

## **Semantic Search and Visualization of Time-Series Data**

**Tatiana Tekušová**

(Fraunhofer IGD, Darmstadt, Germany  
tatiana.tekusova@igd.fraunhofer.de)

**Viktor Voss**

(Technische Universität Darmstadt, Darmstadt, Germany  
viktor.voss@gris.informatik.tu-darmstadt.de)

**Jörn Kohlhammer**

(Fraunhofer IGD, Darmstadt, Germany  
joern.kohlhammer@igd.fraunhofer.de)

**Abstract:** In the economic and financial analysis domain a quick access to the right information plays a major role. Using current systems, the search for and presentation of data is very cumbersome. The data, mostly in form of time-series, is stored in various databases. In order to retrieve the searched data, the analysts need to know where to search and sometimes even the structure of the database and its coding. Then it is required to export the data, process the data and create a chart to view the data. This might take time from tens of minutes to hours. In our work we present a first prototype of an integrated search engine that takes as input a natural language query and offers graphic and text output depending on the user task. The system automatically identifies the time-series answers, types of graphical data presentation and shows the results in a web browser and in Excel. The knowledge-based expert system uses domain ontologies for extraction of economic terms in the search queries and specially built data type taxonomy with user task and chart type ontologies for identification of graphic output.

**Keywords:** Semantic Data Retrieval, Search Result Visualization, Information Visualization, Expert Systems, Economic and Financial Data, Time Series Visualization

**Categories:** H.3.1, H.5.2, H.5.m, J.4, M.6

### **1 Introduction**

In the economic and financial analysis a quick access to the right information plays a major role. Often large amounts of data have to be evaluated in short time. The data are usually stored in heterogeneous systems and in various databases. In their work, analysts combine numerical data from various sources with text (such as news, reports, etc.), and expertise that exists in the company.

In order to collect all the needed information, analysts spend a lot of time every day with information search. Analysts usually simply browse the data sources, or use data search options available by the data provides. These conventional systems for searching for financial and economic information, however, do not offer naturally formulated search queries such as “GDP composition in Germany” with direct graphical output.

The finding of relevant data does not finish the analysis task. In the analysis process, after the data is sought, the data is copied by hand into another system for further processing (e.g. into Excel). Finally, appropriate graphics for data presentation are created. The graphic construction requires from the analyst, in addition to the domain expertise, also skills in information visualization and graphic design.

In this paper, we introduce a first prototype of a semantic-based search engine which takes as input a natural language query, searches over financial information in across multiple financial data services and displays the results in a way that suits best the user tasks (e.g. as table, text or graph). A preliminary user study to examine the common analytic tasks with matching answers was conducted. In later stage, selected end-users were asked for feedback. Results of both studies are presented.

The paper is structured as follows: Section 2 discusses relevant work about automatic graphic data presentation. The third section describes current user search tasks and used systems for data search. In the following section, our framework is presented. Section 5 provides examples of search output visualization. Finally we conclude and discuss further work.

## 2 Related Work

Automatic generation of graphical data representation goes back to the work of MacKinlay [Mackinlay, 86]. The Automatic Presentation Tool (APT) defines graphic representation based on description of visual attributes. According to visualization design methodology, we can divide current systems into two categories - constructive bottom-up and template-based top-down method [Kamps, 99]. The comparison of the two techniques can be found in [Lang, 06].

The development of data type and visualization taxonomy, was based on work of Shneiderman [Shneiderman, 96] and by Tory et al. [Tory, 04] which match different data types with visualization driven by user tasks. A general overview of analytic tasks in visualization is provided by Amar et al. [Amar, 06]. Amar et al. [Amar, 04] also describe a knowledge task-based framework for design and evaluation of information visualizations. Fujishiro et Al. [Fujishiro, 00] describe a taxonomic approach to semi-automatic design of information visualization applications using modular visualization environment. An overview of visualization techniques specialized for time-series can be found in [Aigner, 07], taxonomy for temporal data visualizations can be found in [Daassi, 06]. Kohlhammer [Kohlhammer, 05] describes the use of domain ontologies for effective visualization.

## 3 User Tasks and Current State of Data Search

### 3.1 Search Tasks

In order to best meet the analytic requirements, we have asked ten financial and economic analysts, from various work domains, which search engines they use to find the data and what systems they use for data presentation. They were asked to provide as many data queries with their desired outcomes and estimation of the time needed to accomplish the analytic task as possible (see Table 1).

More than two thirds of the searches are aimed at getting data in tables, single data observations or data charts (see Figure 1, left). Chart responses included line, column, stacked column and scatterplot charts. The respondents use financial and internal databases. They wish to copy the data to create a chart in Excel.

The time needed to accomplish the task (from data search to production of the relevant data presentation) varies strongly according to the desired output and the expertise in the searched data sources. Searching for novel information, not used by the user on a regular basis, is much more time demanding than a search for data where the data source is already known. On average, users spend 18 minutes when creating charts and only 3 minutes when looking for a single value. However, the span between minimum and maximum time is very high (see Figure 1, right) reaching a maximum time spent for one task of 4 hours.

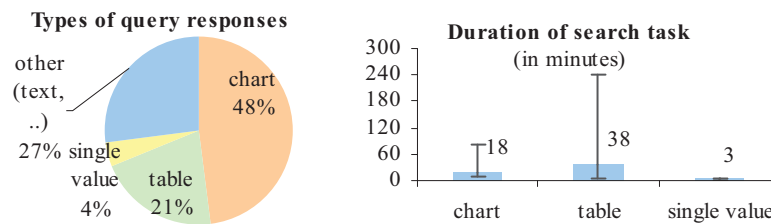


Figure 1: Required search query responses. Left: Types of search responses, Right: Time needed for response by type (average, minimum and maximum)

Question	Expected Output	Expected Duration
Show me EUR/SKK development in the last week	Intra-day line chart.	5mins
DJI index in the past year	Tick data, line chart	5mins
Show me GDP composition in Germany and France?	2 pie charts showing GDP components	30 mins

Table 1: Example Questionnaire answers

### 3.2 Data Sources and Currently Used Search Engines

The sources of information needed are spread across multiple data bases from various data providers (e.g. Bloomberg, Reuters, DataStream, Eurostat, ECB or internal sources). Each data provider uses own data classification system and data description system making the search for data across databases very complicated. Search in such databases is usually constrained to explorative overview of database entries by category. Some providers offer search engines which should facilitate the search for data. These search functions, in general, offer as outcome long lexicographic ranking (lists) of the names of the available time series (see Figure 2).

ESA Germany	Gross domestic product at market price - Current prices - EC/Euro - No adjustment	ESA.DE.N.0000.B1.G000.1000.TTTT.V.UA	100
AME Fin Germany	Gross domestic product at current market prices - Met EURO-DEM - AMECO data class: Data at current prices	AMEA.DEU.1.0.0.0.L0V0D	100
AME Fin Germany	Gross domestic product at 2000 market prices - Met EURO-DEM - AMECO data class: Data at constant prices	AMEA.DEU.1.1.0.0.OV0D	100
ESA Germany	Gross domestic product at market price - Constant prices - National currency (Euro) - No adjustment	ESAA.DE.N.0000.B1.G000.1000.TTTT.Q.N.A	100
ESA Germany	Gross domestic product at market price - Constant prices - National currency (Euro) - No adjustment	ESA.Q.DE.N.0000.B1.G000.1000.TTTT.Q.N.A	100

Figure 2: Example of search results using ECB search engine (<http://sdw.ecb.int>).

## 4 Semantic based Time-Series Search and Visualization

### 4.1 Search and Visualization Framework

Our knowledge-based system for integrated data search and visualization consists of three major parts (see Figure 3): in the first stage the input query is processed in order to extract relevant search criteria, then a suitable set of responses is compiled, extracted from the database and finally the responses are shown in a web browser. Our system is based on Conweaver framework [Conweaver]. ConWeaver provides automated knowledge network construction, semantic integration and intelligent search in portals and intranets.

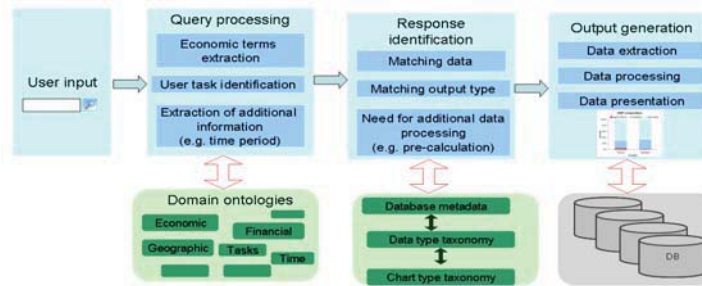


Figure 3: System framework

### 4.2 Query processing and Response Identification

In the first stage, the input query is processed in order to extract economic and financial terms in the query. In our work, we have created modules that identify economic and financial terms in the data descriptions, by looking for longest terms first. For example “GDP per capita” is a different term than “GDP”. Furthermore synonyms (consumer prices and inflation) and abbreviations (e.g. GDP = gross domestic product) can be extracted.

For defining the suitable data presentation type, we have analyzed visualization types used in the financial and economic domain and user study. A template based system for graph generation, which can be adjusted by the users, is applied. In the system, each chart type is described by type of data input, visualization parameters and analytic purpose (keywords).

For each chart type identified in the user questionnaire (line, column, pie, scatter plot, etc.), a description was created and saved in the chart type ontology. It describes types of data that the specific chart can use as input and the user tasks for each chart

type. Each chart parameter (each chart axis) is described via the type of data that it takes as input (quantitative, nominal, ordinal, etc.) and its cardinality. For example a line chart has 3 “axis”: X axis that usually takes quantitative data of cardinality  $>5$ , Y Axis takes quantitative data of cardinality  $\geq 1$ , and Legend takes nominal data of cardinality 1-10 (more than 10 series usually make the graph overcrowded). Chart is also characterized by “keywords” indicating user task connected to the specific chart type. The keywords were identified via matching search queries with expected responses (see in Table 1). For example, “composition” indicates pie chart, “development” indicates line chart.

### 4.3 Output generation

In the output generation stage, the identified suitable response data are collected from the database. The output type is influenced by the rank of matching data, the goodness of fit of the data to specific chart types and user task match with the chart keywords... The output data type and the identified user task are matched with the best possible visualization type. The chart type match (see Figure 4) uses weighted data type hierarchy for defining the best possible output type for a given data set to be shown. The output match is quantitatively expressed by the product of the data type hierarchy step weights. The weights are pre-set parameters, which can be adjusted. Additionally the user tasks influence the output type. Additionally, the the weight of each chart type is influenced by match of chart keywords with the search query. A ranking list of output types is thereby created. The top ones are used in the output visualization.

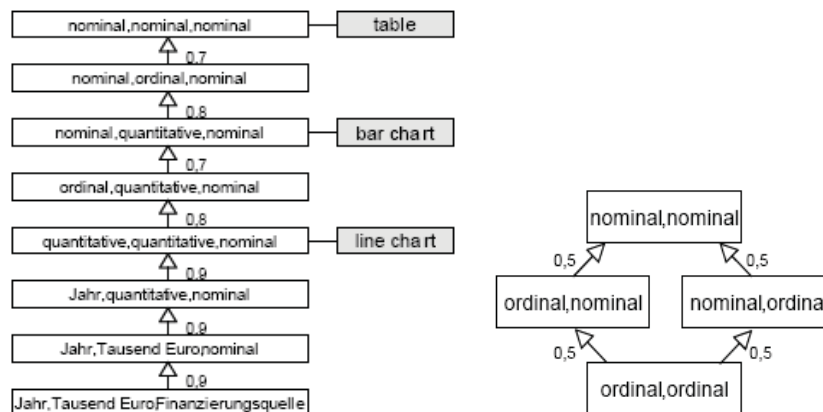


Figure 4: Semantic search for visualization type. Data set type hierarchy.

## 5 Search System Output and Evaluation

For testing the search and visualization framework, real world data set from the Federal Reserve Bank of St. Louis (<http://research.stlouisfed.org/fred2/>) were used. The FRED® database contains about 14000 economic and financial time series in flat text files. It provides raw data and time series descriptions.

The engine displays the search results on a Web frontend in the form of dynamically generated HTML pages based on search results. The data that match best the user query are displayed in a graphical form determined in the response identification phase (see Figure 5). In our prototypic application, the user may click on each chart to see the underlying data table in Excel (see Figure 6). The users can then further process the data results or design further charts in this application. On demand other export formats can be employed. In addition to numeric data, the search engine offers a list of documents searched on the internet using common text search engines (e.g. Google). Using the ConWeaver system [Conweaver] we could have easily included also search in internal documents, however we did not have any such data at disposal. This provides a composite of relevant information in one window.

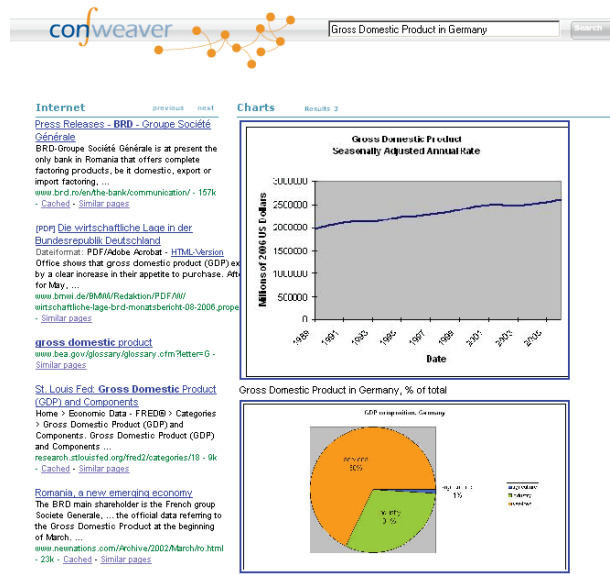


Figure 5: Example search result using our prototype tool

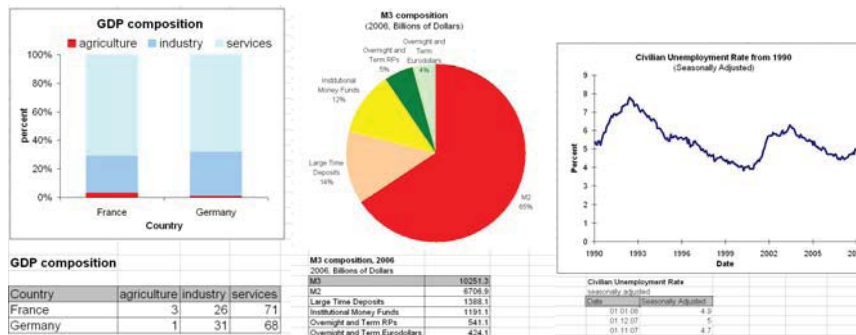


Figure 6: Selected user responses exported into Excel

The prototype user queries were used to test the results of our search engine. The users liked very much the possibility to see an overview of the search responses in graphic/table format. It enables them to quickly identify major data developments without cumbersome data export and graph creation for each single data series. This quick and easy way of preliminary data analysis saves the analysts' time which they can use for more in depth data analysis using e.g. data mining methods.

## **6 Conclusions**

A prototype system for search in economic time-series databases which takes as input a natural language query, and graphically visualizes the search results has been developed. The choice of the proper visualization type is task-dependent using data taxonomy and visualization type ontology. Domain ontology for economic data is used for finding suitable search response. The system offers time saving to the potential users when looking for and presenting economic and financial data allowing them to concentrate on the data analysis part of the analytic process.

## **7 Future Work**

In near future we would like to explore the effect of different parameter set-up for hierarchy type match. It would be also needed to expand the used financial and economic ontologies using new data sources (e.g. further classifications) and newly developed publicly available ontologies. This would allow for more accurate search responses. We would like to include automatic data processing using financial functions (e. g. calculation of indices) and widen the spectrum of used visualization techniques. It would be interesting to combine search for numeric data with search in news feeds. For example, particular patterns (peaks, strong decreases) could be matched with the relevant stock market interpretation.

## **Acknowledgements**

The authors thank the ConWeaver Company, especially Richard Stenzel and Thomas Kamps for their support and the possibility to use their engine for semantic search. We are thankful to the user study participants for their activity and helpful input. We would like to thank Tobias Schreck for helpful comments. This research was partly funded by the German Federal Ministry of Economics and Technology within the THESEUS project (<http://theseus-programm.de>).

## References

- [Aigner, 07] Aigner W., Miksch S., Muller W., Schumann H., Tominski Ch.: Visualizing time-oriented data--A systematic view, *Computers & Graphics*, Vol. 31, No. 3. (June 2007), pp. 401-409, 2007.
- [Amar, 04] Amar, R.; Stasko, J.: A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations, In Proc. IEEE Symposium on Information Visualization, 10-12 Oct 2004, pp. 143-150, 2004
- [Amar, 06] Amar, R., Eagan, J., Stasko, J.: Low-Level Components of Analytic Activity in Information Visualization. Georgia Institute of Technology, 2006
- [Audersch, 05] Audersch, S., Flach, G.: Semantic Web Technologien für die visuelle Exploration und Fusion multivariater Datenbeständen. In Proceedings of Berliner XMLTage, 2005
- [Berners-Lee, 01] Berners-Lee T., Hendler, J. und Lassila, O.: The Semantic Web – A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 2001
- [Card, 99] Card, S.K., Mackinlay, J.D., Shneiderman, B.: Readings in Information Visualization. Using Vision to Think. Morgan Kaufmann, 1999
- [Codd, 70] Codd E. F.: A Relational Model of Data for Large Shared Data Banks, *Communications of the ACM*, Volume 13, Number 6, pp. 377-387, 1970
- [Conweaver] [www.conweaver.de](http://www.conweaver.de), accessed 7<sup>th</sup> April 2008
- [Cox, 01] Cox, K., Grinter, R.E., Hibino, S.L., Jategaonkar Jagadeesan, J., Mantilla, D.: A Multi-Modal Natural Language Interface to an Information Visualization Environment. *International Journal of Speech Technology*, 2001
- [Daassi, 06] Daassi C., Nigay L., Fauvet M.-C. : A taxonomy of temporal data visualization techniques, *Revue en Sciences du traitement de l'Information*. Cépaduès Editions., vol5, no2, pp. 41-63, 2005.
- [Date, 00] Date, C.J: An Introduction to Database Systems, 7th Edition. Addison- Wesley, Reading, MA, USA, 2000
- [Fujishiro, 00] Fujishiro, I.; Ichikawa, Y.; Furuhashi, R.; Takeshima, Y.: GADGET/IV: a taxonomic approach to semi-automatic design of information visualization applications using modular visualization environment, In Proc. IEEE Symposium on Information Visualization, 2000. InfoVis 2000. 9-10 Oct. 2000 Page(s):77 – 83, 2000
- [Kamps, 99] Kamps, T. M.: A Constructive Theory for Diagram Design and its Algorithmic Implementation, PhD Thesis, Technische Universität Darmstadt, 1999
- [Kohlhammer, 05] Kohlhammer, J., Encarnação, José L. (Referent), Knowledge Representation for Decision-Centered Visualization, (Dissertation, Technical University of Darmstadt), GCA-Verlag, Reihe: Forschung und Wissen - Informatik, Herdecke, 2005
- [Lange, 06] Lange, S., Nocke, T., Schumann, H.: Visualisierungsdesign – ein systematischer Überblick, German, „Visualization design – a systematic overview“, Universität Rostok, 2006
- [Mackinlay, 86] Mackinlay, J.: Automating the Design of Graphical Presentations of Relational Information. *Transactions on Graphics*, Vol. 5, Nr. 2 , April VI 1986

[Shneiderman, 96] Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization. Proc. 1996 IEEE Conference on Visual Languages, 1996

[Sowa, 91] Sowa, J. F.: Principles of semantic networks. Explorations in the representation of knowledge. Morgan Kaufmann, 1991

[Tory, 04] Tory, M.; Moller, T.: Rethinking Visualization: A High-Level Taxonomy, In Proc. IEEE Symposium on Information Visualization, 10-12 Oct. 2004 Page(s):151 – 158, 2004

[Ware, 99] Ware, C.: Information Visualization: Perception for Design, Morgan Kaufmann, 1999