumulation of critical points. In fact we cannot recognize how many critical points occur (cf. fig. 2).

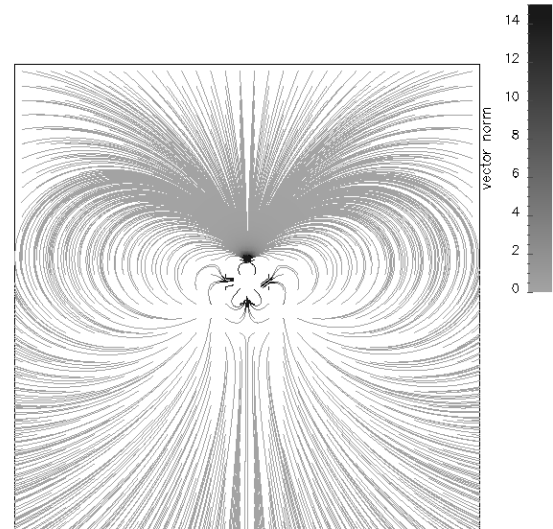


Figure 2. Streamline visualization of the electrostatic field of a water molecule (vector norms are mapped to colors of streamlines) measured by the "Konrad - Zuse" institute of Berlin: overview visualization

However, meta data extraction of critical points gave us the unexpected number of 12 critical points<sup>3</sup>. It was used to extract interesting regions and visualize them in more detail (see fig. 3).

Another example for applying meta data is the reduction of the data set by combining user's exploration tasks, the user's knowledge and automatic cluster algorithms [14]. This combination allows a flexible, easy-to-use **extraction of structures and of interesting subsets**. Beyond the pure cluster data (e.g. an n-tree for hierarchical clustered data), meta data describing outliers, equal data sets and properties of the single clusters according to the attribute values can be specified as well. This information can be applied in different ways for visualization problems. For example outliers can be marked or highlighted visually. Another example is the use of tree distributions for visual compression of complex sub-clusters (cf. [15]).

<sup>3</sup>A possibility would be the direct display of critical points. This may not be sufficient, because in general the topological information is not unique. Critical points can be applied to flow visualization techniques for defining regions of interest. Furthermore, critical points can be displayed in combination with such standard techniques.

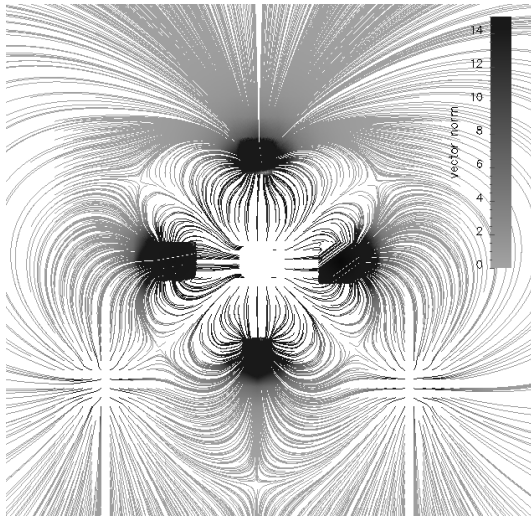


Figure 3. The same visualization of the electrostatic field of a water molecule: visualization of the region of interest in the middle of the vector field, applying a higher number of streamline starting points for that area

Moreover, we define meta data, which allow **gathering of further meta data** (e.g. we use semantic information like strings for an automatic extraction of scale types).

### 3 Gathering of meta data

The process of gathering meta data is not trivial. Determination of all useful meta data can be expensive, so it is absolutely necessary to find an effective way to gather meta data.

This can be achieved by the following strategies:

- predefinition of a **suitable order** of gathering processes, by avoiding repeated calculations and by utilizing processes which supplement each other
- **dividing** the gathering processes into automatic, semi-automatic calculations and user interactions with different levels of user support. Depending on the user profile the interaction level can vary from a high degree of interaction and parameter input to a high degree of using default values.
- using **time frames** for meta data extraction (e.g. fast analysis, ..., no time limits for analysis) for appropriate parameterizations of slow processes and/or for avoiding long-drawn-out extractions

Figure 4 shows the main sequence in the process of gathering meta data. After reading the data set different gathering processes extract different types of meta data. First the general meta data for variables and the specific meta data for attributes are determined. This is of special importance for the following processes, because especially the scale type and the data type of the variables/attributes

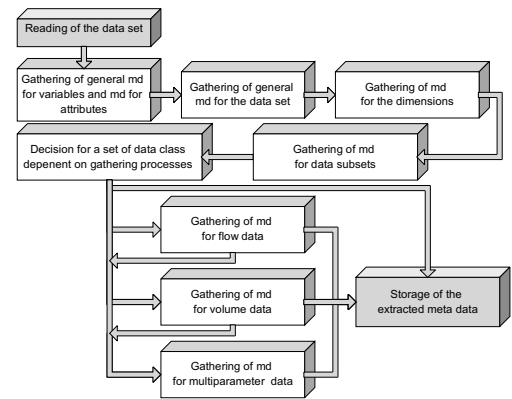


Figure 4. Process sequence of gathering meta data (short md); main sequence

are relevant for later reading and using data values. Another example for reusing meta data is the extraction of histograms to speed up following steps. Next, step is the extraction of general meta data for the data set. This includes statistic calculations to find correlations or the classification of dependent and independent variables. Next extraction methods obtain meta data for the dimensions. Meta data for special data subsets and meta data for the specific data classes are the least to be gathered.

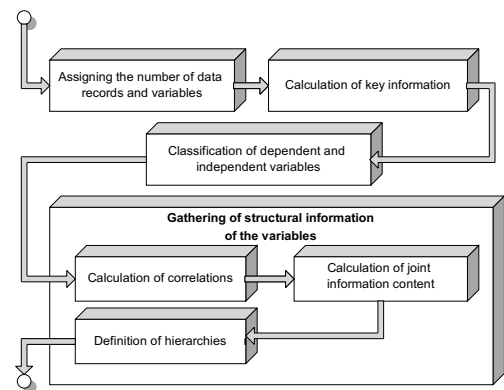


Figure 5. Process sequence of gathering meta data (short md); sequence for gathering the general meta data for the whole data set only

Figure 5 shows the gathering process of the general meta data for the whole data set. We wish to illustrate this using a Baltic Sea data set:

- We assume that discrete and continuous variables had been classified, information content and cardinalities calculated. The variables with the highest information content (i.c.) and the highest cardinality (card.) of ranges of values were temperature (i.c.: 0.772418; card.: 1438) and salinity (i.c.: 0.703109; card.: 1274).

A further result of this extraction was that the values 99 and 999 are frequent outliers in different variables. They were classified as missing values.

- The acquisition of numbers of variables and of data records results: the data set contains 9097 data records and 17 variables. The next step – the analysis of key information – extracted a relatively long minimal key of 9 variables (id of station of measurement, longitude in minutes, longitude in degrees, latitude in minutes, latitude in degrees, depth, year, month and day). This analysis offers an initial idea of possible simplifications, which can be described by subsequent gathering of variable hierarchies. In the next step these key variables are classified as dimensions (independent) and the others as attributes (dependent).
- To analyze the structure of the variables correlations between combinations of variables were calculated. Results showed that most of the analyzed pairs had a correlation near to 0. But there were also some pairs with a high correlation value (-0.888622, 0.717924, 0.964653). The subsequent extraction of joint information content used a threshold to speed up the calculation. Altogether we integrated 13 variables in 28 pairs and 6 triples with relevant information content.
- Finally a variable hierarchy was specified interactively (e.g. dependencies between the dimensions year, month and day, between longitude in minutes and longitude in degrees as well as between latitude in minutes, latitude in degree).

#### 4 A framework for gathering meta data

A flexible system has been developed to integrate the concepts described. Our framework has the following functionality:

- Management of user profiles
  - Interaction level (interactive, semi-automatic and automatic analysis)
  - Time frame (fast analysis, no time limits for analysis)
- Integration of a flexible data management
  - Reading data from files and databases
  - Reading and writing meta data files: meta data file format (binary or ascii)
- Integration of various techniques for gathering meta data
- Integration of display techniques showing the state of the current meta data acquisition
- Inclusion of textual and visual representations of meta data and a context-sensitive help system.

A strict separation of control, interaction and gathering processes allows the system to be portable. Furthermore, concepts for handling and analyzing of coupled data sets have been developed and integrated into our framework.

Figure 6 shows a dialog for the interactive definition of meta data for variables supported by default values.

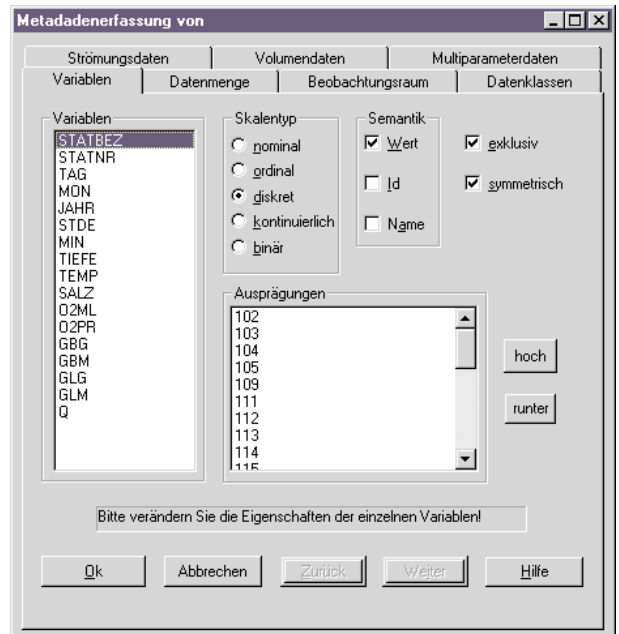


Figure 6. Screenshot from the framework: interactive definition of meta data for variables

The names of variables are displayed on the left. For each variable a default scale type has been defined automatically. This scale type can be binary, nominal, interval-scaled and ratio-scaled. Furthermore, variable's scale types can be declared as multi-valued. Binary variables can be defined as symmetric or asymmetric. The window at the bottom lists the parameter values of each variable. If the variable's scale type is ordinal the definition of an order between the variable's values is supported. Furthermore, special aspects can be defined, concerning semantic information for visualization purposes (e.g. names for annotation).

Figure 7 shows a screenshot of some meta data of a data set of US cities (9 variables for 329 data records). Different windows show different meta data information. The window in the middle left for example shows meta data for the whole data set in a textual form (e.g. relevant keys, joint information content, correlations, number of data sets and variables). Below this outliers of the data set are displayed. The window lower right shows a visualization of the joint information content of different variables. The window above it displays meta data for the dimensions.

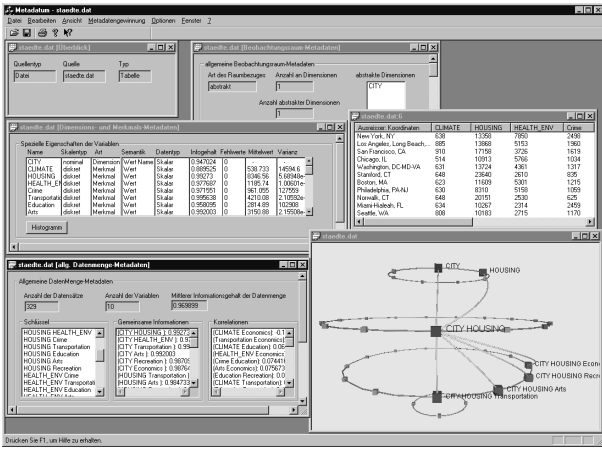


Figure 7. Screenshot from the framework: Overview of types of different meta data for a US cities data set

## 5 Visualization of meta data

Using textual output to interpret complex meta data is often not effective. Visualization can help to improve the understanding and interpretation of meta data. On the one hand known visualization techniques can be applied for analysis and representation of meta data. We adapted some existing techniques in this context, for instance histograms or visualization techniques for hierarchical structures. On the other hand we also developed new techniques, especially for the analysis of meta data, such as visualization techniques for regions of interest or visualization techniques showing joint information content. Figure figures 8 and 9 illustrate two of these techniques.

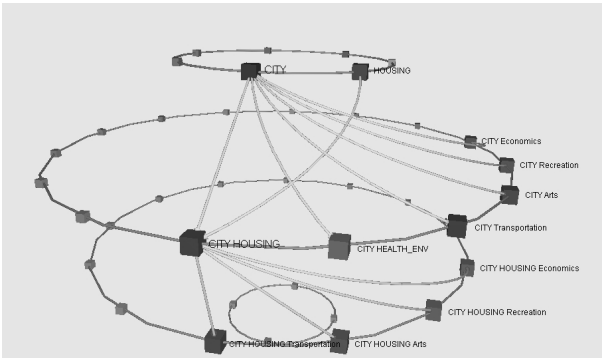


Figure 8. Visualization of meta data in our framework: visualization of joint common information of the US cities data set; the variables city and housing their relevant combinations with other variables are annotated)

Figure 8 shows a developed technique for the representation of joint information content of variables. This technique visualizes the relevant joint information over a threshold  $s$  of any combination of the variables<sup>4</sup>. The prin-

<sup>4</sup>The relevant joint information content was calculated applying an en-

tuple can be summarized as follow: tuples of variables with joint information content are mapped onto cubes which are laid out into a hierarchy of circles in 3D:

- The information content of each single variable is mapped onto the top level (e.g. the information content for the variable city (cf. fig. 8)),
- the second level contains pairs of variables with a relevant common information content (e.g. for the pair of the variables city and housing),
- the third level contains the triples of variables (e.g. the triple city, housing and arts,
- the fourth level contains the relevant 4-tuples (not annotated in fig. 8))
- and so on.

The color of a cube represents the information content of the according tuple. Furthermore, relationships between several tuples<sup>5</sup> can be mapped onto curves in 3D<sup>6</sup>. Using integrated navigation and selection operations, cubes can be faded in and out as well as labeled suitably.

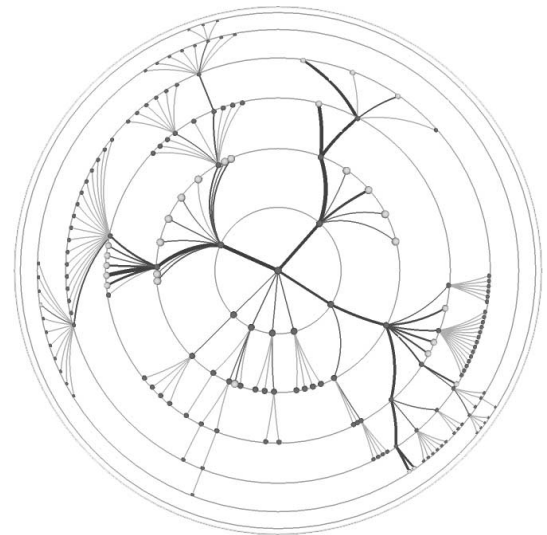


Figure 9. Visualization of meta data: visualization of hierarchical clustering using the technique MagicEyeView from [17]

Figure 9 shows one possibility for representing hierarchical clusters with the Focus+Context technique MagicEyeView [17]. This technique makes it possible to illustrate both hierarchies and special cluster properties. Different colors are used to distinguish between cluster nodes and object nodes within the hierarchy tree. Furthermore, the

trophy measure (cf. e.g. [16]).

<sup>5</sup>Two tuples are related, precisely then, if the elements of one tuple are a subset of the elements of the other tuple.

<sup>6</sup>The same procedure can be done in 2D, too, with a hierarchy of lines, and tuples connected by arrows.

size of a cluster, i.e. the number of objects, is mapped to the cluster edge's size and color intensity. Additional cluster properties such as cluster centroids can be displayed using cylinder icons [14]. A variety of interaction and navigation operations enables zooming and focusing hierarchies in combination with semi-automatic hiding of clusters.

To summarize, the visualization of meta data can be applied similarly to feature visualization to represent important characteristics of the data instead of the whole data set. This allows a deeper insight into the data and its structures. Our intention has been to support the visual data mining process. Depending on the user's intentions, more and more concise and detailed meta data can be gathered and displayed in an iterative refinement and parameterization process.

## 6 Conclusions and future work

Meta data describe important characteristics of a data set and can be applied for visual mining tasks. In this paper we specified a variety of meta data supporting these tasks. Furthermore, we developed an effective gathering strategy for meta data and included it in a general framework.

However, there are still challenges for future work to tackle. First, further evaluation of the introduced framework needs to be performed to determine its effectiveness and to verify its general applicability. Moreover, the integration of further visualization techniques into the process of meta data gathering is a task well worth performing. Visualization of extracted meta data as well as suitable visualization of raw data can be combined to gain an insight into data characteristics.

## References

- [1] C. Westphal and T. Blaxton. *Data Mining Solutions - Methods and Tools for Solving Real-World Problems*. John Wiley & Sons, Inc, New York, 8 edition, 1998.
- [2] D. Keim. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.
- [3] M. Kreuseler and H. Schumann. A Flexible Approach for Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), January-March 2002.
- [4] M. Ankerst. Visual Data Mining with Pixel-oriented Visualization Techniques. In *Proceedings of ACM SIGKDD Workshop on Visual Data Mining'01; San Francisco*, 2001.
- [5] K. W. Brodli et al. *Scientific Visualisation*. Springer-Verlag, Berlin, 1992.
- [6] P. C. Wong and R. D. Bergeron. 30 Years of Multidimensional Multivariate Visualisation. In: *G. Nielson, H. Hagen, H. Mueller: Scientific Visualisation. IEEE Computer Society, Los Alamitos*, pages 3–33, 1997.
- [7] H. Schumann and W. Mueller. *Visualization, Basics and general Methods*. (in german), Springer-Verlag, Berlin, Heidelberg, 1st edition, 2000.
- [8] P. K. Robertson and M.A. Hutchins. An Approach to Intelligent Design of Color Visualisation. In: *G. Nielson, H. Hagen, H. Mueller: Scientific Visualisation. IEEE Computer Society, Los Alamitos*, pages S. 179–190, 1997.
- [9] D. Silver. Feature Visualisation. In: *G. Nielson, H. Hagen, H. Mueller: Scientific Visualisation. IEEE Computer Society, Los Alamitos*, pages 279–293, 1997.
- [10] F. Reinders, F.H. Post, and H.J.W.Spoelder. Visualization of Time-Dependent Data using Feature Tracking and Event Detection. *The Visual Computer*, 17(1):55–71, February 2001.
- [11] V. Petoukhov, A.Ganopolski, V.Brovkin, M. Claussen, A. Eliseev, C. Kubatzki, and S. Rahmstorf. CLIMBER-2: a climate system model of intermediate complexity. Part I: model description and performance for present climate. *Climate Dynamics (2000)*, Springer Verlag, 16, 2000.
- [12] R. Rew, G. Davis, and S. Emmerson. NetCDF user's Guide. An Interface of Data Access. UCAR, 1993.
- [13] T. Nocke. Gathering and Specification of Meta Data for Visualization Decisions. Diploma Thesis; University of Rostock, Department of Computer Science; <http://www.icg.informatik.uni-rostock.de/~nocke/diplom.html>, 2000.
- [14] M. Kreuseler, T. Nocke, and H. Schumann. Integration of Clustering and Visualization Techniques for Visual Data Analysis. *Proceedings of the 25th annual Conference of the German Classification Society'01*, 2001.
- [15] I. Herman, M. Delest, and G. Melancon. Tree Visualisation and Navigation Clues for Information Visualisation. *Computer Graphics Forum*, 17(2):153–165, 1998.
- [16] H. Theisel. Analysis and Visualization Support for multi-dimensional scientific Data. *Informatik, Forschung und Entwicklung (in german)*, 10:91–98, October 1995.
- [17] M. Kreuseler, N. Lopez, and H. Schumann. A Scalable Framework for Information Visualization. *Proceedings of IEEE Information Visualization'00; Salt Lake City; Utah*, 2000.