Semantics on the Cloud: Toward an Ubiquitous Business Intelligence 2.0 ERP Desktop

Diletta Cacciagrano*, Emanuela Merelli*
*UNICAM - School of Science and Technology
Camerino, Italy
{name.surname}@unicam.it

Leonardo Vito*∓
∓Logical System srl
Jesi, Italy
leonardo.vito@gmail.com

Andrea Sergiacomi, Serena Carota
Marche Region,
Ancona, 60125, Italy
{name.surname}@regione.marche.it

***Abstract*—Adequate information management requires more than persistently storing data. Owl-$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ (to read either *owl-mining* or *owl-meaning*) is an expandable 'Business Intelligence 2.0' Enterprise Resource Planning (ERP) prototype, with the aim to lead Public Administration toward Business Intelligence and information maturity. Designed for the Marche Region, Owl-$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ allows transforming, analysing and mining distributed and heterogeneous knowledge through semantic-driven GUI (Graphical User Interface)-based components, integrated on a common semantic knowledge model and embedded in a Cloud-based middleware. Such an architecture puts Owl-$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ beyond the actual expert-oriented semantic computing and makes it a user-friendly environment, where also naive users can easily edit, monitor, execute and store transformation, analysis and mining operations as new, reusable and semantically consistent business process knowledge.**

Capabilities of (i) encoding operational knowledge into a declarative format and (ii) producing new and complex operational knowledge by composition of simpler declarative one allows realizing in Owl-$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ processes of *externalization* (i.e., converting tacit knowledge into explicit one) and *combination* (i.e., creating new explicit knowledge from existing explicit one). An example of externalization on the top of the Marche Region's data warehouse is proposed to show how exploiting Owl-$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ for converting implicit knowledge's intangible character in its successful understanding and sharing.

Keywords* - *Semantic Web; Ontology; Business Intelligence; Data Warehouse; Data Mining.

I. Introduction

Knowledge represents the intellectual principal of any company. This is particularly evident nowadays, where a nontrivial extraction of implicit, previously unknown, potentially useful information from data and its efficient use by effective Business Intelligence (BI) methods can undoubtedly promote business competition and opportunities.

Enterprise Resource Planning (ERP) systems are designed to provide such methods, with the aim of integrating all company business facets. The umbrella term ERP refers to the processes of data transformation (e.g., Extraction, Transformation and Loading—ETL), analysis (e.g., Online Analytical Processing—OLAP) and mining (e.g., querying and clustering), as well as to terms such as data quality, data enrichment, data warehouse (DW), data mart and operational data store.

ERP systems are multi-module software applications that help companies to manage important backbone operations. ERP's major objectives are (i) integrating all company departments and functions onto a single system that can serve all of the company needs, and (ii) enabling companies to present one face to their customers via integrated business processes, DWs and easy access to updated operational data.

On the one hand, ERP provides a valuable conceptual basis. On the other hand, any ERP implementation has to address several factors: information distribution, semantics heterogeneity, impossibility to test and reuse logic from existing transformations (as it is buried in source-specific code), information redundancy (when the same source feeds different data marts, being extracted and transformed by separately coded routines), absence of constrained information (complex descriptions of terms are not retained in the DW dimension tables and, as a consequence, values matching particular criteria and additional information about a term cannot be found without directly inspecting the data source), lack of user support during the mining model specification phase.

Such factors often discourage companies from fully exploiting ERP solutions, restricting their use to trivial operations (e.g., for checking and conveying known information in a more digestible manner, confirming known trends and relationships, automatically providing data for a what-if analysis still dependent on experts' manual judgments).

Major problems arise in the Public Administration (PA), where factors like low interoperability levels among information systems, budgetary restrictions, technological know-how deficits and a latent change resistance worsen the above scenario, moving PAs away from the idea to invest on ERP solutions and on BI techniques.

A. *Contribution of the paper: toward a 'BI 2.0' ERP desktop for PAs and private companies*

Until recently, theoretical research on applying ontology to data mining was carried out by several studies: for dealing with the issue of incorporating ontology in the knowledge discovery process [1], [2], [3], [4], for integrating OLAP and

 ISBN: 978-1-61208-240-0

information retrieval from DWs [5] and for multiple source integration for DW OLAP construction [6].

Taking inspiration from the literature, **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ (to read either *owl-mining* or *owl-meaning*) aims at providing an expandable ERP system with 'BI 2.0' [7] capabilities, where:

- Decisions, facts and context are developed through crowdsourcing.
- Data and reports incorporate context information supplied by users.
- Data have a more direct linkage with action. Exceptions, alerts and notifications are based on dynamic business rules that learn about user's business and what he is interested in.
- User can directly act on information.
- Business decisions can be monitored and hypotheses about business tactics can be integrated into the decision support system.
- Visualizing data and complex relationships is easier and more intuitive models of info-graphics become mainstream.
- The ability to detect complex patterns in data through automated analytic routines or intelligent helper models is built into analytic applications.
- Finding information is easier and search results provide context. Anyone looking at the same data can see that context when viewed.
- Linkages with unstructured contents as well as a previously acquired knowledge base is the key to ensuring collective knowledge and collaboration.

Owl-$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ was born as a UNICAM ICT-outsourcing product for answering Marche Region's demand of semantically unlocking earned information and ensuring high-quality and homogeneous internal decision making processes. However its modular software architecture - a mash up of Semantic Web, workflow techniques, Cloud and Agent computing embedded in a fully web- and GUI (Graphical User Interface)-based environment - makes **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ a low-cost and easily customizable solution for any PAs and private companies.

Owl-$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ consists of several fully semantic-driven and GUI-based components (currently, knowledge and workflow management, semantic annotation and visual query systems), integrated on a common semantic knowledge model and embedded in a Cloud-based middleware. This architecture makes **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ not only an innovative ERP system for transforming, analysing and mining distributed and heterogeneous knowledge, but also a user-friendly environment, where semantics helps naive users to edit, monitor, execute and store transformation, analysis and mining operations as new, re-usable and semantically consistent business process knowledge.

The semantic layer also allows filtering more specific search spaces, minimizing the possibilities of illegal settings of mining models, storing and sharing user's mining work, discriminating between usual and newly acquired knowledge.

The possibility of improving **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ through an incremental and non-invasive refinement process - thanks to the **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ Cloud-based platform where new BI components can be plugged in a compositional way - can lead therefore toward the realization of:

- An integrated knowledge space (instead of a set of isolated and heterogeneous knowledge resources) that will unify different perspectives and interpretations of knowledge resources and will enable their treatment on a far more fine grained level, allowing for more sophisticated applications and services.
- A collaborative BI working environment (instead of a single person decision making process) that will bring every user to the same level of effectiveness and productivity and will ensure more efficient knowledge sharing by providing, at the same time, the reliability and the consistency of the decision making process.
- A change management system (instead of ad-hoc management of changes) that will ensure harmonisation of requests for changes, resolution of changes in a systematic way and their consistent and unified propagation to the collaborative and knowledge space, in order to ensure the high quality of the decision-making process.
- A platform for proactive delivery of knowledge (instead of an one-way knowledge access) that enables creation of an adaptable knowledge sharing environment through learning from the collaboration between users and their interaction with the knowledge repository and supporting in that way full empowerment and acceptance of users. A strong involvement of employees and stakeholder representatives is crucial, since defining the corporate vision is often the first step toward manifesting strategic thinking in PAs and enterprises.
- An ubiquitous assistive mining environment for storing/changing/extending/generalizing mining rules, filtering more specific search spaces by concept-based queries, minimizing the possibilities of illegal settings of mining models, storing/re-using user's work.

Owl-$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ can be tried at the link `http://resourceome.cs.unicam.it/eyeOS/`**(11/08/2012)** (User: owlmining, Passw: tryowlmining) .

B. Plan of the paper

The remainder of this paper is organized as follows: Section II presents the **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{in}\mathscr{G}$ overall architecture, giving details about each components - a declarative and operational knowledge management environment (*Resourceome*), a semantic annotation component (*DataSMart*) and a semantic-driven visual query editor (*OWLEye*). Finally, Section III closes the paper, with a sketch of the ongoing implementation results and intended future work.

II. The **Owl-**$\mathcal{M}_i^{ea}\mathcal{N}^{fn}\mathcal{G}$ Architecture

Owl-$\mathcal{M}_i^{ea}\mathcal{N}^{fn}\mathcal{G}$ is conceived as a semantic ERP platform, pivoting on a semantic DW that stores ontology-based semantic annotations, along with semantic-driven mechanisms for the definition and the execution of transformation and meaning processes over the stored data (see Fig. 1). It is based on a pluggable architecture exploiting and integrating techniques from diverse areas such as Cloud Computing, databases, machine learning, cognitive science, Semantic web, and others.

Currently, **Owl-**$\mathcal{M}_i^{ea}\mathcal{N}^{fn}\mathcal{G}$ embodies as services several fully semantic-driven and GUI-based components - namely, a declarative and operational knowledge management environment (*Resourceome*), a semantic annotation component (*DataSMart*) and a visual query editor (*OWLEye*) - pivoting on a common hybrid OWL/SKOS-based multi-layered knowledge model for the semantic annotation of *resources* and *activities* (see Fig. 2) [8]. A Cloud-based middleware (*EyeOS*) provides the needed integration mechanisms between each **Owl-**$\mathcal{M}_i^{ea}\mathcal{N}^{fn}\mathcal{G}$ component and the knowledge base (modeled as in Fig. 2) and among **Owl-**$\mathcal{M}_i^{ea}\mathcal{N}^{fn}\mathcal{G}$ components themselves, allowing also further meaning services (developed on the top of a knowledge model as in Fig. 2) to be plugged in **Owl-**$\mathcal{M}_i^{ea}\mathcal{N}^{fn}\mathcal{G}$ without changing its current architecture.

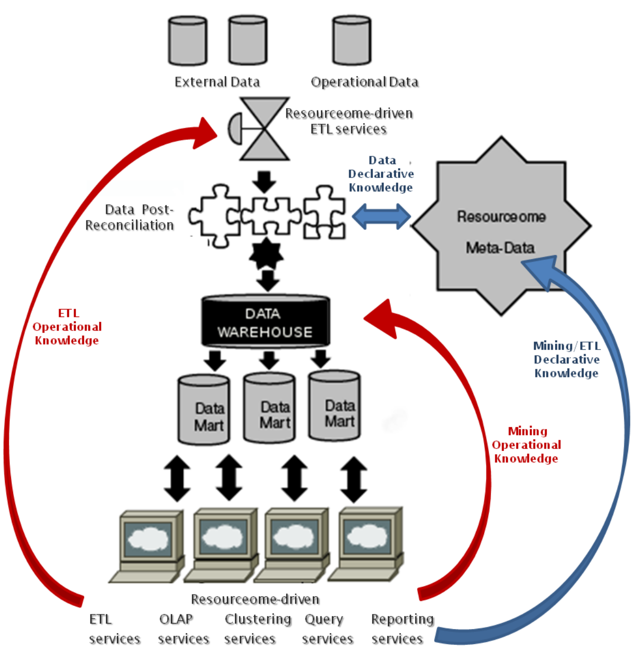

Figure 1: **Owl-**$\mathcal{M}_i^{ea}\mathcal{N}^{fn}\mathcal{G}$ conceptual view.

A. Resourceome + DataSMart*: the semantic ERP kernel*

Resourceome and *DataSMart* are the semantic core of **Owl-**$\mathcal{M}_i^{ea}\mathcal{N}^{fn}\mathcal{G}$. Both components work on the top of a specific knowledge model (Fig. 2) that, when applied to a knowledge base, allows contextualizing (i) resources w.r.t. a given domain and (ii) activities w.r.t. given resources. Usually, the knowledge base is represented by a DW, but can be also the integration of the DW with local and remote sources.

Requirement (i) is satisfied by splitting the Domain Ontology in [9] into two separate ontologies - a *Domain Ontology* conceptualizing the chosen domain instance and a *Resource Ontology* conceptualizing the resource space - and connecting them by abstract relations. Abstract relations also connect Domain and Resource Ontologies to a *Task Ontology* conceptualizing the activity hyperspace; such relations allow any activity to be linked to its working context and the involved roles and resources, thereby satisfying requirement (ii).

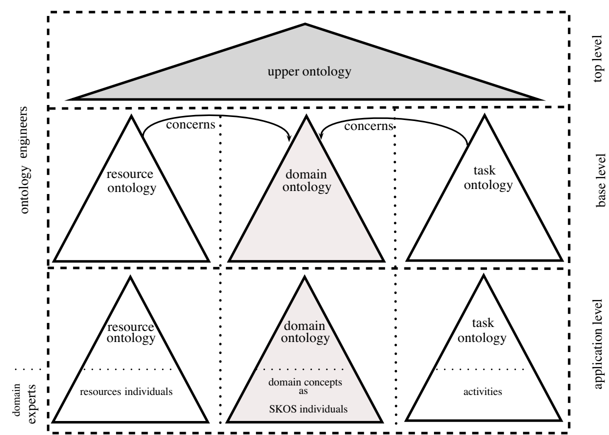

Figure 2: The knowledge model.

On the one hand, *DataSMart* is a *BioMart*[10]-based database federation system that makes it possible to present geographically distributed data sources as federated data in an integrated database, as well as to access and to cross-reference data from these data sources using a single user interface.

However, differently from *BioMart*, it can be also exploited as a data warehousing platform enabling ETL, OLAP and other mining operations (see Fig. 3). Most important, *DataSMart* is also a semantic annotation system based on a drag-and-drop interface, which allows imported data and attributes to be linked to a given knowledge model instance (see Fig. 4).

On the other hand, *Resourceome* [8] provides a web-based integrated environment for (i) managing distributed and heterogeneous knowledge as ontology concepts (e.g., as declarative knowledge); (ii) designing semantically consistent ETL/mining operations; (iii) running ETL/mining operations as distributed and mobile agent systems (e.g., as

 ISBN: 978-1-61208-240-0

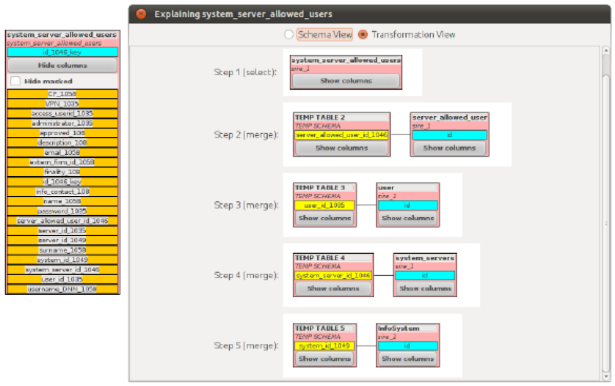

Figure 3: ETL-Transform and Load in *DataSMart*.

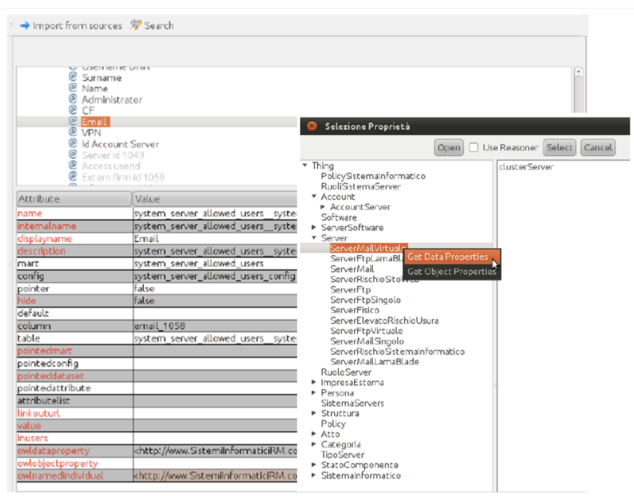

Figure 4: Semantic annotation by *DataSMart*.

operational knowledge); (iv) storing ETL/mining operations as ontology concepts (e.g., as declarative knowledge). Functionalities (i)-(iv) and (ii)-(iii) are provided respectively by a Knowledge Management System (KMS) and a Workflow Management System (WMS), both working on the knowledge base (modeled as in Fig. 2). Fig. 5 presents the final screenshot of an analysis process on financial data, edited and executed through *Resourceome* WMS and visualized by a *Resourceome*-driven reporting service.

B. A dragging-and-dropping environment for conceptual queries

Formulating non-ambiguous queries is often a too demanding task for users as they do not have the overview on the semantics of data stored in the system. Without complete comprehension of the schema and domain related knowledge, end users may develop a query based on their experience or intuition. Therefore, users' formulation of queries can possibly fall into some improper pits. This may lead to incorrect and redundant mining data space or mining

Figure 5: Example of analysis process output visualized by a *Resourceome*-driven reporting service.

results and waste the efforts accordingly.

The goal of *OWLEye* is to overcome this problem by providing an ontology-based information view of the data available in the knowledge base, integrated with a visual querying environment oriented to unskilled users.

OWLeye is equipped with a Query Design component allowing the graphical rendering of SPARQL queries by graphical constructs of the vSPARQL language [11]. This has necessarily entailed the development of a set of graphic notations - based on SPARQL syntax specification - supporting the visual representation of SPARQL query components.

Many of the vSPARQL constructs, once rendered, are selectable objects that can be edited using a popup menu. The menu allows users to define filtering, ordering and grouping information for the selected object. The design canvas itself can be zoomed and panned to view the entire query at different levels of resolution.

The possibility of browsing the knowledge model - embedded in the knowledge base through the *DataSMart* semantic annotation - insulates inexperienced users from the complexity of the query language and guides them in the process of query formulation. When the knowledge model is constructed correctly, the user can formulate semantically correct queries in a very intuitive way: dragging-and-dropping graphical elements allows user to browse the knowledge base and to select specific concepts of the knowledge model, while "stretching" edges permits to select properties and relations of interest (those associated to the stretched edges). Finally, query results can be visualized through several view layouts. An illustrative example of such features is given in Fig. 6.

Example II.1. We show how **Owl-**$\mathscr{M}_{i}^{ea}\mathscr{N}^{in}\mathscr{G}$ has been exploited to capture implicit knowledge from the Marche Region's knowledge base. As a single scenario cannot cover all the application possibilities, we focus on a specific Marche Region's request: knowing what *Struttura* (i.e., departments) are *StrutturaVulnerabile* (i.e., vulnerable department). For Marche's Region a department is considered vulnerable when at least five of its *SistemaInformaticoAmministrato*

 ISBN: 978-1-61208-240-0

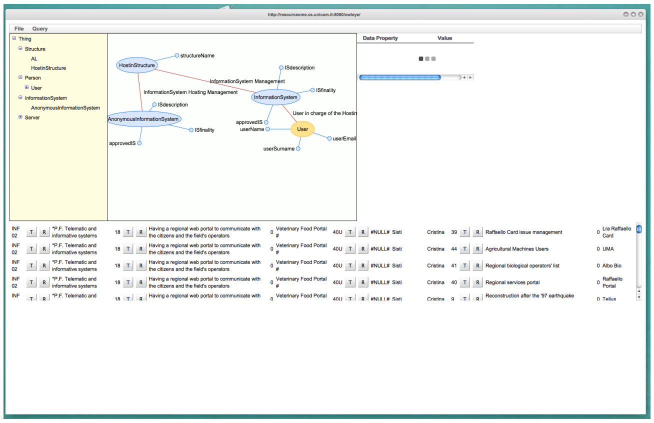

Figure 6: Example of query through *OWLEye*.

(i.e., information systems) manage personal and sensitive data.

This fact can be expressed by the OWL rule shown in Fig. 7, involving the concepts of *Struttura*, *SistemaInformaticoAmministrato*, *PolicySistemaInformatico* and their relations. Fig. 8 shows the query formulated by *OWLEye* and the list of inferred vulnerable departments.

$$\begin{aligned}&\mathit{StrutturaVulnerabile} \equiv \mathit{Struttura} \sqcap\\ &\quad \geq 5\ \mathit{strutturaAmministraSistemaInformatico}.(\mathit{SistemaInformatico} \sqcap\\ &\quad \exists\ \mathit{sistemaInformaticoAssociatoPolicySI}.(\mathit{PolicySistemaInformatico} \sqcap\\ &\quad \exists\ \mathit{policySistemaInformaticoAssociatoPolicy}.(\mathit{Policy} \sqcap\\ &\quad \exists\ \mathit{datiPersonaliPolicy}.\{\mathsf{true}\} \sqcup \exists\ \mathit{datiSensibiliPolicy}.\{\mathsf{true}\})))\end{aligned}$$

Figure 7: Rule for inferring vulnerable departments.

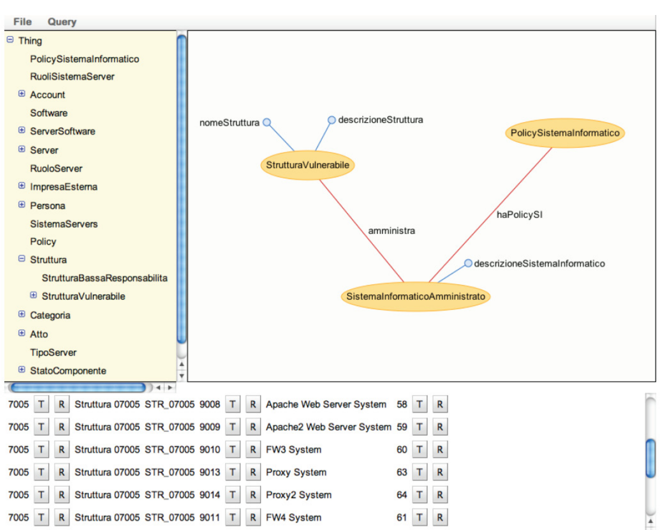

Figure 8: Query in *OWLEye* and list of inferred vulnerable departments.

III. State-of-the-art Implementation and Future Work

There are a number of additional application areas for **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{m}\mathscr{G}$ that we are exploring as part of our current and future work. In particular, we are studying the possibility to exploit **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{m}\mathscr{G}$ for rule creation, information integration and knowledge acquisition.

It is well-known that SPARQL CONSTRUCT queries can be used for information integration and interoperability, since this kind of queries effectively modify and extend (perhaps multiple) knowledge bases according to the presence of information detected from one or more information sources. Since *OWLEye* supports the visualization of (among else) SPARQL CONSTRUCT queries, we argue that *OWLEye* can be used in **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{m}\mathscr{G}$ also for editing rules and for representing the semantic mappings (or ontology alignments) between ostensibly disparate ontologies.

Another interesting point concerns the possibility to exploit **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{m}\mathscr{G}$ for knowledge acquisition. Cluster mining is usually applied to discover groups in large amounts of data using large flat files as input source and, as a consequence, mining techniques are simply seen as tools trying to discover patterns.

As in the case of query-based mining, putting semantics into cluster mining allows to make explicit the conceptual knowledge structures of data, to take advantage of knowledge acquired in the previous knowledge discovery process stages, to provide users with further semantics that improves the understanding of the system, as well as to abstract from specific issues (platform, algorithms, parameters, etc).

For this reason, we plan to integrate in **Owl-**$\mathscr{M}_i^{ea}\mathscr{N}^{m}\mathscr{G}$ a *Resourceome*-driven clustering service equipped with a smart drag-and-drop based editor as *OWLEye*. Such a service shall embed a clustering algorithm with a level of accuracy similar to corpus-based ones but retaining the low computational complexity of path-based ones. At this aim, we are studying a weighted and ontology-based variant of the k-means algorithm [12], where weights are assigned on both data properties and relations and represent the importance level (see Fig. 9). The variant relies on a similarity measure defined as below:

$$sim_g = \left(\sum_{k}^{m} w_k (v_{k,i} - v_{k,j})^g\right)^{\frac{1}{g}}$$

where $r_i = \{v_{i,1}, v_{i,2}, \ldots, v_{i,k}\}$ denotes the value list of the $i-$th record in the dataset $D = \{r_1, r_2, \ldots, r_n\}$, $S = \{a_{1,1}, .., a_{1,n_1}, a_{2,1}, .., a_{2,n_2}, .., a_{k,1}, .., a_{k,n_k}\} = \{a_1, a_2, .., a_N\}$ the attribute set (with $\sum_{i=1}^{k} a_{i,n_i} = N$), $a_{i,k}$ the $k-$th attribute of the $i-$th table, $w_i \in (0..1]$ the weight of a_i.

Notice that sim_g can express the absolute distance $(g = 1)$, the euclidean distance $(g = 2)$ and the Chebyshev distance $(g \to \infty)$.

 ISBN: 978-1-61208-240-0

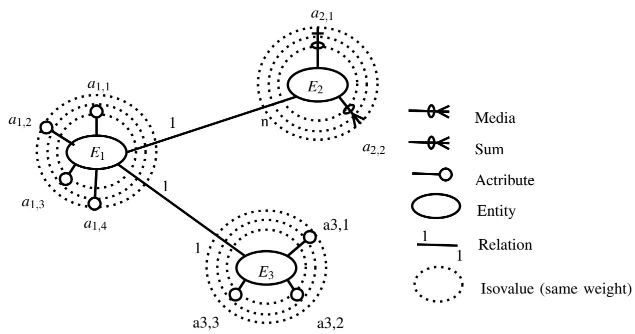

Figure 9: Clustering visualization system.

ACKNOWLEDGEMENTS

A special thank to Sergio Villarreal and Constatino Giuliodori (Marche Region), Sauro Silvestrini (General Impianti Srl) and Pietro Liò (University of Cambridge), that gave an important contribution to the **Owl-**$\mathscr{M}^{ea}_{i}\mathscr{N}^{in}\mathscr{G}$ design.

REFERENCES

[1] J. Aronis, F. Provost, and B. Buchanan, "Exploiting background knowledge in automated discovery," in *In the Proc. of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 355–358.

[2] S. Sharma, Osei-Bryson, and Kweku-Muata, "Framework for formal implementation of the business understanding phase of data mining projects," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 4114–4124, Mar. 2009.

[3] M. Domingues and S. Rezende, "Using taxonomies to facilitate the analysis of the association rules," in *In the Proc. of the Second international workshop on knowledge discovery and ontologies*, 2005.

[4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The kdd process for extracting useful knowledge from volumes of data," *Commun. ACM*, vol. 39, no. 11, pp. 27–34, Nov. 1996.

[5] T. Priebe and G. Pernul, "Ontology-based integration of olap and information retrieval," in *In the Proc. of the 14th intern. workshop on database and expert systems applications*, 2003, pp. 610–614.

[6] W. Lin, M. Tseng, and C. Wu, "Ontology-incorporated mining of association rules in data warehouse," *Journal of Internet Technology*, vol. 8, no. 4, pp. 477–485, 2007.

[7] G. S. Nelson, "Business Intelligence 2.0: Are we there yet?" in *In the Proc. of SAS Global Forum 2010*, Paper 040-2010, 2010.

[8] D. Cacciagrano, F. Corradini, E. Merelli, L. Vito, and G. Romiti, "Resourceome: a multilevel model and a Semantic Web tool for managing domain and operational knowledge," in *The Third International Conference on Advances in Semantic Processing (SEMAPRO 2009)*, P. Dini, J. Hendler, and J. Noll, Eds. IEEE Computer Society, 2009, pp. 38 – 43.

[9] N. Guarino, *Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6-8, 1998, Trento, Italy*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 1998.

[10] J. Zhang, S. Haider, J. Baran, A. Cros, J. Guberman, J. Hsu, Y. Liang, L. Yao, and A. Kasprzyk, "BioMart: a data federation framework for large collaborative projects." *Database: the journal of biological databases and curation*, vol. 2011, no. 0, 2011.

[11] M. Shaw, L. T. Detwiler, N. Noy, J. Brinkley, and D. Suciu, "vsparql: A view definition language for the semantic web," *Journal of Biomedical Informatics*, vol. 44, no. 1, pp. 102 – 117, 2011.

[12] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, L. M. L. Cam and J. Neyman, Eds., vol. 1. University of California Press, 1967, pp. 281–297.