Project ICT 287534
Start: 2011-09-01
Duration: 36 months
Co-funded by the European Commission within the 7th Framework Programme

SEMANCO Semantic Tools for Carbon Reduction in Urban Planning

Deliverable 5.2
Tools for Energy Analysis

Revision: 8
Due date: 2013-05-30 (m21)
Submission date: 2013-07-31
Lead contractor: HAS

Dissemination level		
PU	Public	X
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Deliverable Administration & Summary					
No & name	**D5.2 Tools for Energy Analysis**				
Status	Final	**Due**	m21	**Date**	2013-05-30
Author(s)	German Nemirovski (HAS), Andreas Nolle (HAS), Michael Wolters (HAS), Alvaro Sicilia (FUNITEC), Leandro Madrazo (FUNITEC)				
Editor (s)	German Nemirovski (HAS)				
DoW	Deliverable 5.2 is the outcome of Task 5.*2 Energy analysis, and optimization and strategic decision tools.* It comprises tools for energy analysis: A package of tools and algorithms for data mining, data and knowledge representation, building and energy performance simulation, energy consumption planning and reasoning. The tools access the data and semantic metadata by means of interfaces exposed by SEIF.				
Comments					

Document history

V	Date	Author	Description
1	2013-05-07	German Nemirovski (HAS), Andreas Nolle (HAS), Alvaro Sicilia (FUNITEC)	First version available on the SharePoint
2	2013-07-09	German Nemirovski (HAS)	Review and editing
3	2013-07-25	German Nemirovski (HAS)	Review and editing
4	2013-07-29	German Nemirovski (HAS)	Review and editing
5	2013-07-29	Leandro Madrazo (FUNITEC)	Editing
6	2013-07-29	German Nemirovski (HAS), Michael Wolters (HAS), Andreas Nolle (HAS)	Review and editing
7	2013-07-30	Leandro Madrazo (FUNITEC), Andreas Nolle (HAS), German Nemirovski (HAS)	Layout, review and editing
8	2013-07-31	German Nemirovski (HAS)	Answers to comments from reviewers

Table of Contents

EXECUTIVE SUMMARY

Introduction

Deliverable 5.2 *Tools for Energy Analysis*, developed within Work Package 5 *Integrated Tools*, summarizes the work done and the results achieved in Task 5.2 *Energy analysis, and optimization and strategic decision tools*, whose main goal is the development of analysis tools which combining visual-based approach with complex data mining techniques. The target group of the created tools are stakeholders and consultants supporting stakeholders in their decision making process.

The tools developed in Task 5.2 have three different purposes:

- ***data retrieval and systematization***, whereby the data sources associated with a particular task are retrieved and transformed to a structure and format facilitating the next task.
- ***statistical data analysis***, for instance targeting prediction of missing values, finding dependencies or relations between values of particular attributes, the classification or clustering of objects represented by their attributes. Unlike the other tools developed in WP 5 which are mostly based on (exact) calculations, the energy analysis are tools based on probabilistic approaches such as approximation or regression.
- ***visualization of data and analysis results***, making the results of analysis transparently understandable for the stakeholder who is about to complete a particular task or who must make a decision.

In order to facilitate these tasks the following components have been developed, whereby as a component we understand either a business process model addressing the requirements of different user categories or a software product that supports such a model and interacts with the user over a GUI or interacts with other software products by means of (web) services:

- *Data analysis focused workflow*
  This workflow supports the decision making process for all user roles and enables each user a slim, light weighted and efficient task flow.
- *Web based GUI for workflow support*
  This GUI supports the communication between users and offers access to web-services implementing single workflow steps which are carried out by the system.
- *Tool for querying RDF data stores for the purposes of data analysis*
  The methods implemented by this tool support the retrieval of RDF data and its transformation to fact tables. It facilitates statistical methods for the analysis and visualisation of this data. The tool is implemented as an extension to RapidMiner.
- *Tool for automated generation of SPARQL queries*
  This tool uses the data semantics specified in the SEMANCO ontology to simplify data retrieval by an end user. Furthermore, it offers a web interface and uses dedicated SEMANA services.
- *Set of services for visualization of data results*
  These services are able to interact with 3DMaps, the tools developed by Agency9. Furthermore, these tools integrate Mondrian and JFreeChart, two external systems for the visualisation of scatter point plots, charts, pie diagrams etc.

- *SEMANA a set of web-services*
  SEMANA comprises a set of web services which are able to support the workflow that focus on data analysis. It includes services for integration of external tools (RapidMiner) as well as services that implement internal functions (flow control, query generations, etc.).

In addition, the following third party software has been integrated:

- *Data mining tools RapidMiner and RapidAnalytics*
  RapidMiner facilitates the design of data mining processes, their running and visualization of their results. However, RapidMiner is a standalone tool and it must be installed on a personal computer for a single user. It cannot be simply integrated in any environment developed by a third party. However, the analysis results obtained using RapidMiner can be integrated by means of another tool, RapidAnalytics. RapidAnalytics enable data mining processes developed using RapidMiner to be declared as services which can in turn be integrated into environments developed by third parties. The integration of the RapidAnalytics services has been achieved through their interaction with SEMANA services that are an essential part of the SEMANCO platform.

- *Data visualization systems Mondrian and JFreeChart*
  Both of these tools allow for the graphical representation of data, e.g. as a scatter plot matrix, parallel coordinate matrix, pie or column chart, etc.

Three data mining processes, one for each case study, have been implemented for demonstration purposes. Currently the energy analysis tools integrated into SEMANCO platform are not available online and cannot be used by project partners and associated organizations for demonstration purposes. They will be made available online during the second iteration of the implementation of the demonstration scenarios to be carried out in Task 8.3.

In Task 4.5 *Integrated Platform* these components will be integrated into the SEMANCO platform, which is basically a web based application that uses as its kernel 3DMaps, a system for presentation of geographic – also urban –landscape using 3d vector graphics.

When developing the tools for energy analysis we followed the methodology presented in Deliverable 1.8 *Project Methodology* (Madrazo et al., 2012):

i) Activities leading to the reduction of carbon emissions and related to a particular use case have been described;

ii) The data and tools required have been identified, whereby use cases have provided the specifications for the tools to be developed;

iii) The implementation of the use cases in the actual working conditions in Task 8.2 *Implementation* has provided additional feed-back to the tools development;

iv) Further application of the energy analysis tools will be reflected in Deliverable 8.3 *Intermediate report on implementation* and therefore contribute to their refinement.

Three consortium partners have participated in the development / integration of the tools for energy analysis: HAS, FUNITEC and Agency9. The work in this task has been led by HAS. HAS also developed the data analysis focussed workflow, SEMANA Services, integrated RapidMiner and RapidAnalytics, extended RapidMiner by a component for RDF data retrieval, developed a shortest path search service for automated query generation and implemented three analytic processes for three case studies for demonstration purposes.

FUNITEC created the web based GUI supporting data analysis focussed workflow, integrated the GUI into the SEMANCO platform, and developed the tool for automated query generation. However an essential part of such integration will be carried out within the Task 5.4 *Integrated Platform*.

Agency9 developed services for visualization of analysis results.

Conclusions and Perspective Work

The tools for energy analysis facilitate data mining using with semantic data in the urban planning domain. Sophisticated querying of RDF triple stores and data visualization techniques have been developed. The experiments conducted with data of the three case studies have shown the usefulness and high potential of the tools.

In the near future we will address a classical business intelligence approach, comprising techniques like star schema data storage, data cube/data mart operations like slicing and dicing as well as to the solutions of well-known problems, such as summarizability. Work in this direction has started and will be further developed during the project.

1 INTRODUCTION

1.1 Purpose and target group

The purpose of this deliverable is to report on the work done in Task 5.2 *Energy analysis, optimization and strategic decision tools.*

The target group of the created tools are users, stakeholders and consultants who are all supported in conducting their decision making process. The user roles targeted are described in detail in Section 3 *Data Analysis Focussed Workflow*.

The work done in this task can be summarized as follows:

- Development of a ***workflow*** that defines user roles, the flow of activities to be carried out and the tools that each user role requires. .
- Development of a ***GUI*** that supports the workflow and collaboration between all user roles by providing a space for interaction. At the same time, by means of the GUI, all developed tools/component are integrated into the common SEMANCO platform. An essential part of this integration will be carried out within the Task 5.4 *Integrated Platform*.
- Development of a set of ***services*** to be invoked from the GUI and implementation of the business logic of the tools for energy analysis.
- ***Integration*** of standard tools such as RapidMiner and RapidAnalytics for data mining and JFreeChart and Mondrian for data visualization.

1.2 Contribution of partners

Tools for energy analysis developed in Task 5.2 and described in this report have been developed by HAS, FUNITEC and Agency9. The work in this task has been led by HAS. In particular, the involved partners have made the following contributions:

- HAS developed the data analysis focused workflow described in Section 3, SEMANA Services (Appendix A), integrated the RapidMiner and RapidAnalytics tools for data mining and publishing of data mining processes as services (Section 4.4), extended RapidMiner by a component for querying RDF data (Section 4.5.2), developed a shortest path search service for automated query generation, and developed three analytic processes for three case studies to demonstrate application of tools for energy analysis.
- FUNITEC created the web based GUI supporting data analysis focussed workflow (Section 4.2), integrated the GUI into the SEMANCO platform, and developed the tool for automated query generation (Section 4.5.1).
- Agency9 has developed services for data and analysis results visualization (Section 4.6).

Furthermore, UOT, CIMNE, FORUM and Ramboll have been involved in the development of data mining processes related to their respective case studies (Sections 2.2, 2.3, 2.4 and 4.3.1). These partners not only delivered data but also made substantial contributions to the process design and to the evaluation of results.

1.3 Relation to other activities in the project

The work carried out in this task follows the methodology presented in Deliverable 1.8 *Project Methodology* (Madrazo et al., 2012). Accordingly, activities leading to the reduction of carbon emissions within a particular use case have been described and data and tools required have been identified. This way, use cases have provided the specifications for the tools to be developed in this Task. The implementation of the use cases in the actual working conditions in each case study, which has been carried out in Task 8.2 Implementation, have provided further feed-back to the tools development. The application of the energy analysis tools in Deliverable 8.3 *Intermediate report on implementation* will further contribute to the refinement of the tools presented in this report.

The tools developed in Task 5.2 will be integrated in the platform being carried out in Task 5.4 *Integrated Platform*.

2 ENERGY ANALYSIS TOOLS: A USER PERSPECTIVE

2.1 Introduction

The tools developed in Task 5.2 and described in this report address the needs of stakeholders involved in the decision making process on different urban scales insofar as they enable them to carry out advanced analysis on energy related data at different scales. For instance at the neighbourhood scale, the tools can help to select materials and technologies to be used for building refurbishment. At the municipal scale, the goal can be to optimize the energy consumption and to reduce the CO_2 emissions in a new planned area or in a refurbishment plan; and at the regional scale, the tools can help to define a range of strategies to achieve the previously established sustainability objectives.

Often stakeholders and decision makers are missing the information they need to carry out tasks such as those described above. This information can refer to figures, classifications, hidden dependencies or rules. The reasons why information is missing vary. In some cases, figures such as energy consumption or CO_2 emission values are unavailable for buildings as they have not been built or refurbished yet. Some attributes of residential buildings such as the inner comfort temperature, the number of appliances or the average time spent at home by the inhabitants may be unknown because capturing them by observation requires significant resources or it is simply not possible due to privacy protection laws. Some other information, such as similarities or dependencies, can be contained in available data but must be made explicit, which is a complex and time-consuming task.

Yet, having this information may be crucial for decisions in urban planning, in particular for comparing of alternative planning scenarios or searching for energy "hotspots", that is, areas whose performance form the point of view of energy consumption or CO_2 emission is critical. Ascertaining such missing data is one of the goals of the data analysis tools described in this report. In this connection the main tasks carried out by the tools are:

- ***to retrieve and systematize data*** available in associated data sources and related to a particular task. For instance, if a task is carried out at the macro level (city/municipality) such data can, among others, comprise classifications of buildings according to their usage (family house, apartment block, educational building, shop, storage, garage, etc.), or the construction year (1931-1950, 1951-1970, 1971-1991, etc.), statistics of energy consumption for each class of buildings, forecasts for energy consumption for selected building classes. Also operations such as aggregation and summarisation across geographic/urban scales is part of data retrieval process. The calculation of average or mean values can be embedded in data queries.
- ***to facilitate statistical data analysis***. If a task, for instance, is carried out at the macro level (city/municipality) a numeric regression algorithm (linear regression) might enable the development of a method for predicting the energy consumption for all classes of buildings to be developed in the future.
- **to *visualize analysis results*** in a way and form easily understandable for the stakeholder and to effectively help him/her to complete a particular task. For instance, if a task is carried out at the meso level (district/block), the results of the energy analysis tools should be visualized on a 3D map, whereas at a district scale each building would be painted in a colour corresponding to its predicted SAP rate (Figure 1).

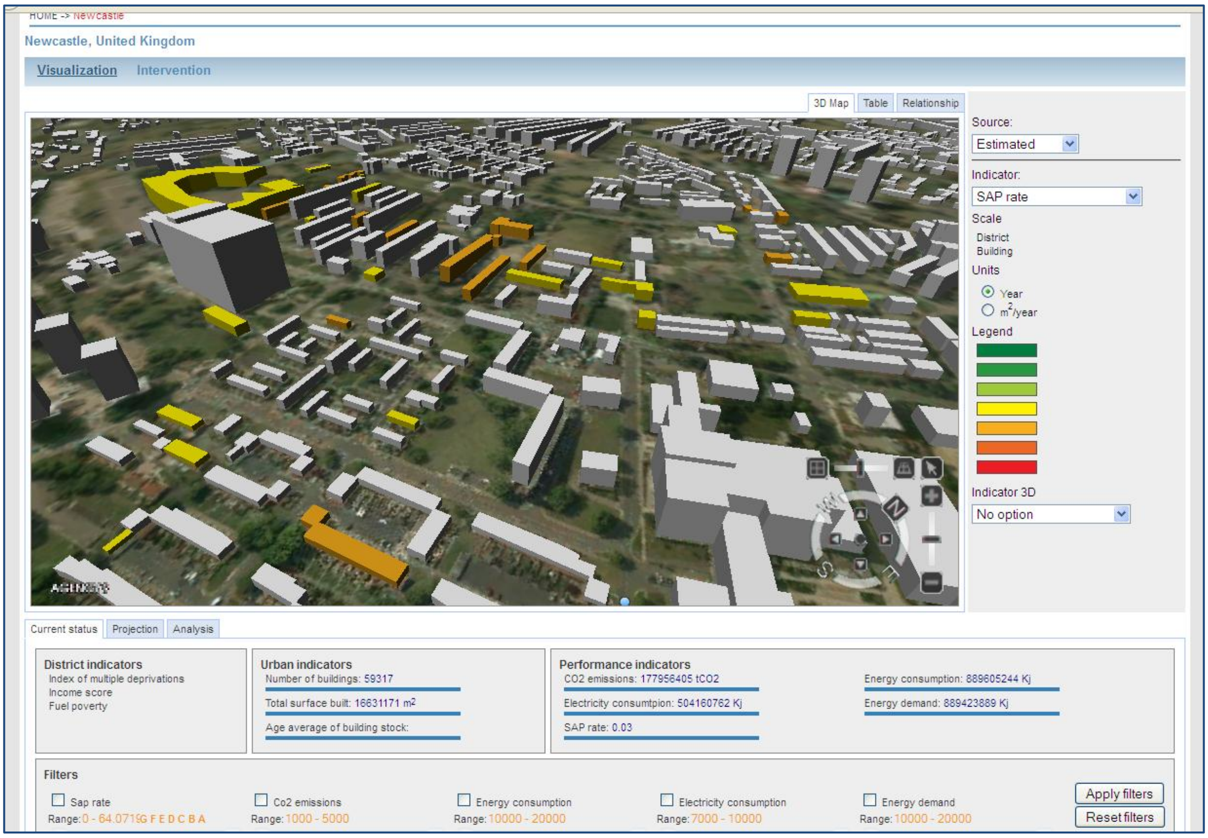


Figure 1. Visualizing buildings according to their predicted SAP rate

The tools have been developed within the context of the three case studies, in particular to the use cases and activities identified in the Deliverable 6.1 (Crosbie & Crilly, 2013). In particular the evaluation of tools have been carried out within the context of the following use cases:

- **Case study:** Newcastle

  Use case UC4-N: *Identify low income (Fuel Poor) households living in energy intensive dwellings with a poor SAP (Domestic Energy Efficiency Rating)*

  Activity A11: *Estimate the energy consumption and* CO_2 *emissions from the existing domestic dwellings*

  In this case study, the goal of the tools is to simplify the SAP calculation methodology. More precisely, to replace the actual calculations by a high probability prediction. The tools are described in Section 2.2.

- **Case study:** Copenhagen (North Harbour)

  Use case UC3-C: *Calculate impacts of alternative energy supply and demand options on CO2 reduction, final energy consumption and primary energy consumption.*

  Activity A19*. Calculation of total energy demand in a baseline and alternative scenario*

  In this case study, the tools for energy analysis help to forecast future energy consumption by different types of buildings. The approach is described in Section 2.3.

- **Case study:** Manresa

  Use case UC4-M: *To compare different alternative urban plans between them and against the baseline.*

In this case study, the tools for energy analysis are applied to identify different archetypes/categories of buildings according to their energy consumption and CO_2 emission, to identify the archetypes of buildings in different scenarios and in doing so to enable the comparison of such scenarios. The tools are described in Section 2.4.

2.2 Newcastle: SAP value prediction

The Standard Assessment Procedure (SAP) is a methodology used by the Department of Energy & Climate Change (DECC) in the United Kingdom whose purpose is to provide accurate and reliable assessments of dwelling energy performance which is needed to for energy and environmental policy initiatives.[1]

The calculation of an estimated SAP value requires a significant number of parameters (25) whose collection is resource and time consuming. In this context arises the need to predict SAP values using data mining/machine learning using less parameters while achieving a comparable accuracy. If the set of required parameters is smaller, then physical inspection of houses might be avoided and the planning of refurbishment simplified.

Machine learning classification algorithms such as neural networks are usually applied for this kind of predictions. Prediction experiments using neural networks have been carried out for a set of Newcastle buildings. For each of these buildings a maximal set of parameters (25) as well as calculated estimations for the SAP values have been used, so that we could experiment with different parameter subsets and compare resulting predictions with the previously calculated estimations.

Figure 2 below shows the results obtained by using a set of 12 parameters, with the average root mean squared error of 6.166 +/- 0.000.

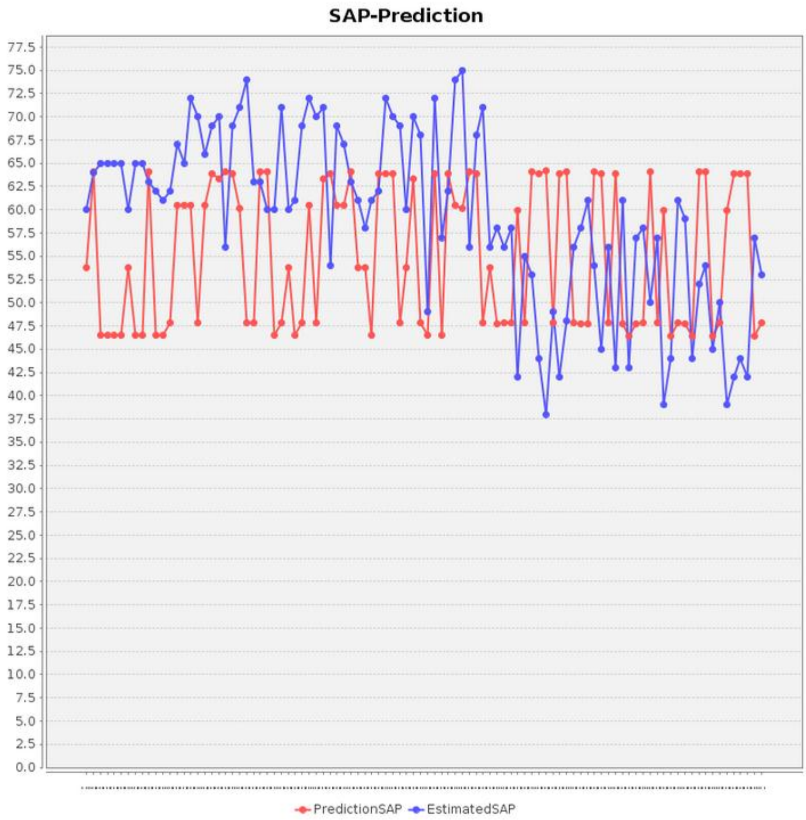


Figure 2. SAP value prediction for the case study Newcastle, compared to estimated SAP values

[1] https://www.gov.uk/standard-assessment-procedure

In the sections below, we will not detail show these results were obtained, which was a matter of careful algorithm selection and tuning. Also, we will demonstrate how the tools developed in Task 5.2 help to i) set up communication between different user roles responsible for these results and ii) simplify the application of complex mathematical algorithms to facilitate the work to users without sophisticated mathematical background.

2.3 Copenhagen: prediction of energy consumption for future buildings

In the case of Copenhagen New Harbour the focus of interest are buildings and districts that have not yet been built. Correspondingly, the analysis task targets the prediction of urban energy performance in future scenarios. In particular, it aims to determine a projection of energy demand until 2035 for three building (use) categories, and energy demand statistics for different building and age classes. Predictions until 2035 are produced for all available building (use) categories.

The figure below shows graphics of energy demand (axis Y) related to year (axis X) for four different building (use) categories. For categories with IDs 120, 140 and 230 which respectively correspond to a family house, an apartment block and an office building, the energy demand statistics from 1930 as well the projection until 2035 are available. For the *target* category, which is in this case *residential*, only energy demand statistics from 1930 to 2007 are available. Using data mining techniques the projection of energy demand until 2035 is predicted for the *target* category (the lower part of the blue curve in Figure 3 below).

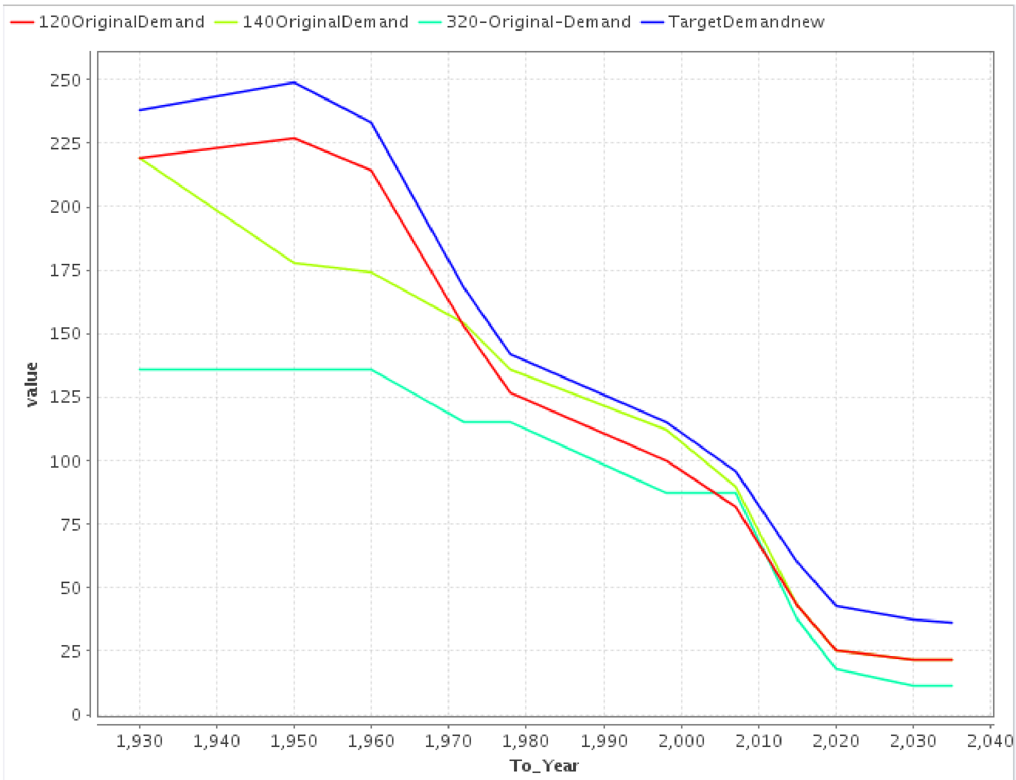


Figure 3. Prediction of energy consumption for the building use category Residential Building

This prediction is done with help of linear regression algorithm. Though linear regression is a comparatively simple form of regression its advantage is the ability to obtain reliable results with relatively small data sets (Martens & Dardenne, 1998). Our experience confirmed that in situations when the available training set is limited, which is obviously the case in the example above, linear regression performs better than more sophisticated alternatives.

2.4 Manresa: Clustering

In order to identify buildings archetypes, several houses in the city have been surveyed and their building geometry described in a form consisting of 34 parameters. The resulting data set has been supplemented by additional attributes retrieved from the census and the cadastre which were presented in Deliverable 3.4 *Interfaces with external tools* (Sicilia, 2013). These additional parameters refer to the number of inhabitants of the whole building, the number of apartments occupied, the number of occupied apartments divided by the total apartment of the building and the mean income (in Euros) of a neighbourhood.

In order to reduce the number of parameters the values have been normalized and processed with a principal component analysis algorithm. Clustering analysis using a K-Means algorithm have been carried out to assign houses to four different groups (clusters), where each cluster unifies houses sharing similar features. It became obvious that four reliable groupings/clustering of houses each with four / five parameters would be sufficient. In some settings it may even be sufficient to know only three parameters. The figure below illustrates the clustering results graphically. The axes x, y and z take values of roof U-value, façade 1 windows glass g-value and façade 1 orientation correspondingly.

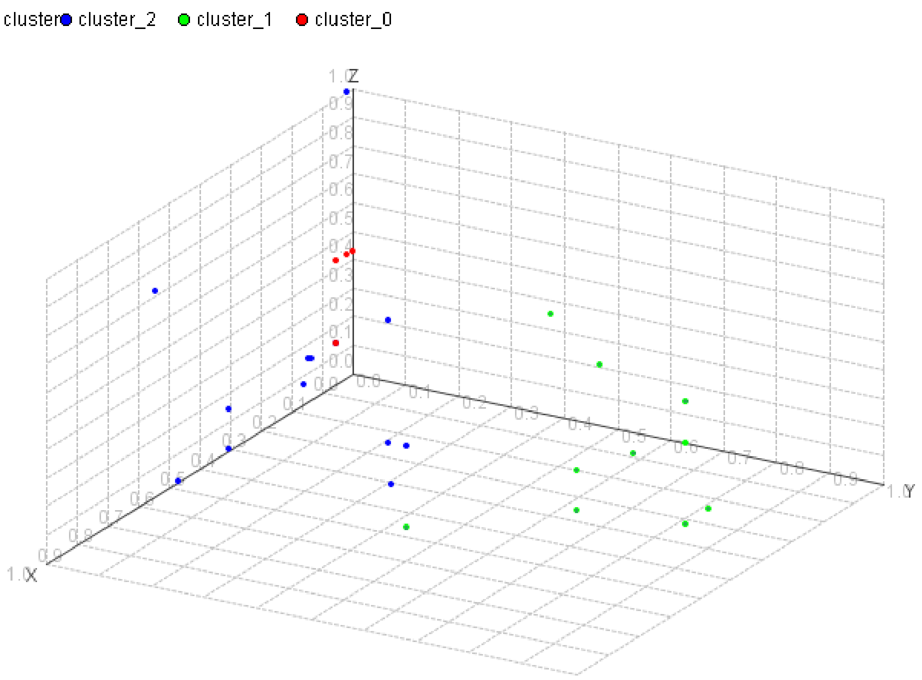


Figure 4. Visualization of clustering analysis Results for case study Manresa

If this hypothesis is confirmed by further experiments, the task of classifying buildings can be simplified significantly with this clustering tool.

3 DATA ANALYSIS FOCUSSED WORKFLOW

Interim evaluations carried out during the elaboration of the developed tools in the context of each case study as shown in examples above have demonstrated that the nature of three tasks mentioned in Section 2.1 (data retrieval, data analysis and visualization) requires knowledge and skills – for example, in using of query languages, data cleansing and transformation, or application of machine learning algorithms – that are beyond the reach of stakeholders who are typically urban planners, energy providers, architects, or politicians.

To facilitate the use of energy analysis tools by these stakeholders an innovative workflow has been developed that involves, apart from stakeholders, three further user roles associated with technical/mathematical/analytic skills. Correspondingly, single workflow steps that encompass the application of certain tools are assigned to particular user roles according to their skills and responsibilities (Figure 5).

The workflow considers two organization structures –represented as pools (grey rectangles) in the figure below– and four user roles – represented as lanes (white rectangles).

Pools:

- User Side (upper grey rectangle): an organization or unit responsible for the generation of decisions related to urban planning, political proposals, investments, etc.
- Consultant Side (lower grey rectangle): an organization that offers technical support to the user side, e.g. a company that possesses data mining and data access expertise and advanced knowledge in the urban planning related topics.

User roles:

- Decision Maker: a user responsible for generated decision, e.g. a director in a urban planning unit or organization, a politician or an real estate owner
- Data Mining Specialist: a user integrated in the organizational structure of the decision maker and having advanced knowledge in data mining.
- Ontology Expert: a user integrated into the structure of the Consultant Side and having the knowledge required for accessing data sources federated my means of semantic technologies for data integration developed in the SEMANCO project and implemented in the SEIF.
- Data Mining Expert: a user integrated in the organizational structure of the Consultant Side and having expert knowledge in data mining.

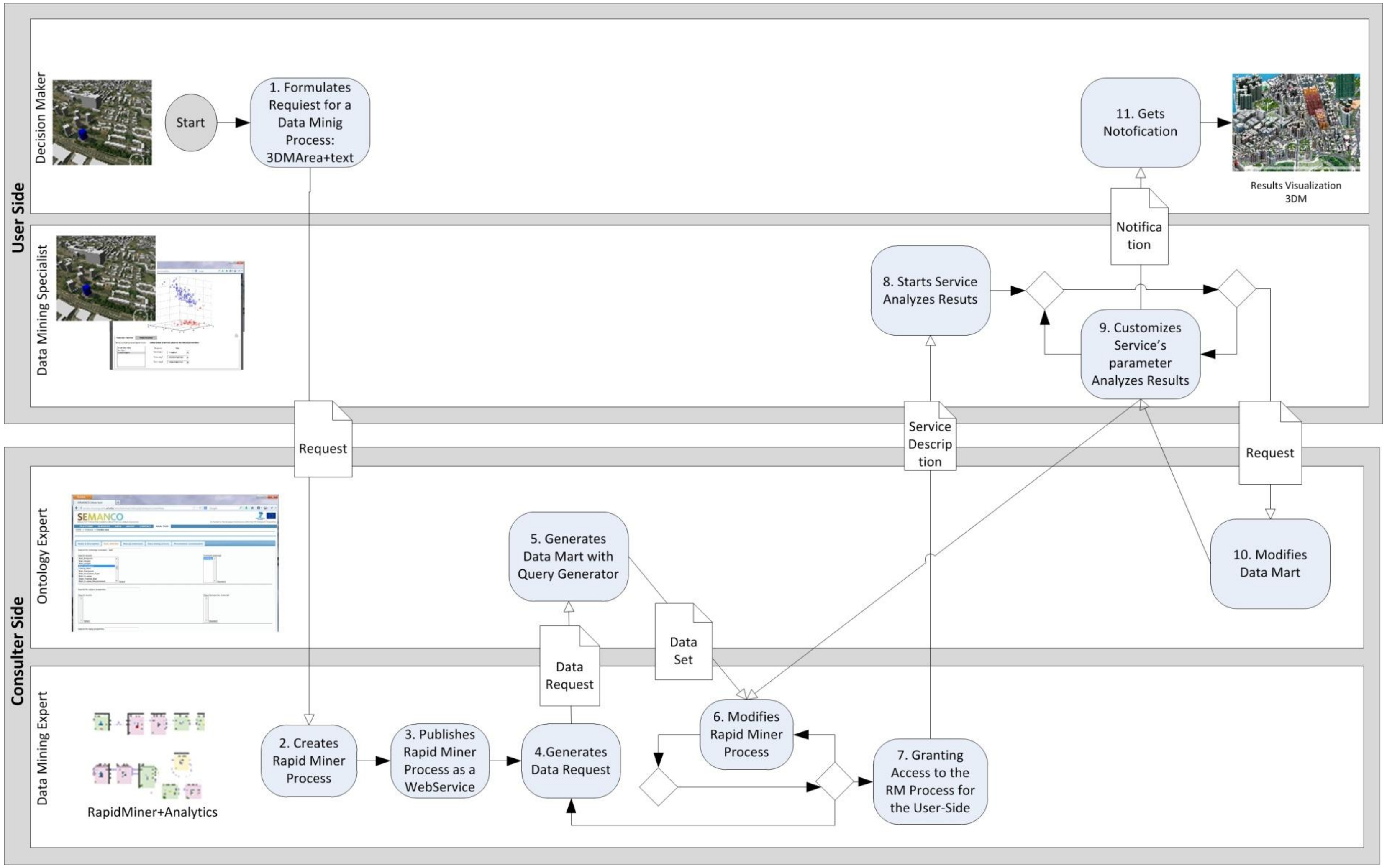


Figure 5. Marking Data analysis focussed workflow

The main idea behind this workflow is to outsource most of technical and technological processes to the Consultant Side in order to organize the decision making on the User Side in a slim, light-weighted and efficient manner. In particular, the data analysis workflow is organized as follows:

- A Decision Maker starts the workflow by formulating a request for data analysis. The request is submitted to a Data Mining Expert on the Consultant Side. The latter uses RapidMiner[2]. With RapidMiner he/she creates a data mining process which is aimed to address the request generated by the Decision Maker. Such a process usually contains from five to fifty different steps for data transformation, cleansing and analysis. Typically, at this stage the process uses a small set of data collected for testing purposes. After the process has been published as a web service it becomes available for other participants of the workflow. Then, a Data Mining Expert formulates a data request and forwards it together with the reference to the web service to the Ontology Expert (Consultant Side).
- An Ontology Expert generates and tests a data query according to the requirements mentioned in the request. The expert uses a query generation tool described in Section 4.5 and the latest version of Semantic Energy and Information Framework (SEIF).
- The generated query is sent to the Data Mining Expert who, using a RapidMiner extension presented in Section 4.5.2, integrates the query in the existing data mining process and evaluates the process basing on the “real” data. Depending on the evaluation results: i) the process can be modified, ii) a new data request can be sent to the Ontology Expert or, iii) if the evaluation is successful, a graphical representation of the analysis results is generated and integrated into the SEMANCO platform. Furthermore, a notification containing the reference to the process service is forwarded to the Data Mining Specialist (User Side). In fact, however, a set of several data mining processes is created to fulfil the requirements formulated in the initial analysis request by the Decision Maker. In this case, such notification would contain a set of references to the corresponding services.
- A Data Mining Specialist customizes the process by changing the values of selected parameters prepares and adjusting the graphical representation of the analysis results. If needed he/she may request that the Consultant Side, i.e. a Data mining Expert or ontology Expert, should modify either the process, the data query, or both.
- In the final stage of the procedure, the data mining process(es) is (are) conveyed to the decision maker who can “play” with it/them by modifying the values of a few selected parameters and watching the analysis results in form of charts, graphics or 3DMaps as one shown in Figure 1 above. The technical aspects of how the end-to-end process interacts with the SEIF to access which data, how data transformations and data mining algorithms are used as well as the interaction aspects between visualization tools and the data mining process are hidden for the Decision Maker and stay in the responsibility of other technology focussed user roles, i.e. Data Mining Expert, Ontology Expert and Data Mining Specialist. Therefore, at this point a Decision Maker can benefit from the simplicity of the process as well as from the expressivity of the visualized results.

All the tools developed and presented in this report have been conceived to support this workflow and are integrated in the GUI built upon the 3DMaps platform created by Agency9.

[2] RapidMiner is a graphical tool for data mining developed at the Technical University of Dortmund, supported by Rapid-I GmbH and previously known under the name YALE (Yet Another Learning Environment).

4 TOOLS FOR ENERGY ANALYSIS

4.1 Survey of the developed and integrated tools

In Task 5.2 the following components have been developed. With the exception of the first component in the list below, which is essentially a business process model, the other components are software products created for this project. Some of these products like the Web-based GUI interact with the user directly, and some of them provide background services that can only be invoked by the user by means of interactive components. Furthermore, we distinguish between components 100% developed for SEMANCO project and those developed by third parties and integrated into SEMANCO platform. The latter are listed at the end of this Section (4.1).

- ***Data analysis focussed workflow*** that has been described in Section 3.
- ***Web based GUI for workflow support*** (Section 4.2)
    - guiding communication between different user roles within the workflow,
    - accessing and invoking web-services that implement/support single steps of the workflow,
    - extending the SEMANCO platform been developed as a result of previous deliverables.
- ***SEMANA a set of web-services*** supporting the data analysis focussed workflow (APPENDIX A. SEMANA, a set of Web Services)
    - including services for integration of external tools and libraries such as RapidMiner/ RapidAnalytics software, Mondrian and JFreeChart library for data visualization,
    - including services implementing internal functions such as flow control, status management and query generation.
- ***Tool for automated generation of SPARQL queries*** for data retrieval via SEIF (Section 4.5.1)
    - using data semantics expressed through the SEMANCO ontology developed in Task 4.2,
    - simplifying the data retrieval for the end user who operates on the basis of requirements formulated in the context of the decision making process in the urban planning,
    - exposing a web interfaces and
    - making use of dedicated SEMANA services.
- ***Tool for querying RDF data stores for the purposes of data analysis*** (Section 4.5.2)
    - implementing methods for retrieving RDF data and its transformation into a fact table representation,
    - facilitating statistical methods for analysis and visualization of the retrieved data,
    - implementing an environment for data mining and machine learning, distributed under the AGPL open source license, as an operator/extension of RapidMiner software.

- ***JavaScript library for visualization of data in a web browser*** (Section 4.6)
    - Interacting with 3DMaps - the tools developed by Agency9 and used as a core presentation tool for the SEMANCO environment – to highlight graphically particular 3d objects, e.g. houses, districts or roads.
    - Integrating external visualisation tools Mondrian and JFreeChart, facilitating visualization of data plots like charts, pie diagrams, scatterplots, etc.
    - Supporting data transfer in different formats, i.e. json, xml, comma separated values (csv) and plain text.
- **Three sample RapidMiner processes** one for each case study. The goals and results of these processes are described in introduction part, in Sections 2.2-2.4. The process related to the Copenhagen case study is described in details in section 4.3.1.

Besides, external tools have been integrated with other tools developed for the project:

- **Data mining tools RapidMiner and RapidAnalytics** (Section 4.3 and 4.4)
    - RapidMiner facilitates the design of data mining processes, their running and the visualization of their results. However, RapidMiner is a standalone tool, e.g. it should be installed on a personal computer for a single user. Furthermore, it cannot be simply integrated in any environment developed by the third party.
    - RapidAnalytics enables the declaration of data mining processes developed using RapidMiner as services which can in turn be comparatively easily integrated into environments developed by third parties.
    - The integration of the RapidAnalytics services has been done through their interaction with the SEMANA services that are integrated as part of the SEMANCO environment.
- **Data visualization tools Mondrian and JFreeChart** (Section 4.6)
    - Both tools implement the graphic representation of data, e.g. scatter plot matrix, Scatterplot matrix or Parallel Coordinate Matrix, pie or column chart, etc.

4.2 Web-based GUI for workflow support

The GUI is an integrated part of the SEMANCO platform and is a facility aiming at supporting and optimizing the data analysis focussed workflow described in Section 3. The aspects of usability, ergonomics, efficiency of interaction and self-evidence of graphical representation have been placed in the focus of the GUI design. The GUI has been implemented using PHP technology. Furthermore, PHP scripts exploit the web services of the SEMANA library described in Section 4.7.

We will illustrate the GUI's functionality using as an example the Newcastle case study previously presented in Section 2.

As already mentioned, the workflow starts with the request for data analysis formulated by the Decision Maker role. For instance, in the case study of Newcastle an urban planner playing this role would deal with buildings that may need refurbishment. One of significant indicators of the need for refurbishment is the SAP rate. However, if only a small number of reliably estimated SAP rates are available then the planner will have to estimate the missing information. The question is how reliable such a prediction might be. In this situation, the urban planner can formulate a request using the web-based GUI in these terms: “I need a reliable data mining process for SAP rates prediction, including the demonstration of its reliability” (Figure 6).

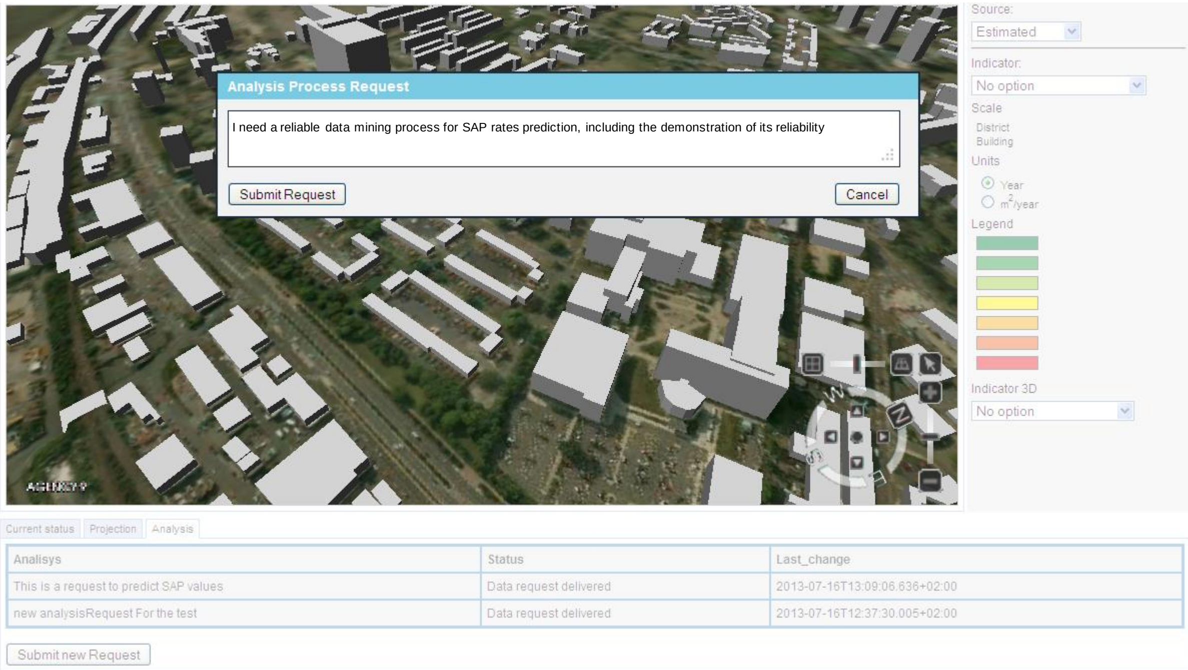
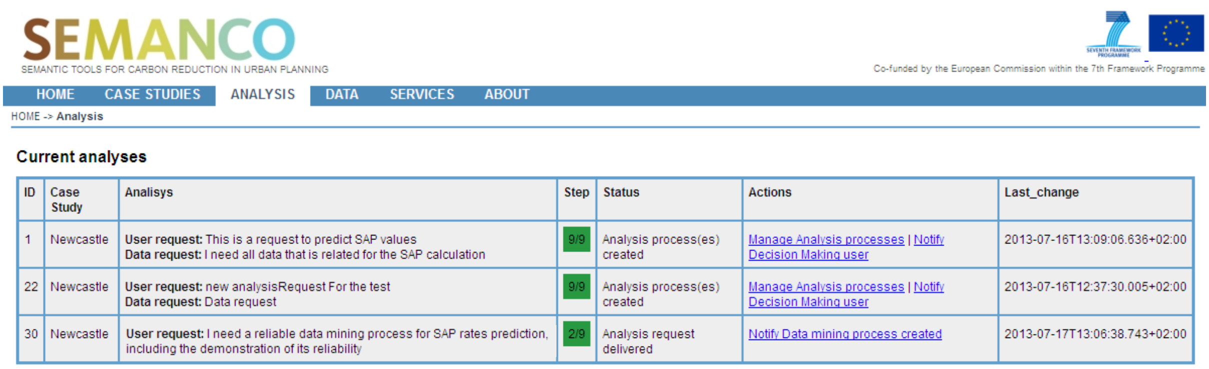


Figure 6. Web based GUI window containing the request of an urban planner

After the request has been formulated, it gets the status “2/9” – that means the second step of nine – and it becomes visible for users logged in the role Data Mining Expert. The corresponding view shown to these users is presented in Figure 7 below. The status is shown in the green box in the column Step (the line with ID 30 in Figure 7).

A Data Mining Expert investigates data required for the fulfilling of request and designs a provisional data mining process. After they have published the process as a service he/she formulates a request for the Ontology Expert in order to get access to the data managed by SEIF. The ontology Expert is shown both the initial request by the urban planner as well as the one just been formulated by the Data Mining Expert. The corresponding green boxes show the status of the requests (the lines with ID 1 and 22 in Figure 7).



ID	Case Study	Analisys	Step	Status	Actions	Last_change
1	Newcastle	**User request:** This is a request to predict SAP values **Data request:** I need all data that is related for the SAP calculation	9/9	Analysis process(es) created	Manage Analysis processes | Notify Decision Making user	2013-07-16T13:09:06.636+02:00
22	Newcastle	**User request:** new analysisRequest For the test **Data request:** Data request	9/9	Analysis process(es) created	Manage Analysis processes | Notify Decision Making user	2013-07-16T12:37:30.005+02:00
30	Newcastle	**User request:** I need a reliable data mining process for SAP rates prediction, including the demonstration of its reliability	2/9	Analysis request delivered	Notify Data mining process created	2013-07-17T13:06:38.743+02:00

Figure 7. Web based GUI window showing the status of different requests

Afterwards, an Ontology Expert generates a semantic query (Section 4.5.1) which can be processed by SEIF and submits the query as an answer of the Data Mining Expert’s request. The corresponding view visible for Data Mining Expert is shown in Figure 8 below. The green boxes show the status of the original query those changes with each next communication step. The Data Mining Expert integrates the query generated by the Ontology Expert into the data mining process as it is shown in Section 4.5.2.

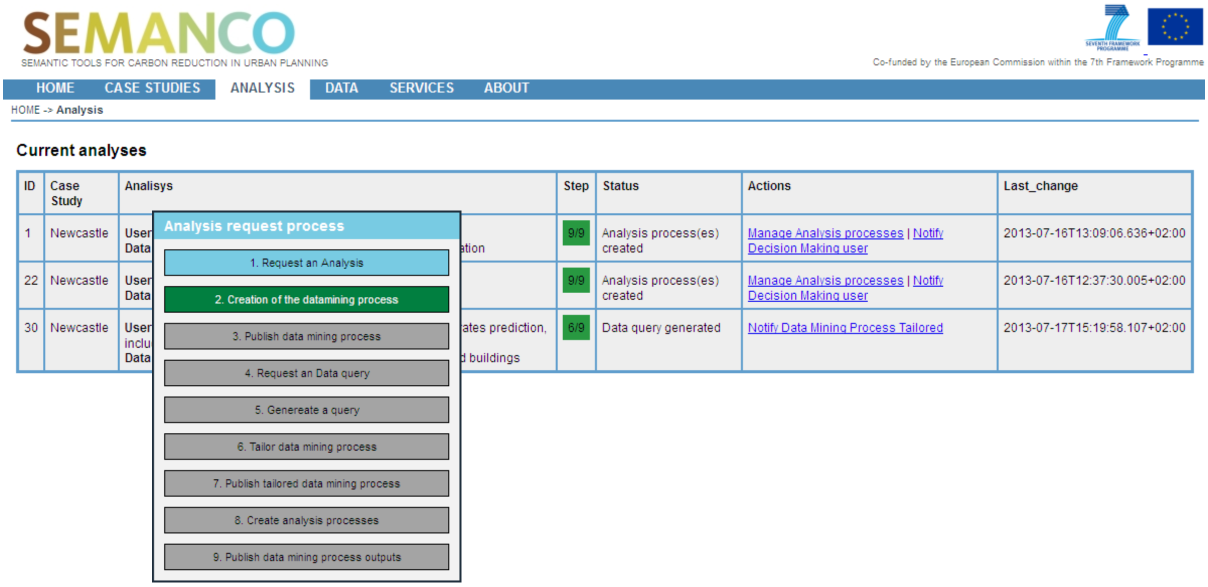


Figure 8. Web-based GUI window showing the view of the Data Mining Expert after a semantic query has been generated and submitted by Ontology Expert

Let us take that after a series of modifications the data mining process using data delivered by the query mentioned above generates satisfying results. Then, the Data Mining Expert sends a notification to the Data Mining Specialist on the User Side. The GUI in Figure 9 now displays the perspective of the Data Mining Specialist who receives the notification.

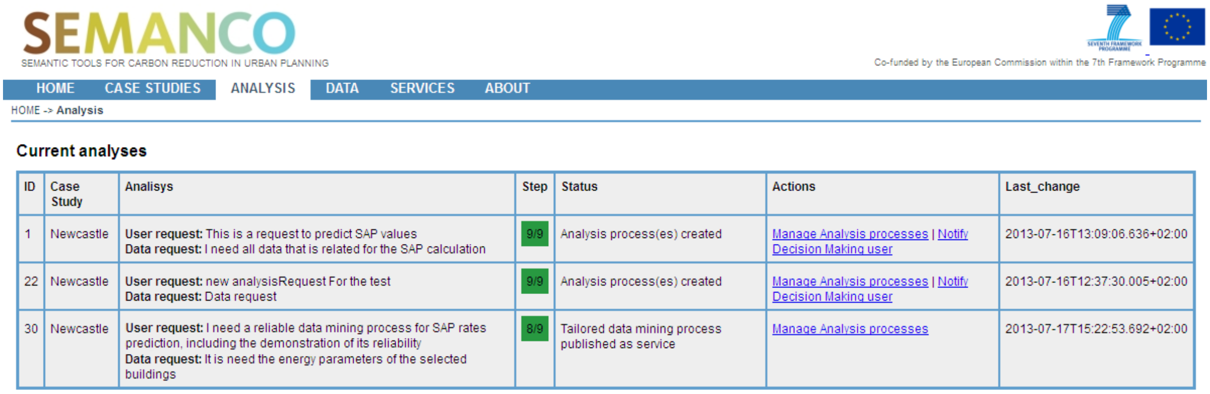


Figure 9. Web based GUI window showing the view of the Data Mining Specialist who has received the notification about a service published

At last the Data Mining Specialist submits the notification about the finalization of the request processing to the urban planner who is in the role of Decision Maker. The urban planner can see the analysis results as in Figure 10. He/she can observe the graphical representation, switch between different representations resulting from different analysis processes related to the request that has initiated the workflow. If appropriate, he/she can also experiment with the analysis processes by changing parameter values before taking a decision (for instance, to select one of many alternatives, to determine the volume of investment or plan a marketing campaign.).

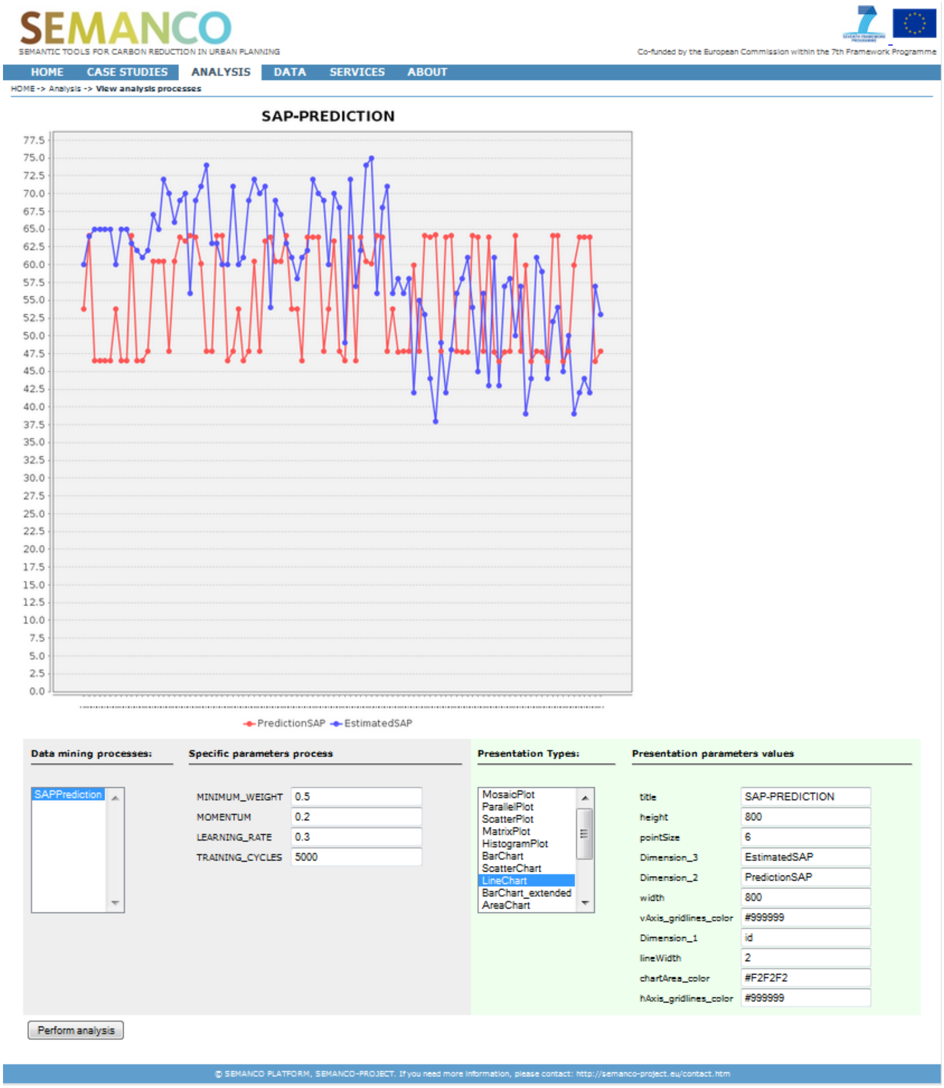


Figure 10. The Urban Planner is shown a graphical representation of the analysis results

It is important to emphasise that at each step of the workflow the actions of different roles, that is, the requests and responses, are always available to the roles responsible for the next step of the procedure. This way each user participating in the workflow can observe the history of the procedure in order to better understand its goals and specific details.

4.3 Data mining, process design and execution

In this section we will demonstrate how RapidMiner is applied within a data analysis focused workflow to design and execute particular data mining processes which address particular requirements of specific tasks, such as the use cases and activities shown in Section 2.

Numerous machine-learning algorithms applicable for data mining purposes are already available. Most of them have been implemented as software products, such as simple programme libraries or complex user-friendly data mining environments. One example of such an environment is the RapidMiner system, which incorporates hundreds of algorithms and data mining techniques. By its integration within the SEMANCO platform we have made all of the algorithms available.

A user in role of Data Mining Expert is expected to apply RapidMiner to design and execute the data mining processes. This is done in an interactive manner by dragging and dropping so-called operators that implement different techniques of data transformations and data mining algorithms onto the work desk and connecting them to flows (Figure 11). Four features of RapidMiner justify its choice:

- quality and comfort of GUI, facilitating design of arbitrary data mining processes,
- availability of numerous data mining algorithms and techniques, also developed by third parties, e.g. WEKA library[3],
- programme code open for extensions: one can develop his/her own RapidMiner operators and us them together with other operators in one data mining process,
- supplementing technology, i.e. RapidAnalytics, which enables export of data mining processes as web-services.

The extension of RapidMiner by an operator facilitating querying of RDF data and integration of data mining services into the SEMANCO platform will be demonstrated in Sections 4.5 and 4.4 correspondingly.

In the following section we will demonstrate on an example related to the case study Copenhagen how a RapidMiner process can be designed.

4.3.1 Case Study Copenhagen: Predicting energy consumption

Having received the request initiated by the urban planner, a Data Mining Expert begins with the request processing. He/she analyses the available data and retrieves a sample data set. Though incomplete such a data set usually contains records representative for the intended purpose of analysis. An example of such dataset related to the case study Copenhagen and generated in the course of processing of urban planner's request shown in Figure 6 (*"I need an energy projection for all building use categories for the period of time till 2035."*) can look as shown in Table 1 below.

Table 1. An example data set, case study Copenhagen

To_ Year	Residential Building	Family House	Apartment Block	Office
1930	197	219	219	136
1950	206	227	178	136
1960	193	214	174	136
1972	140	153	154	115
1978	119	127	136	115
1998	115	100	112	87
2007	96	82	90	87
2015		43	43	37,4
2020		25,5	25,5	18,1
2030		21,5	21,5	11,1
2035		21,5	21,5	11,1

As stated in Section 2.3, in the case of Copenhagen for all building categories statistical data from the last 80 years is available. In addition, for three categories only projections of energy consumption until year 2035 exist. For other categories, these projections are missing. The energy consumption values for these categories are shown in the three last columns in the table above. For other building categories, only statistics for the past exist and no projections for the future have been made. One of such categories is called *Residential Building*. The

[3] http://www.cs.waikato.ac.nz/ml/weka/

energy consumption values for buildings of this category are listed in the greyed column. The first column of the table shows the last year of the age class that corresponds to the measured or projected values. In order to fulfil the request of the urban planner, energy consumption values predicted for the future for all categories of building use are needed. Such predictions can be done using a classification algorithm based on the three available predictions.

The proposed data mining process designed using RapidMiner uses linear regression to predict the values missing in the second column of the table above. The process flow (Figure 11) is designed as follows:

1. The data set is made available by the operator *Retrieve*.
2. The data set is copied two times by the operator *Multiply*.
3. The first copy undergoes filtering (*Filter Example Set*), whereby only first seven lines of the data set shown in Table 1 (the energy consumption statistics until 2007) are selected. This data is used as training set for the linear regression.
4. The learned model and the second copy of the original data set are used as input for the operator *Apply Model*. Here an additional table column is generated. It contains predicted energy consumption values for all age classes (from 1930 to 2035).
5. Due to some technical aspects, the following three operators are needed to select values predicted for years 2008-2035 and to prepare the merge (*Operator Append*) with statistics for the years 1930-2007 taken from the third copy of the original data set.
6. Finally, with the last operator the *To_Year* Column is specially labelled for presentation purposes.

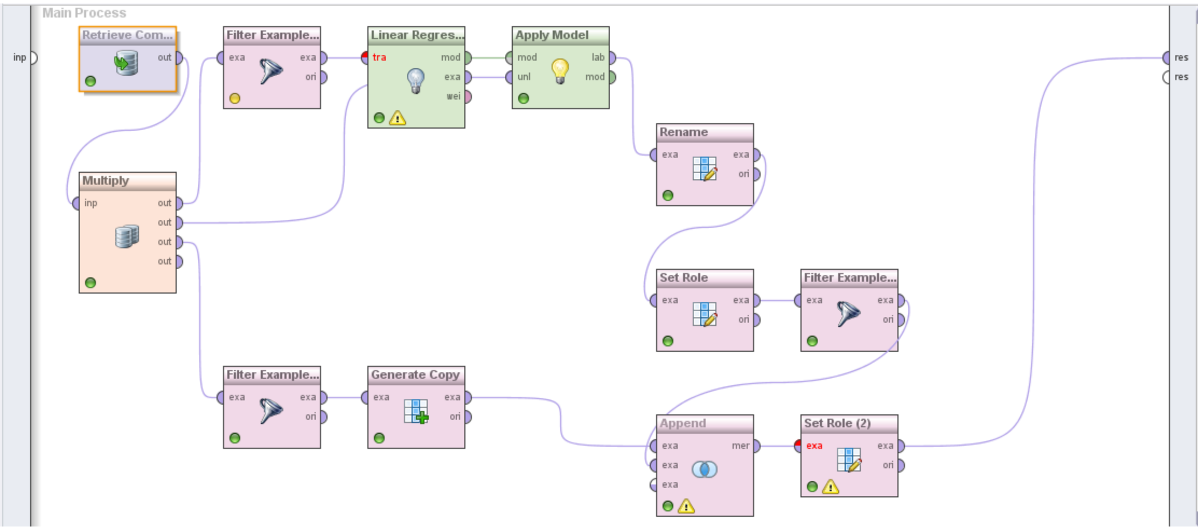


Figure 11. Workflow for prediction of energy consumption, case study Copenhagen

After the data mining process has terminated its results can be observed in the *results perspective* of RapidMiner (Figure 12). Visually, the result values seem to be reasonable. This hypothesis can be confirmed after a small modification of the process by adding the performance measurement operator right after the *Apply Model* operator.

In this particular case the root mean squared error measured by the formula

$$\sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}},$$

where x_i is a known value, y_i is a predicted values and *n* is total number of values, achieves the value of 0.713, which, if taking into account energy performance values between 25 and 250, is a really good result.

Figure 12 illustrates the prediction made using the same workflow for another building use category, i.e. *Garage*. The intuitively expected difference in energy consumption between a Garage and a Residential Building can be observed by comparison of Figure 3 and Figure 12.

Figure 12. Prediction of energy consumption for the building use category Garage

4.4 Publishing data mining process as a service

In order to facilitate access by other user roles to the data mining process such as the one shown in the previous section, a Data Mining Expert publishes the process as a service. To do so she/he applies RapidAnalytics, another tool supported by Rapid-I, the same provider as RapidMiner. RapidAnalytics offers a repository where all data sets can be analysed and all data mining processes developed by Data Miner Expert can be centrally stored. Furthermore, RapidAnalytics enables the publication of each process as a web service. When a process is published it becomes a URL which can be used by a web client. The analysis results can be returned in one of the common formats like XML, JSON, HTML, plain text and some others (Figure 13).

An important feature of RapidAnalytics is its ability to parameterise the service being published[4], that dependently on comprised algorithms/operators accepts as input parameters such as the number of iterations, min/max values and mathematical methods to calculate thresholds. Consequently, the values of these parameters can be set at the moment when the service is invoked. This feature is used in the SEMANCO platform to facilitate the User Side, basically a Data Mining Specialist, the possibility to experiment with the published processes.

[4] A detailed description of RapidAnalytics functionalities can be found at http://www.e-lico.eu/tl_files/elico/software/rapidanalytics/release/RapidAnalytics-Manual-1.1.014.pdf

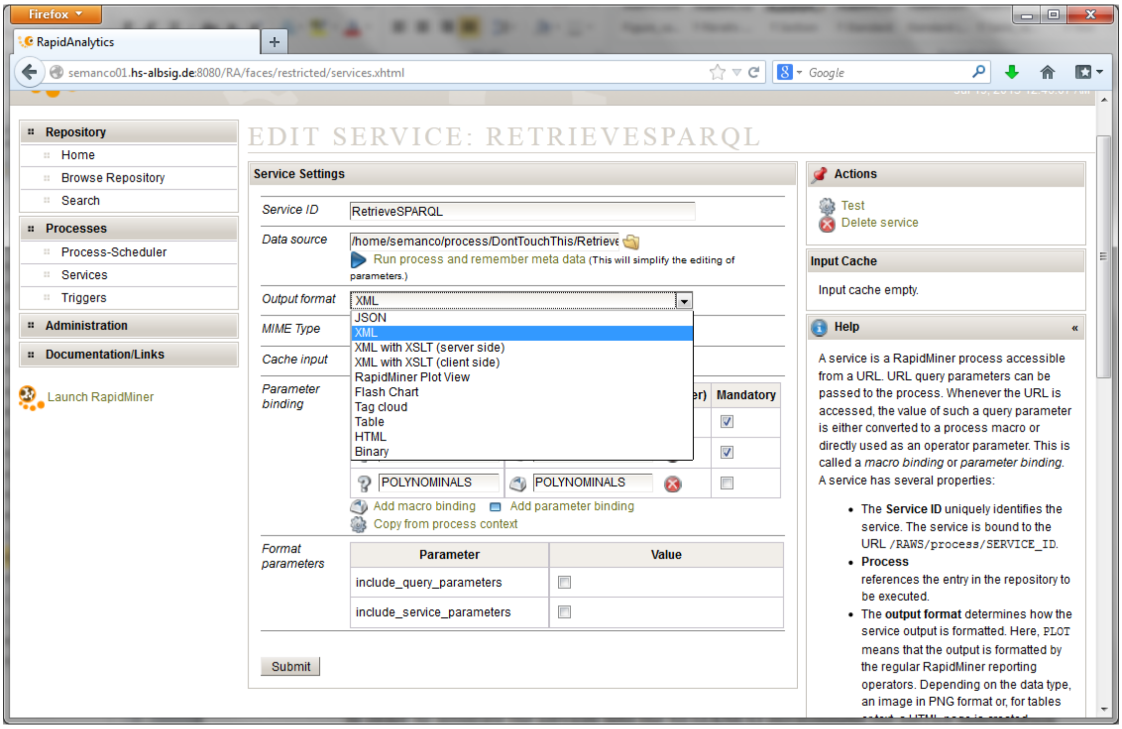


Figure 13. Publishing of a data mining process as a web service

In order to integrate the services into the SEMANCO platform, one of the SEMANA services (Section 4.7) looks up to the repository of RapidAnalytics, identifies the services published and presents them to the Data Mining Expert (Figure 14). They then select the service corresponding to the Decisions Maker's request which is being processed and associates the service to a specific visualization (as shown in Section 4.6).

Figure 14. selecting one of the data mining services by the Data Mining Expert role

After the publishing process is completed a data mining processes developed in RapidMiner and the corresponding visualization of analysis results are integrated in the SEMANCO platform as an *analysis process* that can be invoked and customized at the User Side (i.e. Data Mining Expert and Decision Maker) by changing parameter values. Often, it makes sense that one data mining process is associated with a set of different visualizations. In this case the data mining process/visualization pair appear in the SEMANCO platform as different analysis processes. In a rule a set of analysis processes is associated with the original request by decision maker.

4.5 Automated query generation and data access

Following the data analysis focused workflow, after having completed an interim data mining process, a Data Mining Expert posts a request for a complete data set. The complete data set contains all of the available data related to the current analysis task and it ensures the completeness of the data mining results.

After being notified, an Ontology expert analyses the request, the structure of the test data set contained in the interim data mining process and formulates a query aiming at retrieval of a complete data set requested by Data Mining Expert. For instance a SPARQL query used to retrieve data for the process related to the case study Copenhagen looks as follows:

```
Prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
Prefix sumo: <http://www.ontologyportal.org/SUMO.owl#>
Prefix Semanco: <http://www.semanco-project.eu/2012/5/SEMANCO.owl#>

SELECT ?Building ?Existing ?Refurbished ?building_UseValue ?toYearValue
?heat_DemandValue ?hotWater_DemandValue

WHERE {

     ?Building rdf:type sumo:Building .

     ?Building Semanco:hasBuilding_Use ?Building_Use .

     ?Building_Use Semanco:building_UseValue ?building_UseValue .

     ?Building Semanco:hasAge ?Age_Class .

     ?Age_Class Semanco:hasTo_Year ?To_Year .

     ?To_Year Semanco:toYearValue ?toYearValue .

     ?Building semanco:hasSpace ?Conditioned_Space .

     ?Conditioned_Space Semanco:hasHeat_Demand ?Heat_Demand .

     ?Heat_Demand semanco:heat_DemandValue ?heat_DemandValue .

     ?Conditioned_Space semanco:hasHotWater_Demand ?HotWater_Demand .

     ?HotWater_Demand semanco:hotWater_DemandValue ?hotWater_DemandValue .

     BIND(EXISTS{?Building semanco:hasConservation_State
     <http://arcdev.housing.salle.url.edu/semanco/repository/northharbourd2r/conse
     rvation_state/Existing>} AS ?Existing)

     BIND(NOT EXISTS{?Building semanco:hasConservation_State
     <http://arcdev.housing.salle.url.edu/semanco/repository/northharbourd2r/conse
     rvation_state/Existing>} AS ?Refurbished)

}
```

Our experience collected during elaboration of the tools has shown that in praxis such queries can be very long. For instance, building a comparatively small data set may require a query of almost 6000 characters such as the one shown in APPENDIX D. An example of a query generated by using of the shorts path calculation algorithm. The manual formulation and evaluation of queries of this length is extremely inefficient and time consuming. Therefore, a system to support this, i.e. to automate this process is strongly required.

4.5.1 Automation of Query Generation

To support an Ontology Expert in query generation two tools with web-based front ends have been developed.

The first tool allows an Ontology Expert to create a query in interactive manner (Nolle et al., 2013). A user starts by formulating a query by determining a concept. He/she types its name, for instance *Building* in the input box (Figure 15). For user support purposes, a suggestion feature has been implemented in form of a drop-down list which is activated while the user is typing a concept name. The drop-down list contains the ontology concept names that contain the text entered by user. A user can select one of the matches to retrieve the object properties which are shown in brackets. When the user selects a concept, its annotations properties (e.g. descriptions, authors, comments, and references) are shown. In turn, each object property of the selected concept can be selected to demonstrate the property's ranges (i.e. other ontology concepts). When selecting one of range concepts user is shown its object and data properties. For example, when the class *BuildingGeometry* is selected its properties are shown, among others NumberOfCompleteStoreys, *NumberOfRooms*, and *GroundFloor*. By clicking one of these object properties, their ranges and further the object properties of the range concepts, the user can navigate through the complete ontology.

Furthermore, a user can select data properties by activating the corresponding check boxes located right to the data property names. The selected data properties are the actual targets of the SPARQL query to be generated. The paths that user has been followed by selecting object properties and their ranges are recorded by the tool. In particular, this path is used to generate the SELECT and WHERE clauses of the SPARQL query. Finally, the generated query is submitted to the SPARQL endpoint to be evaluated, the retrieved data are shown to the user who basing on this view is able to modify the query and repeat its evaluation.

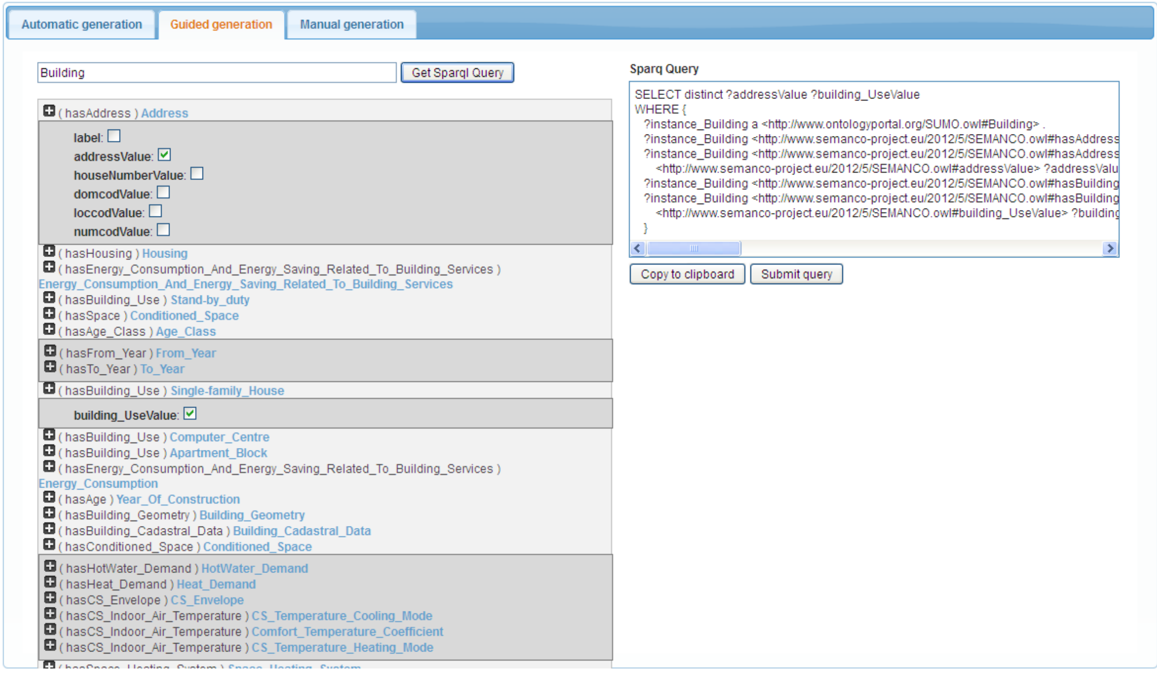


Figure 15. Web based tool for automation of Query Generation (1)

However, the nesting of properties the number of hierarchy levels that have to be taken into account in a query can be very high. In this case, the methodology described above that works excellently for comparatively compact relationship graphs can become awkward and uncomfortable. For such cases a second tool has been developed, that takes a completely different strategy: it requires as input only the names of concepts and the object and data properties that have been requested (used in the data interim mining process) by a Data Mining Expert. In the next step the tool invokes a service of the SEMANA framework

(APPENDIX A. SEMANA, a set of Web Services) to generate the paths between these concepts. These paths are converted in SPARQL and added to the query being generated.

In order to get reliable results at maximum efficiency the service mentioned above exploits the Dijkstra algorithm for calculation of the shortest paths in a graph. On this purpose, the OWL ontology used to evaluate the query is transformed into a proprietary data structure that represents an oriented graph. By doing so, in particular cases we have managed to reduce the time needed for the query generation from five hours when the *findShortestPath* method of Jena framework's class *OntTools*[5] has been used to ten seconds. The web-based interface of this tool is shown in Figure 16.

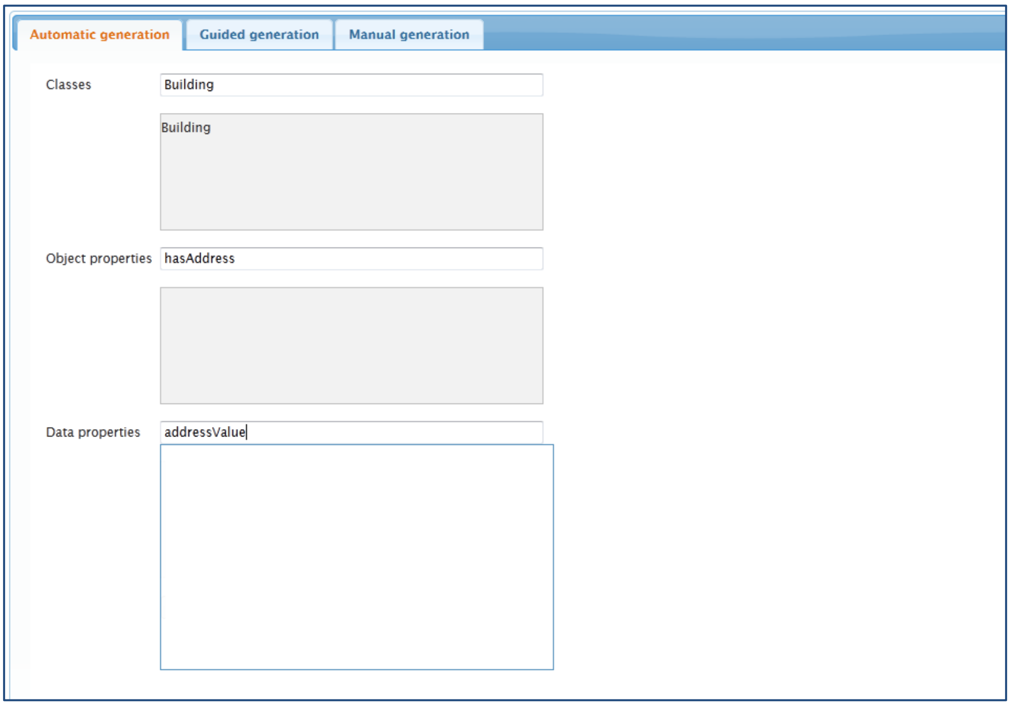


Figure 16. Web based tool for automation of Query Generation (2)

So far, the experience of applying both tools is insufficient to make a statement about their advantages or disadvantages. We hope that in further applications of the software in WP 8 will help to know which of the two tools for query generation is better and for which application cases.

4.5.2 Integration of queries in a data mining process

A query generated by an Ontology Expert is sent to a Data Mining Expert, who integrates it into the process previously developed. On the purpose of such integration, i.e. to facilitate querying of RDF data sources and use the retrieved data in data mining processes we have developed a new RapidMiner operator that is a de facto extension of the RapidMiner environment.

As with all other RapidMiner operators the operator can be found in the operators palette, dragged and dropped onto the working desk and connected by flow arrows with follower operators (as it is shown in Figure 11).

The operator has three parameters (Figure 17). The first parameter specifies the URL of the SPARQL endpoint to be queried. The second parameter contains the SPARQL query itself

[5] http://jena.apache.org/documentation/ javadoc/jena/com/hp/hpl/jena/ontology/OntTools.html

and the third one is to specify variables of the SPARQL query that have to be interpreted as a RapidMiner type polynomial, i.e. taking their values from a certain final set of values.

Figure 17. Input parameters for the RDF retrieval operator

To query more than one SPARQL endpoints a federation engine like ELITE (Nolle & Nemirovski, 2013) can be used. Another option for querying multiple data sources simultaneously is to use SPARQL 1.1's federated queries[6].

After a query has been evaluated by a SPARQL endpoint and a set of RDF triples, has been retrieved, it is transformed into the internal representation of RapidMiner. The data type of each result variable is analysed and transformed into the corresponding RapidMiner type. For instance, Boolean variables are transformed into binomials and URIs of individuals are transformed into strings. Since polynomial variables cannot be identified based on the RDF data type analysis, additional information is required to solve this issue. To determine polynomial variables the operator provide a third parameter (Figure 17) by which the names of concepts can be set. To identify these concepts SEMANA analyses the semantics of data, i.e. particular conceptualization aspects expressed over the ontology's TBox and sets the parameter value on each RapidAnalytics service call. For this purpose, we have defined a special ontology concept that subsumes all concepts whose individual's URIs should be converted to polynomial variables.

The data retrieved after execution of the SPARQL query mentioned above and transformed into the internal representation of RapidMiner is shown in Figure 18 and Figure 19 demonstrates metadata which specifies each corresponding table column.

ExampleSet (98 examples, 1 special attribute, 5 regular attributes)

ID	Building	numberCSValue	numberOfRoomsValue	GFAreaValue	GFHeightValue
0	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/1	1	4	60	3
1	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/10	1	4	61	3
2	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/11	2	4	54	3.500
3	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/12	2	4	40	3
4	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/13	2	5	64	3
5	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/14	2	5	50	3.500
6	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/15	2	5	68	3
7	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/16	2	5	53	3.500
8	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/17	2	5	65	3.500
9	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/18	2	5	68	3
10	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/19	2	5	50	3.500
11	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/2	2	5	39	3
12	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/20	2	5	56	3.500
13	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/21	2	5	56	3.500
14	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/22	2	5	55	3
15	http://arcdev.housing.salle.url.edu/semanco/repository/sap/building/23	2	4	54	3.500

Figure 18. Retrieved data transformed in RapidMiner representation format

[6] http://www.w3.org/TR/2013/REC-sparql11-federated-query-20130321/

ExampleSet (98 examples, 1 special attribute, 5 regular attributes)

Role	Name	Type	Missings
id	ID	integer	0
regular	Building	text	0
regular	numberCSValue	integer	0
regular	numberOfRoomsValue	integer	0
regular	GFAreaValue	real	0
regular	GFHeightValue	real	0

Figure 19. Meta data of the retrieved data transformed in RapidMiner representation format

After being transformed into the internal representation of RapidMiner, the data can be further processed with conventional operators.

To avoid recurrent execution of SPARQL queries, which may be time consuming, we store the retrieved data as a temporary data mart in a relational database (Figure 20, Part A). By this means, the data retrieved from the RDF sources can be efficiently reloaded and processed (Figure 20, Part B).

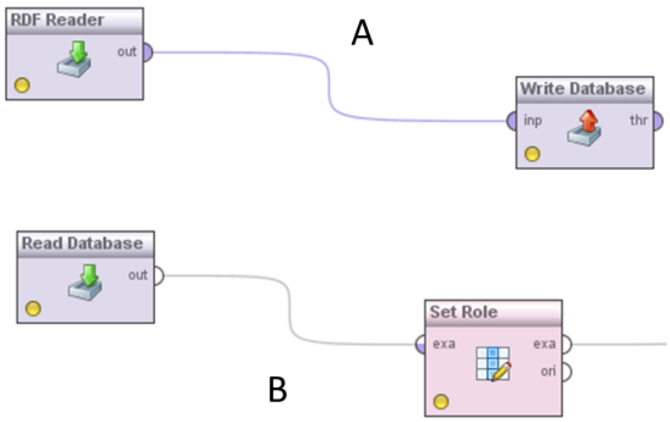


Figure 20. RapidMiner processes to retrieve RDF data and store it temporarily (A), RapidMiner processes for reloading the temporarily stored data (B)

4.6 Data visualization

According to the data analysis focussed workflow, after the development of the data mining process and design of the SPARQL query for data retrieval are completed by the Consultant Side the access to the process is granted to the User Side, in particular to the Data Mining Specialist. The latter prepares, that is, parameterizes or customizes the process, and configures its presentation settings in order to simplify the task of Decision Maker who originally initiated the current workflow.

While a Data Mining Expert (Consultant Side) uses RapidMiner representation of analysis results, a Data Mining Specialist (User Side) deals with the representation integrated in the SEMANCO environment. Only this representation is available to a Decision Maker who customizes the data analysis results working in the integrated platform front-end. Preparing this representation is the task of the Data Mining Expert (Consultant Side). It is in fact the last activity carried out before granting the access to Data Mining Specialist (User Side).

To facilitate the presentation of data visualizations within the SEMANCO platform two data visualization systems, i.e. JFreeChart[7] and Mondrian[8] have been integrated. A JavaScript API

[7] http://www.jfree.org/jfreechart

[8] http://stats.math.uni-augsburg.de/mondrian

has been created to generate image based Portable Network Graphics (PNG) or Scalable Vector Graphics (SVG) charts enhanced by interactive feedback in the form of tooltips. By means of this API, charts can be integrated into HTML code and shown in a browser which facilitates their use in the SEMANCO platform. The interactive feedbacks is implemented by means of HTML ImageMaps.

The following chart types are supported:

- MosaicPlot
- ParallelPlot
- ScatterPlot
- MatrixPlot
- HistogramPlot
- BarChart
- ScatterChart
- LineChart
- BarChart
- AreaChart
- PieChart

Figure 21 shows four examples of data visualization charts available in SEMANCO platform.

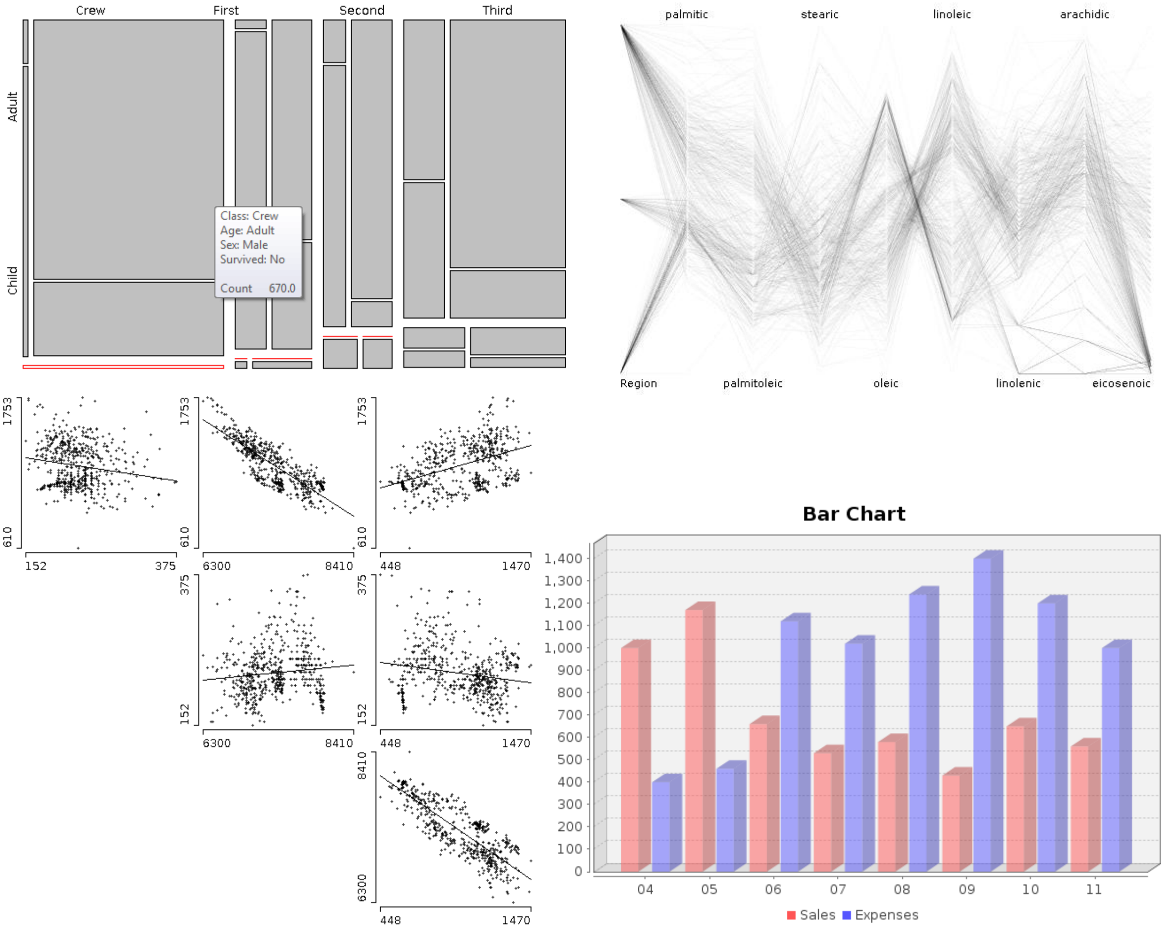


Figure 21. Example of data visualization charts integrated into SEMANCO platform (Mosaic Plot, Scatter, parallel plot, Plot Matrix and Bar Chart)

Furthermore the following code determines the chart selection and “writes” the generated image into the HTML code:

```
var chart = new a9.chart.Chart("MosaicPlot");
…
chart.draw(function(imageUrl, imageMap, imageMapId) {
  document.write(imageMap);
  document.write("<img src='" + imageUrl + "' usemap='#" + imageMapId + "'
/>");
}, function(errorMsg) {
  alert(errorMsg);
});
```

Data access can be enabled either by the specification of a web service that delivers data, i.e. service URI and the format of data delivered by the service (Ascii, Excel, XML or JSON):

```
chart.setDataService('http://10.7.1.102:8080/mondrian/test/Titanic.txt',
'Ascii');
```

or by declaration in the JavaScript code an array containing data to be visualized:

```
var chartData = new a9.chart.DataTable([
  ['Class','Age','Sex','Survived'],
  ['First','Adult','Male','Yes'],
  ['First','Adult','Male','Yes']
]);
```

In the case of SEMANCO platform the former method (web service specification) is used. One of the services published by means of RapidAnalytics software is used as a data source. The association of a service with a visualization method is done by Data Mining Expert in interaction with the platform. Figure 22 illustrates the process. A Data Mining Expert selects a service (left part of the figure) and then defines the parameters required for this type of chart (right part). After the “save” button (left bottom corner) is clicked the corresponding analysis process comprising a data mining process published as a service and its visualization becomes available to the User Side, i.e. to the Data Mining Specialist.

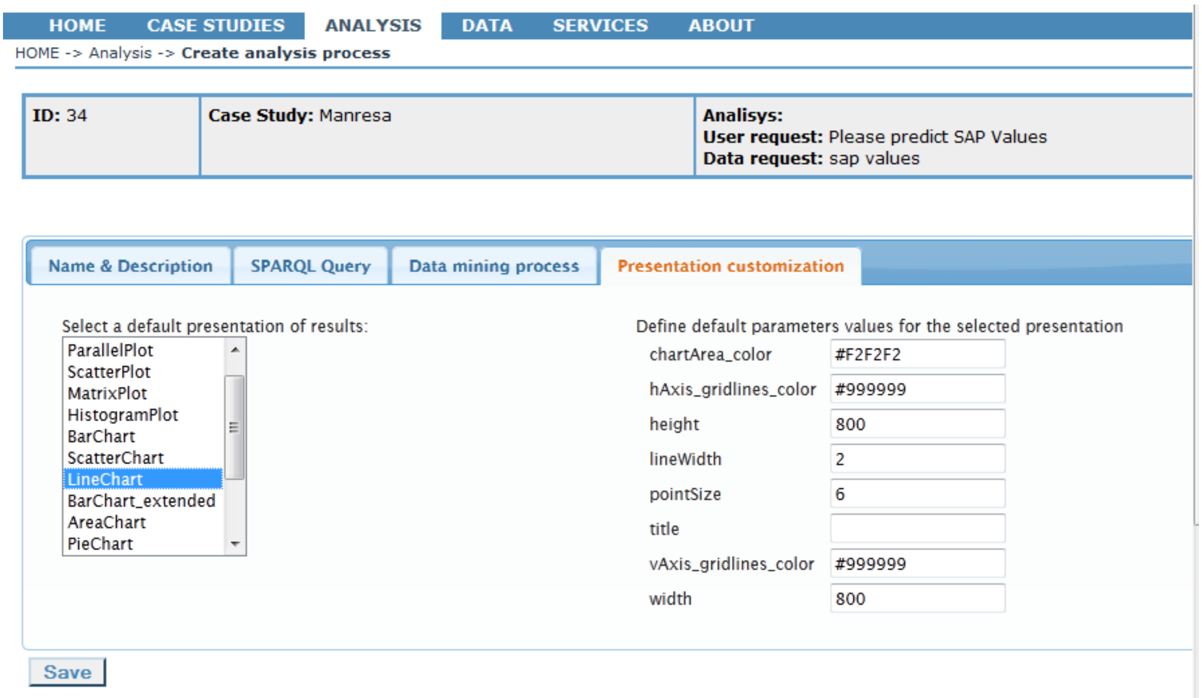


Figure 22. Selection of a visualization chart for a data mining process

One service can be associated with different visualization charts. Such associations are available to a Data Mining Specialist as different analysis processes associated with the same request originally formulated by Decision Maker. The user in this role can i) perform the process, ii) see its visualized results, iii) modify parameters values or delete the process.

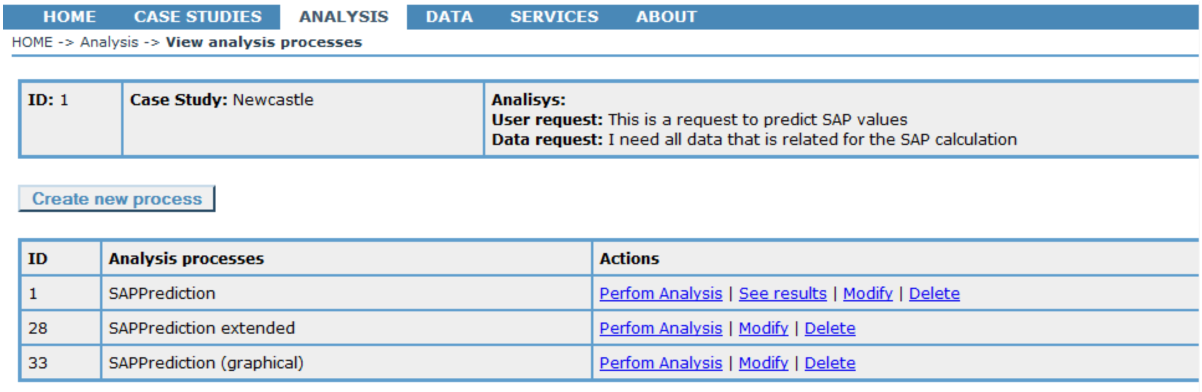


Figure 23. Interface of the Data Mining Specialist

4.7 SEMANA services

SEMANA is a set of web-services that has been developed to support the data analysis focussed workflow. SEMANA services can be subdivided in i) services that can be invoked at different stages of the data analysis focussed workflow and serve to the aim of immediate support for this workflow, and ii) additional services which theoretically can be invoked at any stage of the workflow.

The descriptions of services are listed in the APPENDIX A. SEMANA, a set of Web Services.

5 Conclusions

5.1 Contribution to overall picture

The tools for energy analysis developed in Task 5.2 fulfil the requirements identified in WP 6, in particular by addressing the specifications as described in the use cases related to three case studies (Newcastle, Copenhagen, Manresa). The tools use data sources specified and collected in WP 3 and accessed over Semantic Energy Information Framework (SEIF) developed in WP 4. The evaluation of the tools is carried out in WP 8.

Together with other tools developed in WP 5, the energy analysis tools make one of the main contributions of the SEMANCO project. However, in contrast to other tools developed in WP 5 the energy analysis tools do not use exact calculations. Instead, they exploit probabilistic algorithms and therefore should be applied in settings with lack of data or in tasks requiring the exploration of "hidden" relationships or dependencies.

In particular, the developed tools fulfil the following three tasks: i) data retrieval and systematization, ii) statistical data analysis and iii) visualization of data and analysis results.

The following components have been developed in connection with this deliverable:

- A data analysis workflow (Section 3)
- A web-based GUI for workflow support (Section 4.2)
- A tool for querying RDF data stores for the purposes of data analysis (Section 4.5.2)
- A tool for automated generation of SPARQL queries (Section 4.5.1)
- A set of services for visualization of data results (Section 4.6)
- A SEMANA a set of web-services (Section 4.7 and APPENDIX A. SEMANA, a set of Web Services)

In Task 5.2. the following tools developed by third parties has been integrated

- Data mining tools RapidMiner and RapidAnalytics
- Data visualization systems Mondrian and JFreeChart

While the workflow determines the overall data analysis process involving four different user roles and serves as a basis for the technical and business-administrative integration of tools and people into particular analysis tasks, all other components are of technical character and have been integrated into the web based SEMANCO platform.

5.2 Impact on other WPs and Tasks

The energy analysis tools will be integrated in the SEMANCO platform under development in Task 5.4. Their integration will be based on the workflow, which defines the way that the tools will be used in the platform by different users.

The data mining processes described in Sections 2.2-2.4 can be used to automate some of the tasks performed by the tools developed in Task 5.1. For instance, instead of calculating the energy efficiency of a set of houses using the tools developed in Task 5.1, it would be possible to predict their performance using less input parameters with the data mining tools presented in this report.

Furthermore, we believe that application of analysis results obtained from three demonstrator data mining processes (Sections 2.2-2.4) and form further processes that may be developed for particular tasks formulated from the context of Task 8.3 will contribute to further elaboration of case studies and use cases presented in Deliverable 6.1.

5.3 Contribution to demonstrations

Currently the energy analysis tools integrated into SEMANCO platform aren´t available online and can´t be used by project partners and associated organizations for demonstration purposes. This will be done in the second iteration of the implementation of the demonstration scenarios to be carried out in Task 8.3.

Three data mining processes, each for one case study, have been implemented for demonstration purposes.

5.4 Other conclusions and lessons learned

The tools that have been developed facilitate data mining using with semantic data in the urban planning domain. The experiments conducted with data of the three case studies have shown the usefulness and high potential of the tools. Sophisticated querying of RDF triple stores and data visualization techniques have been developed.

However classical business intelligence techniques like star schema support, data cube/data mart operations like slicing and dicing as well as the well-known problems of data warehouses, such as summarizability have not been addressed yet. Work in this direction has started and it will be developed further during the project.

6 References

Crosbie, T. & Crilly, M. (2013) *SEMANCO. Deliverable 6.1: Stakeholder requirements analysis*. Retrieved July 30, 2013 in Web site: http://semanco-project.eu/index_htm_files/SEMANCO_D6.1_20130430.pdf

Madrazo, L., Gamboa, G., Nemirovskij, G., Sicilia, A. & Crosbie, T. (2012) *SEMANCO. Deliverable 1.8: Project Methodology*. Retrieved July 30, 2013 in Web site: http://semanco-project.eu/index_htm_files/SEMANCO_D1.8_20120921.pdf

Martens, H. & Dardenne, P. (1998) *Validation and verification of regression in small data sets*. Chemom. Intell. Lab. Syst., 1998, 44 (1-2), pp. 99-122.

Nolle, A., Nemirovski, G., Sicilia, A. & Pleguezuelos, J. (2013) *An Approach for Accessing Linked Open Data for Data Mining Purposes*: In Proceedings of RapidMiner Community Meeting and Conference (RCOMM 2013), Porto.

Nolle, A. & Nemirovski, G. (2013) *ELITE: An Entailment-based Federated Query Engine for Complete and Transparent Semantic Data Integration*: In Proceedings of International Workshop on Description Logic 2013, Ulm.

Sicilia, A. (2013) *SEMANCO. Deliverable 3.4: Ontology repository with migrated data*. Retrieved July 30, 2013 in Web site: http://semanco-project.eu/index_htm_files/SEMANCO_D3.4_20130430.pdf

7 APPENDICES

APPENDIX A. SEMANA, a set of Web Services

SEMANA contains the following services that are ordered to the workflow statuses in which they can be invoked.

1. ***Status: Analysis request delivered***

   At this stage a Decision Maker creates a new Analysis request by formulating her/his intentions as plain text associated with a project, location or case study, for which this analysis request is related. At this stage, the following service is available:

   - *AnalysisRequestService.CreateAnalysisRequest* takes a string for the formulated analysis request as well as a string for the related project (case study) and returns the id of the generated analysis request.

2. ***Status: Data mining process created***

   A Data Mining Expert observes all analysis requests in an overview table.

   - *AnalysisRequestService.GetAllAnalysisRequestsForDME* returns all analysis requests that can be changed by a DME (Data Mining Expert)

   After she/he has created a data mining process for a request, the user can change the status of this request accordingly.

   - *AnalysisRequestService.NotifyDataMiningProcessCreatedFor* takes an analysis request id and changes the status of this request

3. ***Data mining process published as service***

   A Data Mining Expert observes analysis requests that should be settled by her/his role in an overview table.

   - *AnalysisRequestService.GetAllAnalysisRequestsForDME* see above

   After she/he has published a data mining process for a request as a web service, she/he can change the status of this request accordingly.

   - *AnalysisRequestService.NotifyDataMiningProcessPublishedFor* takes an analysis request id as input and changes the status of this request accordingly

4. ***Status: Data request delivered***

   After a data mining process is published as a web service a Data Mining Expert is able to formulate a data request as plain text.

   - *AnalysisRequestService.CreateDataRequestFor* takes the id of an analysis request as input and a string for the formulated data request and changes the status of this request accordingly

5. ***Status: Data query generated***

   An Ontology Expert can find all analysis requests with data requests that should be settled by her/his role in an overview table.

   - *AnalysisRequestService.GetAllAnalysisRequestsForOE* returns all analysis/data requests that can be processed by an OE (Ontology Expert). An OE generates a SPARQL query by using the existing query generation service or manually.

- *AnalysisRequestService.SetSPARQLQueryFor* takes the id of an analysis request and a string for the formulated SPARQL query and changes the status of this request accordingly. After an OE has generated a SPARQL query according to the request, the status of the request will be changed accordingly.
- *AnalysisRequestService.GenerateSPARQLQueryFor* takes the id of an analysis request, selected classes, selected object properties, selected data properties, query conditions, query select extension, query where extension and query post extension and returns a task id that enables to monitor the automated query generation
- *AnalysisRequestService.GetResultForGeneratedSPARQLQuery* takes the task id of the query generation process, changes the status of the corresponding analysis request accordingly and returns the generated SPARQL query.

After a data request is posted or a query is generated a Data Mining Expert is able to change the existing data request using the following operation:

- *AnalysisRequestService.ChangeDataRequestFor* takes the id of an analysis request and a string for the (re)formulated data request and changes the status of this request back to “Data request delivered” accordingly

6. ***Status: Data mining process tailored to delivered data:***

A Data Mining Expert can observe all analysis requests that should be processed by this role in an overview table.

- *AnalysisRequestService.GetAllAnalysisRequestsForDME* (see above)

After a Data Mining Expert has tailored a data mining process with respect to the delivered results of the SPARQL query, she/he can change the status of this request accordingly:

- *AnalysisRequestService.NotifyDataMiningProcessTailoredFor* takes an analysis request id and changes the status of this request accordingly

7. ***Status: Tailored data mining process published as service***

A Data Mining Expert can find all analysis requests that should be processed by this role in an overview table.

- *AnalysisRequestService.GetAllAnalysisRequestsForDME* (see above)

After a Data Mining Expert has published the tailored data mining process as a web service, she/he can change the status of this request accordingly.

- *AnalysisRequestService.NotifyTailoredDataMiningProcessPublishedFor* takes an analysis request id and changes the status of this request accordingly

8. ***Status: Analysis process created***

After a tailored data mining process is published as a web service a Data Mining Expert is able to generate new analysis processes for an analysis request by selecting the defined data mining process and by determining the type of visualization chart as well as the default values for all possible parameters (presentation parameters and data mining process parameters).

- *AnalysisService.CreateAnalysis* takes a name, a description, an analysis request id, a RapidAnalysis service name, additional parameter values, a presentation, presentation parameter values, condition class which is the ontology class used to formulate the conditioned where extension, which is a SPARQL code snipped to define e.g. the selected area or selected buildings, as input and changes the status of the given analysis request accordingly and returns the id of the generated analysis process.

9. ***Status: Analysis evaluated***

After an analysis process is generated, the user in role Data Mining Specialist (User Side) can find all analysis processes that should be modified or interpreted by her/his role in an overview.

- *AnalysisRequestService.GetAllAnalysisRequestsForDMU* returns all analysis requests that can be changed by a DMS (Data Mining Specialist). She/he is able to change parameter values for the presentation parameters but also for the data mining parameters for each analysis process assigned to the current analysis request.
- *AnalysisService.PerformAnalysis* takes an analysis process id that should be performed with the parameters already defined for this process and returns a task id that enables to monitor the execution of this analysis process
- *AnalysisRequestService.GetResultForPerformedAnalysis* takes the task id of a performed analysis process as input and returns the generated code snippet (result) of this process.
- *AnalysisService.performCustomizedAnalysis* takes an analysis object to perform an analysis process with customized values as input and returns a task id that enables to monitor the execution of this customized analysis process
- *AnalysisService.getResultsForPerformedCustomizedAnalysis* takes the task id of a performed customized analysis process as input and returns the generated code snippet (result) of this process.

The last to services enables the user to play with process parameters without modifying the stored parameter values. If she/he finds a good constellation she/he can save the changed parameter values for this process.

- *AnalysisService.UpdateAnalysis* takes an analysis object with customized parameter values as input and assigns those to an analysis process of an analysis request.
- *AnalysisService.CopyAnalysis* takes an analysis id as input and returns the id of the new copy of the given analysis process. The user is able to copy the current analysis process and change the parameters in this copy.

After saving one or more analysis processes of an analysis request with customized parameter values she/he can also formulate any additional evaluation results as plain text.

- *AnalysisRequestService.CreateEvaluationResultsFor* takes the id of an analysis request as input and a string for the formulated evaluation results to an analysis request and changes the status of this request accordingly

After an analysis request is evaluated the user (Decision Maker) is able to read any defined plain texts and to 're-execute' all analysis processes assigned to this specific analysis request with the defined customized parameter values.

- *AnalysisRequestService.PerformEvaluatedAnalysesFor* takes an analysis request id for which all assigned analysis processes should be performed each with the customized parameters already defined for this process and returns a task id that enables to monitor the execution of this analysis request process
- *AnalysisRequestService.GetResultsForPerformedEvaluatedAnalyses* takes the task id of a performed analysis request process and returns all generated code snippets (results) of all assigned analysis processes.

Additional services to administrate analysis requests:

- *AnalysisRequestService.DeleteAnalysisRequest* takes an analysis request id and returns the corresponding analysis request object
- *AnalysisRequestService.GetAnalysisRequestById* takes an analysis request id and deletes the corresponding analysis request
- *AnalysisRequestService.GetAllAnalysisRequests* returns all analysis request objects

Additional services to administrate analysis processes:

- *AnalysisService.DeleteAnalysis* takes an analysis process id and returns the corresponding analysis process object
- *AnalysisService.GetAnalysisById* takes an analysis process id and deletes the corresponding analysis process
- *AnalysisService.GetAllAnalyses* returns all analysis process objects

Additional services to search ontology objects used for the query generation:

- *OntologyService.SearchClasses* takes a search string and returns all classes that contains this search string
- *OntologyService.SearchObjectProperties* takes a search string and returns all object properties that contains this search string
- *OntologyService.SearchDatatypeProperties* takes a search string and returns all datatype properties that contains this search string

Additional services for RapidAnalysis services:

- *RAAnalysisService.GetAllRAAnalysisServices* returns all available data mining processes that are published as a web service on RapidAnalytics and compatible with SEMANA
- *RAAnalysisService.GetRAAnalysisService* takes an id of an defined RapidAnalytics analysis service and returns the corresponding data object containing all specific information about that service

Additional services for presentations:

- *PresentationService.GetAllPresentations* returns all available presentations that can be selected for an analysis process and contains all presentation parameters as well as its value types
- *PresentationService.GetDefaultPresentationParamValues* returns all available presentation parameters for which a default value is defined

Additional services for running tasks

- *TaskService.GetTask* takes the id of a running task and returns the corresponding object containing the current status of this task
- *TaskService.GetAllTasks* returns all running tasks and the corresponding information about that tasks

APPENDIX B. Agency9 Chart API Reference based on JFreeChart

Agency9 Chart API is a tool for generating image based (PNG) or SVG charts from your data using JavaScript. User interactive feedback is enabled in the form of tooltips using HTML ImageMaps.

Setup

To start using the API include these two JavaScript libraries in your document header using:

```
<script type="text/javascript" src="http://haleh.agency9.se:8080/chart/jquery-1.8.2.min.js"></script>
<script type="text/javascript" src="http://haleh.agency9.se:8080/chart/a9-chart-api.js"></script>
```

Then initialize the API by calling the a9.chart.setApiProvider function with the URL to the providing server:

```
a9.chart.setApiProvider('http://haleh.agency9.se:8080/chart/');
```

Following classes are available:

a9.chart.Chart

Constructor

```
a9.chart.Chart(chartType)
```

Creates a new chart.

chartType: String that specifies the type of chart to create. For a list of valid chart types.

Methods

```
setChartType(chartType)
```

Set the type of chart to be generated.

chartType: String that specifies the type of the chart.

```
setDataTable(dataTable)
```

Set the data to be visualized by this chart.

dataTable: Instance of DataTable that includes the data to be visualized.

```
setDataService(serviceUrl, dataFormat)
```

An alternative to setDataTable. By setting a DataService data can be loaded from an external URL if the format is supported.

serviceUrl: The external URL to the data source.

dataFormat: The data format of the source. Supported formats are:

- 'Ascii'
- 'Excel'
- 'RapidAnalytics'

```
setOptions(options)
```

Specify the options parameters to use for chart generation.

options: JSON object with options. For a list of available options see Options Parameters.

```
draw(callback, [onErrorCallback])
```

Generates the chart as a PNG image and responds with a callback to the user.

callback: Function to be called when the chart has been generated. It receives three parameters, the URL to the generated image, the HTML code for the image map and the identifier of the image map.

onErrorCallback: Optional function to be called if an error occurs. It receives one parameter, a string with the error message.

```
drawSVG(callback, [onErrorCallback])
```

Generates the chart as SVG and responds with a callback to the user.

callback: Function to be called when the chart has been generated. It receives one parameter, the generated SVG markup.

onErrorCallback: Optional function to be called if an error occurs. It receives one parameter, a string with the error message.

a9.chart.DataTabke

Constructor

```
a9.chart.DataTable([dataArray])
```

Creates a new DataTable populated with the content of dataArray.

dataArray: A multi-dimensional array of data. Each sub-array corresponds to one row in the table.

Methods

```
addColumn(type, label)
```

Adds a column to the DataTabel.

type: A string with the data type of the values of the column. Valid types are 'string', 'number', 'boolean', 'date', 'datetime' or 'timeofday'.

label: The name associated with this column.

```
addRow(rowData)
```

Adds a row of data to the DataTabel.

rowData: An array of values that each corresponds to one column in the DataTable.

Example Usage

Generate chart from static data

```
a9.chart.setApiProvider('http://10.7.1.102:8080/chart/');

var chartOptions = {
  title: 'Chart Title',
  width: 600,
  height: 400
};

var chartData = new a9.chart.DataTable([
  ['x', 'x^2', 'x^3'],
  [ 1,   1,   1],
  [ 2,   4,   8],
  [ 3,   9,   27]
]);

var chart = new a9.chart.Chart("ScatterChart");
chart.setDataTable(chartData);
chart.setOptions(chartOptions);

chart.draw(function(imageUrl, imageMap, imageMapId) {
```

```
  document.write(imageMap);
  document.write("<img src='" + imageUrl + "' usemap='" + imageMapId + "' >");
}, function(errorMsg) {
  alert(errorMsg);
});
```

Load data from providing service

Supported data fromats at the moment are 'Ascii', 'Excel' and 'RapidAnalytics'.

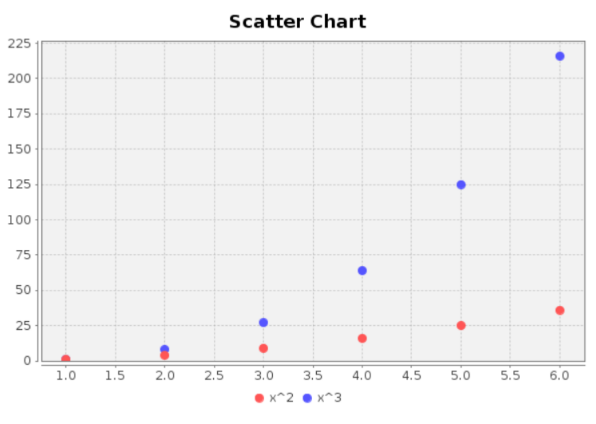
```
a9.chart.setApiProvider('http://10.7.1.102:8080/chart/');

var chartOptions = {
  title: 'Chart Title',
  width: 600,
  height: 400
};

var chart = new a9.chart.Chart("ScatterChart");
chart.setDataService('http://10.7.1.102:8080/chart/test/Data.txt', 'Ascii');
chart.setOptions(chartOptions);

chart.draw(function(imageUrl, imageMap, imageMapId) {
  document.write(imageMap);
  document.write("<img src='" + imageUrl + "' usemap='#" + imageMapId + "' />");
}, function(errorMsg) {
  alert(errorMsg);
});
```

Supported Chart Types

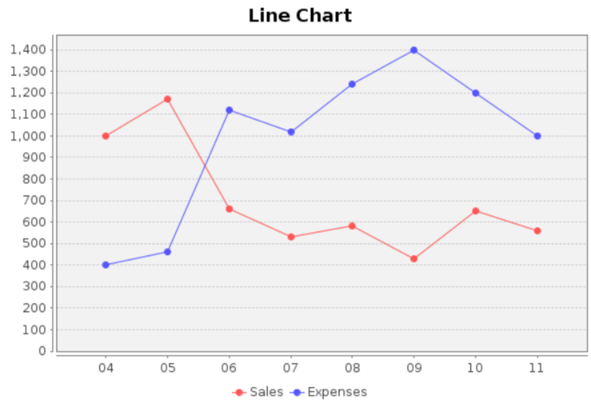
Supported values of chartType are:

- ScatterChart
- LineChart
- BarChart
- AreaChart
- PieChart

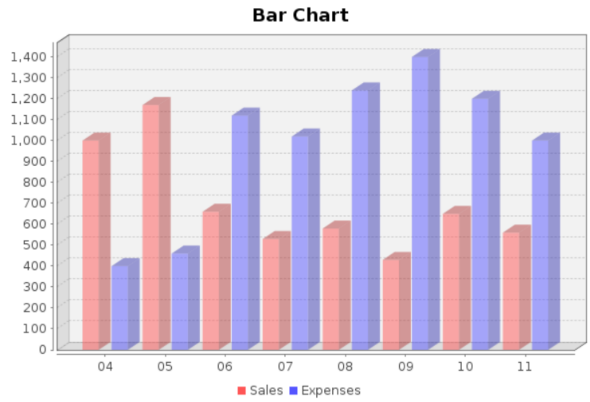
Demos



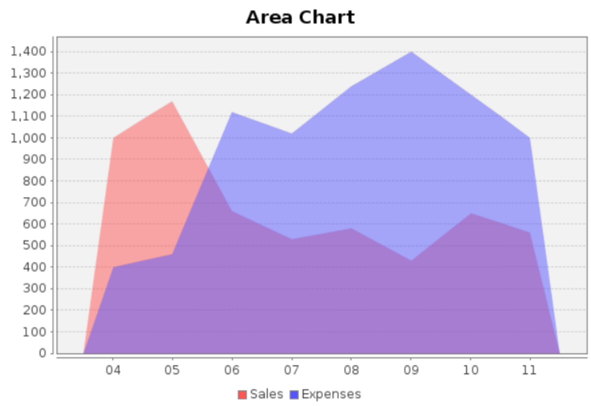

```
var chartOptions = {
  width: 500,
  height: 450,
  vAxis: { gridlines: { color: '#999999'
} },
  hAxis: { gridlines: { color: '#999999'
} },
  pointSize: 8,
  chartArea: { color: '#F2F2F2' }
};

var chartData = new a9.chart.DataTable([
  ['x', 'x^2', 'x^3'],
  [ 1,  1,         1],
  [ 2,  4,         8],
  [ 3,  9,        27],
  [ 4,  16,       64],
  [ 5,  25,      125],
  [ 6,  36,      216]
]);
```



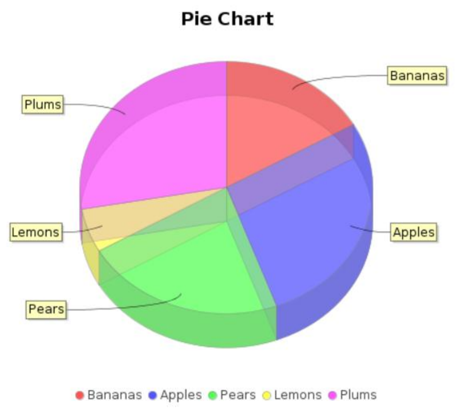

```
var chartOptions = {
 title: 'Line Chart',
 width: 500,
 height: 450,
 vAxis: { gridlines: { color: '#999999' }
},
 chartArea: { color: '#F2F2F2' },
 pointSize: 6
};

var chartData = new a9.chart.DataTable([
 ['Year', 'Sales', 'Expenses'],
 ['04', 1000,   400],
 ['05', 1170,   460],
 ['06', 660,    1120],
 ['07', 530,   1020],
 ['08', 580,   1240],
 ['09', 430,   1400],
 ['10', 650,   1200],
 ['11', 560,   1000]
]);
```




```
var chartOptions = {
 title: 'Bar Chart',
 width: 500,
 height: 450,
 vAxis: { gridlines: { color: '#999999' }
},
 chartArea: { color: '#F2F2F2' },
 pointSize: 6,
 is3D: true
};

var chartData = new a9.chart.DataTable([
 ['Year', 'Sales', 'Expenses'],
 ['04', 1000,   400],
 ['05', 1170,   460],
 ['06', 660,    1120],
 ['07', 530,   1020],
 ['08', 580,   1240],
 ['09', 430,   1400],
 ['10', 650,   1200],
 ['11', 560,   1000],
]);
```




```
var chartOptions = {
 title: 'Area Chart',
 width: 500,
 height: 450,
 vAxis: { gridlines: { color: '#999999' }
},
 chartArea: { color: '#F2F2F2' },
 pointSize: 6
};

var chartData = new a9.chart.DataTable([
 ['Year', 'Sales', 'Expenses'],
 ['04', 1000,   400],
 ['05', 1170,   460],
 ['06', 660,    1120],
 ['07', 530,   1020],
 ['08', 580,   1240],
 ['09', 430,   1400],
 ['10', 650,   1200],
 ['11', 560,   1000],
]);
```




```
var chartOptions = {
 title: 'Pie Chart',
 width: 500,
 height: 450,
 chartArea: { color: '#ffffff' },
 is3D: true
};

var chartData = new a9.chart.DataTable([
 ['Fruits',      'Quantity'],
 ['Bananas',     60],
 ['Apples',      100],
 ['Pears',       80],
 ['Lemons',      20],
 ['Plums',       100]
]);
```

Options Parameters

Common

Parameter name	Type	Description
Title	string	The chart title.
Width	number	Width in pixels.
Height	number	Height in pixels.
Padding	number or object	Padding in pixels. If supplied as a number the padding will be set the same for all directions.
padding.left	number	Left padding in pixels.
padding.right	number	Right padding in pixels.
padding.top	number	Top padding in pixels.
padding.bottom	number	Bottom padding in pixels.
Color	string	Default color of the chart text. Supplied as a hexadecimal color value, for example: '#ff00aa'.
fontName	string	Default font family used for text rendering.
fontSize	number	Default font size used for text rendering.
titleTextStyle	object	Modify the text styling of the title.
titleTextStyle.color	string	Color of the title text. Supplied as a hexadecimal color value.
titleTextStyle.fontName	string	Font family used for the title.
titleTextStyle.fontSize	number	Font size used for the title.
backgroundColor	string or object	Color of the background as a hexadecimal color value.
backgroundColor.fill	string	Color of the background as a hexadecimal color value.
backgroundColor.stroke	string	Stroke color as a hexadecimal color value.
backgroundColor.strokeWidth	number	Width of the background border, default is 0.
chartArea	object	Specific settings for the chart area.
chartArea.color	string	Color of the chart area background as a hexadecimal color value.
Legend	object	Settings for the chart legend.
legend.position	string	Set to 'none' to remove the legend.
Colors	array of	Array of colors that represent each series of data,

	strings	for example: ['#ff00ff', '#ff0000'].

Chart Axes

Parameter name	Type	Description
hAxis	object	Modify the horizontal axis for charts that support axes.
hAxis.visible	boolean	True if the axis should be visible.
hAxis.minValue	number	Minimum value for the axis.
hAxis.maxValue	number	Maximum value for the axis.
hAxis.color	string	Default color of the axis text as a hexadecimal color value.
hAxis.fontName	string	Default font family used for axis text rendering.
hAxis.fontSize	number	Default font size used for axis text rendering.
hAxis.title	string	Title of the axis.
hAxis.titleTextStyle	object	Modify the text styling of the axis title.
hAxis.titleTextStyle.color	string	Color of the axis title text as a hexadecimal color value.
hAxis.titleTextStyle.fontName	string	Font family used for the axis title.
hAxis.titleTextStyle.fontSize	number	Font size used for the axis title.
hAxis.gridlines	object	Modify the gridlines for the axis.
hAxis.gridlines.count	number	Number of gridlines.
hAxis.gridlines.color	string	Color of the gridlines as a hexadecimal color value.
hAxis.minorGridlines	object	Modify the minor gridlines for the axis.
hAxis.minorGridlines.count	number	Number of gridlines.
hAxis.minorGridlines.color	string	Color of the minor gridlines as a hexadecimal color value.
vAxis	object	Modify the vertical axis for charts that support axes.
vAxis.visible	boolean	True if the axis should be visible.
vAxis.minValue	number	Minimum value for the axis.
vAxis.maxValue	number	Maximum value for the axis.
vAxis.color	string	Default color of the axis text as a hexadecimal color value.
vAxis.fontName	string	Default font family used for axis text rendering.
vAxis.fontSize	number	Default font size used for axis text rendering.
vAxis.title	string	Title of the axis.
vAxis.titleTextStyle	object	Modify the text styling of the axis title.
vAxis.titleTextStyle.color	string	Color of the axis title text as a hexadecimal color value.
vAxis.titleTextStyle.fontName	string	Font family used for the axis title.
vAxis.titleTextStyle.fontSize	number	Font size used for the axis title.
vAxis.gridlines	object	Modify the gridlines for the axis.
vAxis.gridlines.count	number	Number of gridlines.
vAxis.gridlines.color	string	Color of the gridlines as a hexadecimal color value.
vAxis.minorGridlines	object	Modify the minor gridlines for the axis.
vAxis.minorGridlines.count	number	Number of gridlines.

vAxis.minorGridlines.color	string	Color of the minor gridlines as a hexadecimal color value.

Bar Chart Specific

Parameter name	Type	Description
orientation	string	Set to 'horizontal' or 'vertical' to specify the bar directions.
is3D	boolean	Display the three-dimensional representation of the chart.

Line Chart Specific

Parameter name	Type	Description
pointSize	number	Size of the data points.
lineWidth	number	Width of the lines.

Scatter Chart Specific

Parameter name	Type	Description
pointSize	number	Size of the data points.

Pie Chart Specific

Parameter name	Type	Description
pieSliceText	string	Set to 'none' to disable rendering of the text descriptions for each slice.
is3D	boolean	Display the three-dimensional representation of the chart.
shadow	boolean	Display shadows for the 3D pie chart.

APPENDIX C. Agency9 Chart API based on Mondrian

Agency9 Chart API is a tool for generating image based (PNG) or SVG charts from your data using JavaScript. User interactive feedback is enabled in the form of tooltips using HTML ImageMaps.

Setup

To start using the API include these two JavaScript libraries in your document header using:

```
<script type="text/javascript" src="http://haleh.agency9.se:8080/mondrian/jquery-1.8.2.min.js"></script>
<script type="text/javascript" src="http://haleh.agency9.se:8080/mondrian/a9-chart-api.js"></script>
```

Then initialize the API by calling the a9.chart.setApiProvider function with the URL to the providing server:

```
a9.chart.setApiProvider('http://haleh.agency9.se:8080/mondrian/');
```

a9.chart.Chart

Constructor

```
a9.chart.Chart(chartType)
```

Creates a new chart.
chartType: String that specifies the type of chart to create. For a list of valid chart types.
Methods

```
setChartType(chartType)
```

Set the type of chart to be generated.

chartType: String that specifies the type of the chart.

```
setDataTable(dataTable)
```

Set the data to be visualized by this chart.

dataTable: Instance of DataTable that includes the data to be visualized.

```
setDataService(serviceUrl, dataFormat)
```

An alternative to setDataTable. By setting a DataService data can be loaded from an external URL if the format is supported.

serviceUrl: The external URL to the data source.

dataFormat: The data format of the source. Supported formats are:

- 'Ascii'
- 'Excel'
- 'RapidAnalytics'

```
setOptions(options)
```

Specify the options parameters to use for chart generation.

options: JSON object with options. For a list of available options see Options Parameters.

```
draw(callback, [onErrorCallback])
```

Generates the chart as a PNG image and responds with a callback to the user.

callback: Function to be called when the chart has been generated. It receives three parameters, the url to the generated image, the HTML code for the image map and the identifier of the image map.

onErrorCallback: Optional function to be called if an error occurs. It receives one parameter, a string with the error message.

```
drawSVG(callback, [onErrorCallback])
```

Generates the chart as SVG and responds with a callback to the user.

callback: Function to be called when the chart has been generated. It receives one parameter, the generated SVG markup.

onErrorCallback: Optional function to be called if an error occurs. It receives one parameter, a string with the error message.

a9.chart.DataTable

Constructor

```
a9.chart.DataTable([dataArray])
```

Creates a new DataTable populated with the content of dataArray.

dataArray: A multi-dimensional array of data. Each sub-array corresponds to one row in the table.

Methods

```
addColumn(type, label)
```

Adds a column to the DataTabel.

type: A string with the data type of the values of the column. Valid types are 'string', 'number', 'boolean', 'date', 'datetime' or 'timeofday'.

label: The name associated with this column.

```
addRow(rowData)
```

Adds a row of data to the DataTabel.

rowData: An array of values that each corresponds to one column in the DataTable.

Example Usage

Generate chart from static data

```
a9.chart.setApiProvider('http://10.7.1.102:8080/mondrian/');

var chartOptions = {
  width: 600,
  height: 400,
  select: ['Age', 'Sex', 'Class']
};

var chartData = new a9.chart.DataTable([
  ['Class','Age','Sex','Survived'],
  ['First','Adult','Male','Yes'],
  ['First','Adult','Male','Yes']
]);

var chart = new a9.chart.Chart("MosaicPlot");
chart.setDataTable(chartData);
chart.setOptions(chartOptions);

chart.draw(function(imageUrl, imageMap, imageMapId) {
  document.write(imageMap);
  document.write("<img src='" + imageUrl + "' usemap='#" + imageMapId + "' />");
}, function(errorMsg) {
  alert(errorMsg);
});
```

Load data from providing service

Supported data formats at the moment are 'Ascii', 'Excel' and 'RapidAnalytics'.

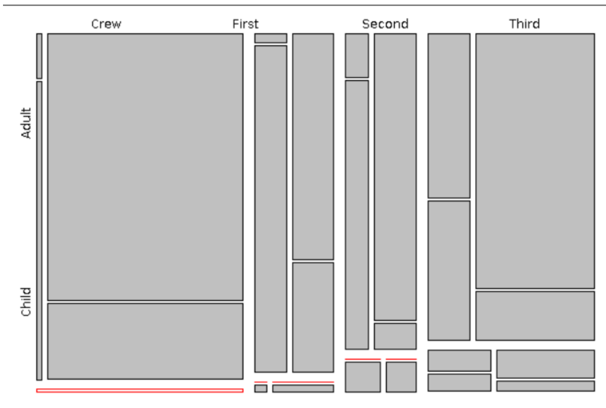
```
a9.chart.setApiProvider('http://10.7.1.102:8080/mondrian/');

var chartOptions = {
  width: 600,
  height: 400,
  select: ['Age', 'Sex', 'Class']
};

var chart = new a9.chart.Chart("MosaicPlot");
chart.setDataService('http://10.7.1.102:8080/mondrian/test/Titanic.txt', 'Ascii');
chart.setOptions(chartOptions);

chart.draw(function(imageUrl, imageMap, imageMapId) {
  document.write(imageMap);
  document.write("<img src='" + imageUrl + "' usemap='#" + imageMapId + "' />");
}, function(errorMsg) {
  alert(errorMsg);
});
```

Supported Chart Types

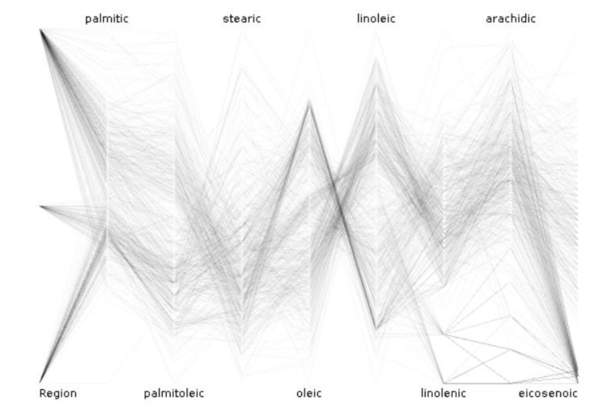
Supported values of chartType are:

- MosaicPlot
- ParallelPlot
- ScatterPlot
- MatrixPlot
- HistogramPlot
- BarChart

Demos

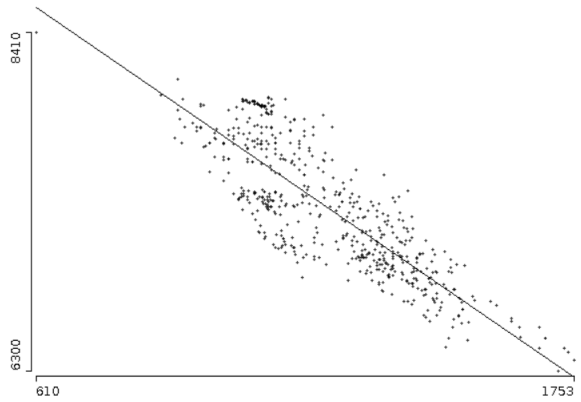



```
var chartOptions = {
        select: ['Class', 'Age', 'Sex', 'Survived'],
        width: 600,
        height: 400
};

var chart = new a9.chart.Chart("MosaicPlot");
chart.setOptions(chartOptions);

chart.setDataService(
        'http://10.7.1.102:8080/mondrian/test/Titanic.txt'
        'Ascii'
);

chart.draw( ... );
```

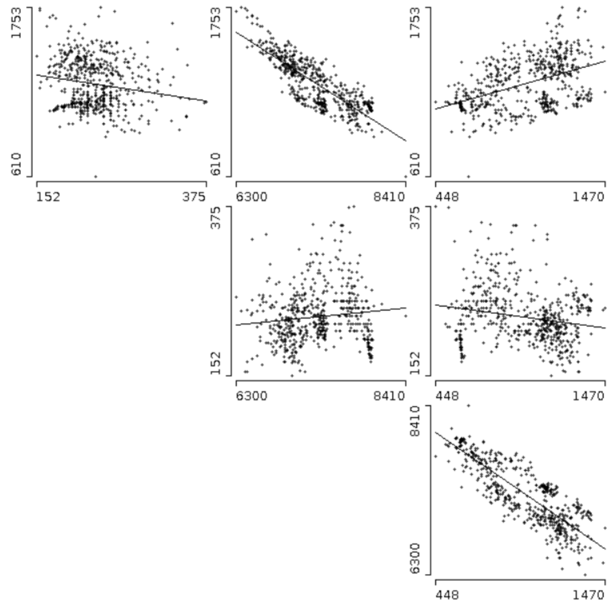



```
var chartOptions = {
        select: ['Region', 'palmitic',
                 'palmitoleic', 'stearic',
                 'oleic', 'linoleic',
                 'linolenic', 'arachidic',
                 'eicosenoic'],
        width: 600,
        height: 400,
        alpha: 0.02
};

var chart = new a9.chart.Chart("ParallelPlot");
chart.setOptions(chartOptions);

chart.setDataService(
        'http://10.7.1.102:8080/mondrian/test/Olive.txt',
```

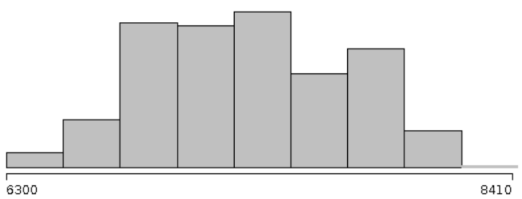
```
        'Ascii'
);

chart.draw( ... );
```



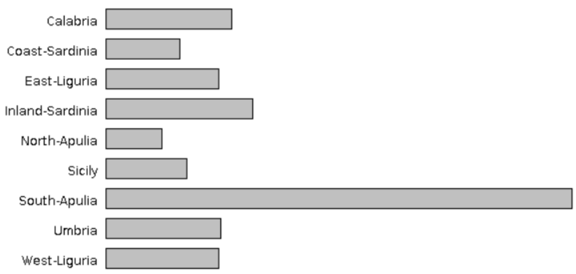

```
var chartOptions = {
        select: ['palmitic', 'oleic'],
        width: 600,
        height: 400,
        lineApprox: true
};

var chart = new a9.chart.Chart("ScatterPlot");
chart.setOptions(chartOptions);

chart.setDataService(
        'http://10.7.1.102:8080/mondrian/test/Olive.txt',
        'Ascii'
);

chart.draw( ... );
```




```
var chartOptions = {
        select: ['palmitic', 'stearic', 'oleic',
'linoleic'],
        width: 600,
        height: 600,
        lineApprox: true
};

var chart = new a9.chart.Chart("MatrixPlot");
chart.setOptions(chartOptions);

chart.setDataService(
        'http://10.7.1.102:8080/mondrian/test/Olive.txt',
        'Ascii'
);

chart.draw( ... );
```




```
var chartOptions = {
        select: ['oleic'],
        width: 600,
        height: 400
};

var chart = new a9.chart.Chart("HistogramPlot");
chart.setOptions(chartOptions);

chart.setDataService(
        'http://10.7.1.102:8080/mondrian/test/Olive.txt',
        'Ascii'
);

chart.draw( ... );
```




```
var chartOptions = {
        select: ['Area'],
        width: 600,
        height: 300
};

var chart = new a9.chart.Chart("BarChart");
chart.setOptions(chartOptions);

chart.setDataService(
        'http://10.7.1.102:8080/mondrian/test/Olive.txt',
        'Ascii'
);

chart.draw( ... );
```

Options Parameters

Common

Parameter name	Type	Description
width	number	Width in pixels.
height	number	Height in pixels.
select	array of strings	The column keys of the data that should be included in the chart.

Scatter and Matrix Plot

Parameter name	Type	Description
lineApprox	boolean	True if a regression line should be approximated for the data points.

Parallel Plot

Parameter name	Type	Description
alpha	number	The opacity that each data line contributes to the plot.

APPENDIX D. An example of a query generated by using of the shorts path calculation algorithm

```
SELECT ?Building ?Age_Class ?Bottom_Floor_Type ?Draught_Lobby ?Type_Of_Wall
?Wall_Insulation_Type ?Roof_Area ?Roof_Tilt ?addressValue
?number_Of_Complete_StoreysValue ?number_Of_RoomsValue ?ground_Floor_AreaValue
?ground_Floor_HeightValue ?building_PerimeterValue
?number_Of_Natural_Ventilation_DeviceValue
?percentage_Of_Window_Door_Draught_StrippedValue ?number_Of_Sides_ShelteredValue
?window_TypeValue

WHERE {

?Building <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.ontologyportal.org/SUMO.owl#Building> .

?Age_Class <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#Age_Class> .

?Building <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasAge> ?Age_Class .

?Bottom_Floor_Type <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.semanco-project.eu/2012/5/SEMANCO.owl#Bottom_Floor_Type> .

?Building <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasSpace>
?Conditioned_Space .

?Conditioned_Space <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#hasCS_Envelope> ?CS_Envelope .

?CS_Envelope <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasBottom_Floor>
?Bottom_Floor .

?Bottom_Floor <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#hasBottom_Floor_Type> ?Bottom_Floor_Type .

?Draught_Lobby <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.semanco-project.eu/2012/5/SEMANCO.owl#Draught_Lobby> .

?Conditioned_Space <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#hasCS_Ventilation> ?CS_Natural_Ventilation .

?CS_Natural_Ventilation <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#hasNatural_Ventilation_Device> ?Draught_Lobby .

?Type_Of_Wall <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.semanco-project.eu/2012/5/SEMANCO.owl#Type_Of_Wall> .

?CS_Envelope <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#hasVertical_Enclosure> ?Vertical_Enclosure .

?Vertical_Enclosure <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasWall>
?Wall .

?Wall <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasType_Of_Wall>
?Type_Of_Wall .

?Wall_Insulation_Type <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.semanco-project.eu/2012/5/SEMANCO.owl#Wall_Insulation_Type> .

?Wall <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasWall_Insulation>
?Wall_Insulation .

?Wall_Insulation <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#hasWall_Insulation_Type> ?Wall_Insulation_Type .

?Roof_Area <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#Roof_Area> .

?CS_Envelope <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#hasHorizontal_Superior_Enclosure>
?Horizontal_Superior_Enclosure .

?Horizontal_Superior_Enclosure <http://www.semanco-
project.eu/2012/5/SEMANCO.owl#hasRoof> ?Roof .

?Roof <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasRoof_Area> ?Roof_Area .
```

```
?Roof_Tilt <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.semanco-project.eu/2012/5/SEMANCO.owl#Roof_Tilt> .

?Roof <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasRoof_Tilt> ?Roof_Tilt .

?Building <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasAddress> ?Address .

?Address <http://www.semanco-project.eu/2012/5/SEMANCO.owl#addressValue> ?addressValue .

?Building <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasBuilding_Geometry> ?Building_Geometry .

?Building_Geometry <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasNumber_Of_Complete_Storeys> ?Number_Of_Complete_Storeys .

?Number_Of_Complete_Storeys <http://www.semanco-project.eu/2012/5/SEMANCO.owl#number_Of_Complete_StoreysValue> ?number_Of_Complete_StoreysValue .

?Building_Geometry <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasNumber_Of_Rooms> ?Number_Of_Rooms .

?Number_Of_Rooms <http://www.semanco-project.eu/2012/5/SEMANCO.owl#number_Of_RoomsValue> ?number_Of_RoomsValue .

?Building_Geometry <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasGround_Floor> ?Ground_Floor .

?Ground_Floor <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasGround_Floor_Area> ?Ground_Floor_Area .

?Ground_Floor_Area <http://www.semanco-project.eu/2012/5/SEMANCO.owl#ground_Floor_AreaValue> ?ground_Floor_AreaValue .

?Ground_Floor <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasGround_Floor_Height> ?Ground_Floor_Height .

?Ground_Floor_Height <http://www.semanco-project.eu/2012/5/SEMANCO.owl#ground_Floor_HeightValue> ?ground_Floor_HeightValue .

?Building_Geometry <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasBuilding_Perimeter> ?Building_Perimeter .

?Building_Perimeter <http://www.semanco-project.eu/2012/5/SEMANCO.owl#building_PerimeterValue> ?building_PerimeterValue .

?CS_Natural_Ventilation <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasNumber_Of_Natural_Ventilation_Device> ?Number_Of_Natural_Ventilation_Device .

?Number_Of_Natural_Ventilation_Device <http://www.semanco-project.eu/2012/5/SEMANCO.owl#number_Of_Natural_Ventilation_DeviceValue> ?number_Of_Natural_Ventilation_DeviceValue .

?Building_Geometry <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasOverall_Window_Surface> ?Overall_Window_Surface .

?Overall_Window_Surface <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasPercentage_Of_Window-Door_Draught_Stripped> ?Percentage_Of_Window_Door_Draught_Stripped .

?Percentage_Of_Window_Door_Draught_Stripped <http://www.semanco-project.eu/2012/5/SEMANCO.owl#percentage_Of_Window-Door_Draught_StrippedValue> ?percentage_Of_Window_Door_Draught_StrippedValue .

?Building_Geometry <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasNumber_Of_Sides_Sheltered> ?Number_Of_Sides_Sheltered .

?Number_Of_Sides_Sheltered <http://www.semanco-project.eu/2012/5/SEMANCO.owl#number_Of_Sides_ShelteredValue> ?number_Of_Sides_ShelteredValue .

?Vertical_Enclosure <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasWindow> ?Window .
```

```
?Window <http://www.semanco-project.eu/2012/5/SEMANCO.owl#hasWindow_Type>
?Window_Type .

?Window_Type <http://www.semanco-project.eu/2012/5/SEMANCO.owl#window_TypeValue>
?window_TypeValue .

}
```