- ORIGINAL ARTICLE -

Automatic Suggestions to Improve the Quality of Scatterplots during its Creation: a Case Study of Ontology and Semantic Reasoning Applied to Visualization

Mejorando la Calidad de los Scatterplots con Sugerencias Automáticas durante su Creación: un Caso de Estudio de Visualización Basada en Semántica.

Dana K. Urribarri[1], Martín L. Larrea[1], and Silvia M. Castro[1]

[1] *VyGLab (UNS-CICPBA), Departamento de Ciencia e Ingeniería de la Computación,*
Universidad Nacional del Sur, Bahía Blanca, CP8000, Argentina
{dku, mll, smc}@cs.uns.edu.ar

Abstract

The process of creating a visualization is a very complex exploration activity and, even for skilled users, it can be difficult to produce an effective visualization. Since the result of such a process depends on the user's decisions along it, one way to improve the probability of achieving a useful outcome is to assist the user in the configuration and preparation of the visualization. Our proposal consists in live suggestions on how to improve the visualization. These live suggestions are based on the user decisions, and are achieved by the integration of semantic reasoning into the visualization process. In this paper, we present a case study for scatterplots visualization that combines ontologies with a semantic reasoner and helps the user in the generation of an effective visualization.

Keywords: ontology, scatterplot, semantic-based visualization, semantic reasoner, visualization quality prediction

Resumen

El proceso de creación de una visualización es una actividad de exploración muy compleja e, incluso para usuarios expertos, puede ser difícil obtener como resultado una visualización efectiva. Dadp que este resultado depende de las decisiones que el usuario toma a lo largo del proceso de visualización, una forma de mejorar la probabilidad de lograr un buen resultado es ayudar al usuario en la configuración y preparación de la visualización. Nuestra propuesta consiste en ofrecer sugerencias sobre cómo mejorar la visualización mientras el usuario la está creando. Estas sugerencias se basan en las decisiones que el usuario tomó y se logran mediante la integración del razonamiento semántico al proceso de visualización. En este artículo, presentamos un caso de estudio para la visualización de *scatterplots* (gráficos de dispersión), en el cual se muestra cómo se asiste al usuario, mediante la combinación de ontologías con un razonador semántico, para logar una visualización efectiva.

Palabras claves: ontología, predicción de calidad, razonador semántico, scatterplot, visualizaciones basadas en semántica

1 Introduction

Huge amount of data is becoming available, ranging from unstructured and multimedia documents to structured data stored in databases. This is extremely useful and exciting, but the ever growing amount of available data generates cognitive overload and even anxiety, especially in novice or occasional users. Nowadays computer technology allows the visual exploration of big data resources; however, one major setback in the visualization process is that, generally, the only way to evaluate the quality of a visualization is by creating the visualization itself.

From a dynamic point of view, the visualization can be perceived as a process that takes data from the user domain (i.e. the input or raw data), processes them, and gives the view back. Although researchers in the area have presented different models, all agree that this process is made up of states and transformations [1, 2, 3, 4]. The quality of a visualization could be measured along the different stages of this process, and the view (the output of the process) is the most straightforward stage to evaluate the result. However, an evaluation of the visualization in this last step implies the generation of the visualization itself, regardless of the complexity of the process, even if it is not going to be an effective one.

One of the goals in a semantic-based visualization process is the development of a visualization model that considers the semantics of both the data and the different stages in the visualization process. By making these considerations, the visualization process will be able to determine the characteristics of an effective visualization and to guide the user through the different stages. This research topic aims to establish a common visualization vocabulary, that includes the underlying semantics. This common vocabulary enables the definition of visualization specifications that can be executed by a visualization engine with ontological support. A visualization ontology defines the vocabulary and, with the addition of inference rules to the system, a reasoner can derive conclusions about visualization properties. These conclusions not only allow the enhancement of the visualization, but also guide the user throughout the entire process toward an effective result.

A widely used and popular visualization is the scatterplot technique. This diagram uses Cartesian coordinates to visualize with points (or glyphs) the values of bidimensional variables.

To analyze the possibilities that semantics offers, we selected a metric that measures the quality of a scatterplot without the need of rendering it [5]; however, as it is presented in the original paper, the interpretation and understanding of the given number are not trivial and inadvisable for non-expert users. Thus, we developed a detailed interpretation of the result and the actions to be taken accordingly. To help non-expert users we propose a two-part solution: the translation of our interpretation of the metric into an ontology and the usage of a semantic reasoner to assist the non-expert user in the decision-making process of the creation a scatterplot. We have implemented this assistant as a web service and web page.

Since configuring a scatterplot could be time-consuming and frustrating, a visibility index [5] was developed. This index measures the quality of the scatterplot without the need of rendering it and its interpretation gives clues on how to improve the visualization. However, the user requires advance knowledge in scatterplots and the visibility index to get a proper interpretation. To help the user in this situation we propose a novel approach: translating this knowledge into an ontology, which is used by a semantic reasoner to assist the non-expert user in the decision-making process of creating a scatterplot.

The remainder of the article is organized as follows. Next section, "Previous Work on Semantics and Visualization", presents the research carried out in the integration of semantics into the visualization process. Then, in section "Case Study: Semantic Reasoning Applied to Scatterplots Design" we present a case study that uses an ontology and a semantic reasoner to guide the user in the configuration of a scatterplot according to a visibility index. Finally, in the last section, we draw some conclusions and outline future work.

2 Previous Work on Semantics and Visualization

The integration between semantics and visualization has been of great interest to the scientific community. Much work has been done to formalize the visualization topics and to define visualization reference models. Duke et al. [6] argued the need for increasing the rigorousness of visualization descriptions to explicitly define a visualization ontology, and also gave some clues about how it can be achieved. In this sense, data and/or task oriented taxonomies are partial views of the concepts of the Visualization field that must be explicitly modeled by every valid visualization ontology. One of the first works in establishing a taxonomy about visualization concepts is [7] where a classification is established on the types of data admissible for a visualization. In [8], a compilation of the literature is made at the time of its publication and a series of rules that seek to formalize the process of creating a visualization is presented. In [9] the work done in [7] is extended and a new taxonomy is proposed. This one aimed to help implementers to understand how to apply and implement information visualization techniques. Another taxonomy based on models of a data set rather than attributes of the data itself was presented by [10]. A taxonomy associated with benchmark datasets and specific visualization tasks for graph visual representation appeared in [11]. The work published in [12] proposes a new visualization notation introducing it in the context of a new visualization reference model, which aims to describe the visualization process in a clear and unambiguous way. All these works show that the need for a formal representation of the concepts involved in visualization has been present for many years.

There are good examples of how these concepts evolved into semantic information in the Visualization field. In [13] the authors created a framework for semantic integration and querying over heterogeneous geospatial data sets distributed over the grid environment. In [14] semantic information was used in the form of ontologies in order to create an ontology-based image retrieval system for asteroideae flower family domain. Thellmann et al. [15] introduced an automatic visualization workflow which guides users through the process of creating visualizations by automatically categorizing and binding data to visualization parameters. The approach was based on an heuristic analysis of the input data structure and a comprehensive visualization model facilitating the automatic binding between data and visualization options. The resulting assignments were ranked and the highest ranking visualization instantiations were presented to the user. Minu & Thyagharajan [16] created an ontology that combined visual features with textual elements in order to represent the asteroideae family flower domain. A semantic reasoner was used to complete the ontology information. Healey et al. [17] presented a semi-automatic visualization assistant that helps users to construct perceptually sound visualizations for large multidimensional datasets. Golemati et al. [18] proposed a framework that uses context information and a set of rules to automatically select a suitable visualization method. Koop et al. [19] presented a method for aiding in the visualization pipeline design. A database of pipeline configurations was used to properly complete the configuration of the user's pipeline. Gilson et al. [20] combined a domain ontology with a visual representation ontology to automatically select a proper visualization for web data. Several efforts to define generalized visualization ontologies have also been made. Brodlie et al. [21] presented a seminal paper in this matter. In that work, a top level ontology was outlined and future tendencies were given. Moreover, Chen et al. [22] discussed the role of knowledge in visualization and identified possible trends. Kalogerakis et al. [23] proposed a graphics ontology representing the semantics for a 3D-scene in order to enhance the retrieval capabilities of search engines.

Additionally, semantic specifications for particular visualization aspects such as the user domain data classification, the visual representation, and the visual mapping were developed. Some examples include size-based data classifications [24, 25], a taxonomy for visualization algorithms that was based on assumptions over their inputs [26], the characterization of visual variables to represent visual representations at a higher level of abstraction [27], the use of fuzzy logic semantics to replace the traditional transfer function setup in illustrative volume rendering [28, 29], and a specific semantic model created by a machine learning mechanism that used representative dataset collections as training sets [30].

Unlike all these works, on this article we focus on how ontologies and semantic reasoning can represent the interpretation of information and aid the user in the creation a visualization. Particularly, in this paper we consider information regarding a visualization metric. We have been working on the subject of semantic-based visualization [31, 32, 2, 33, 34, 35] by integrating semantic information and reasoning into the visualization-generation process and also using a semantic reasoner in the selection of the colors in a visualization [32].

3 Case Study: Semantic Reasoning Applied to Scatterplots Design

Based on the visibility index [5], we developed an ontology that encapsulates the knowledge of the index's interpretation. Using a semantic reasoner to infer from this knowledge, we can suggest to the user proper modifications to the scatterplot configuration without the need of actually creating the visualization. In the following subsections we detail the work done.

3.1 The visibility index

The visibility index [5] was defined as a specific metric for scatterplots. Given a scalar dataset and the window's and glyph's dimensions, it estimates the expected percentage of glyphs that are not completely overlapped with other glyphs (there exists at least one pixel of the glyph which is not overlapped with another glyph), that is, the expected number of glyphs that are always visible despite the rendering order. The mathematical model that defines the visibility index assumes that the data is normally distributed and the scatterplot will be rendered with square glyphs in a square window. This model allows to estimate the index before achieving the final step in the visualization process.

Given a normally distributed dataset of cardinality x visualized in a square window of size h with square glyphs of size p, the visibility index is defined as:

$$f(x,h,p) = \frac{1}{1+e^{a\ln(x)+b\ln(h)+c\ln(p)+d}}$$

where,

$$a = 1.86056686 \qquad b = -3.25349985$$
$$c = 2.91520408 \qquad d = -0.68834377$$

A right combination of visual parameters must be chosen in order to get an appropriate visualization of the dataset. As the user knows the dataset, and therefore the amount of data, the goal of this metric is to guide him while choosing these right parameters, in particular, the window's and glyph's sizes.

Given a specific dataset, if the user chooses to visualize it with a scatterplot with certain glyph's and window's sizes, and the obtained estimation is not promising (for example, below 25% of visibility), the user may try to improve that estimation reducing the glyph's size or increasing the window's size.

However, even though this model contemplates windows as big or glyphs as small as necessary, in practice, the size of the window should not be bigger than the size of the available display and the glyph cannot be smaller than 1 pixel. Therefore, it would not be possible to obtain a better result than the one with the larger possible window and the minimum possible glyph.

3.2 Index's interpretation

The index value can provide information on how to improve, if possible, the visualization itself. A right interpretation of this value can give an expert valuable information on how to improve the visualization. However, this interpretation is not only absent in the original article but also not trivial and inadvisable for non-expert users.

Let $\tilde{p}$ be the inflection point of f, that is the value of p where $\frac{\partial f}{\partial p}$ has a local extremum (in particular, it has its minimum), $\frac{\partial^2 f}{\partial p^2}$ is equal to zero, and $\frac{\partial^3 f}{\partial p^3}$ is different from zero (see Figure 1). Then,

$$\tilde{p} = \left(\frac{(c-1)e^{-d}x^{-a}h^{-b}}{c+1}\right)^{1/c}$$

and

$$f(x,h,\tilde{p}) = \frac{c+1}{2c} \approx 0.671515.$$

The value 0.67 is assumed as the minimum acceptable visibility index for a visualization.

For a given dataset of size x, if the user selects a window size of W and a glyph size of G, then there are four values involved in the index interpretation:

- cWG: this value is the visibility index calculated with the user's selected dimensions for the window and for the glyphs, that is $f(x,W,G)$.

- $cW_{max}1$: this value is the visibility index calculated with the dimension of the biggest possible window and of 1-pixel glyphs, that is $f(x,W_{\max},1)$.

- $cW_{max}G$: this value is the visibility index calculated with the dimensions of the user's selected glyph and of biggest window possible, that is $f(x,W_{\max},G)$.

- $cW1$: this value is the visibility index calculated with the dimensions of the user's selected window and of 1-pixel glyphs, that is $f(x,W,1)$.

A value greater than or equal to 0.67 implies that the combination of window and glyph sizes used in the calculation are suitable for the amount of data. However, a value smaller than 0.67, implies that the combination is not appropriate.

There are six combinations of the cWG, $cW_{max}1$, $cW_{max}G$ and $cW1$ values that lead to different interpretations:

- *good*: If cWG is greater than or equal to 0.67 then the user's selected window and glyph dimensions will produce a fair visualization.

- *bad W*: If cWG and $cW1$ are smaller than 0.67, but $cW_{max}G$ is not, then the user's selected window should be modified.

- *bad G*: If cWG and $cW_{max}G$ are smaller than 0.67, but $cW1$ is not, then the user's glyph dimensions should be modified.

- *bad W or G*: If cWG is smaller than 0.67, but $cW_{max}G$ and $cW1$ are not, then there is a problem with the user's selected window or glyph dimensions. One of those two should be modified.

- *bad W and G*: If cWG, $cW1$ and $cW_{max}G$ are smaller than 0.67, but $cW_{max}1$ is not, then both the user's selected window and glyph dimensions should be modified.

- *bad*: If $cW_{max}1$ is smaller than 0.67 then it is not possible to create a fair visualization.

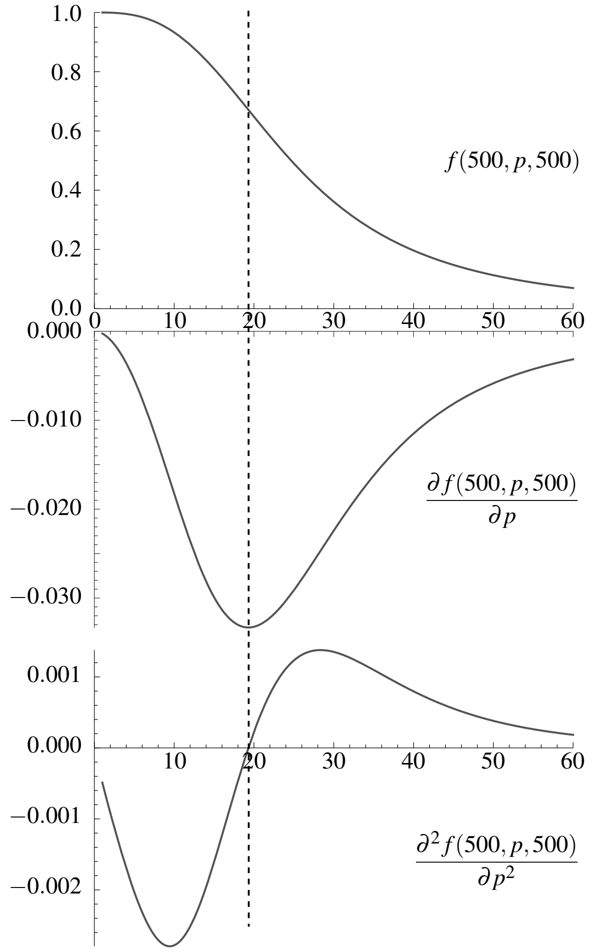

Figure 1: The inflection point $\tilde{p}$ of f is the point where $\frac{\partial f}{\partial p}$ has a minimum, that is $\frac{\partial^2 f}{\partial p^2}$ is equal to zero, but $\frac{\partial^3 f}{\partial p^3}$ is different from zero.

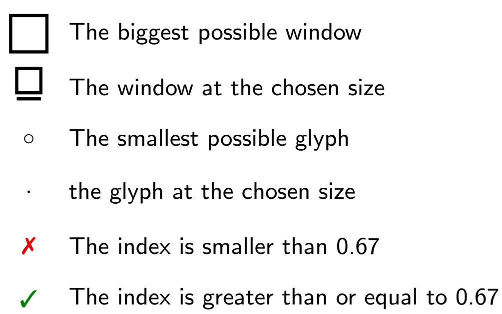

Figure 2: Iconography used in the visual representation of the index's interpretation.

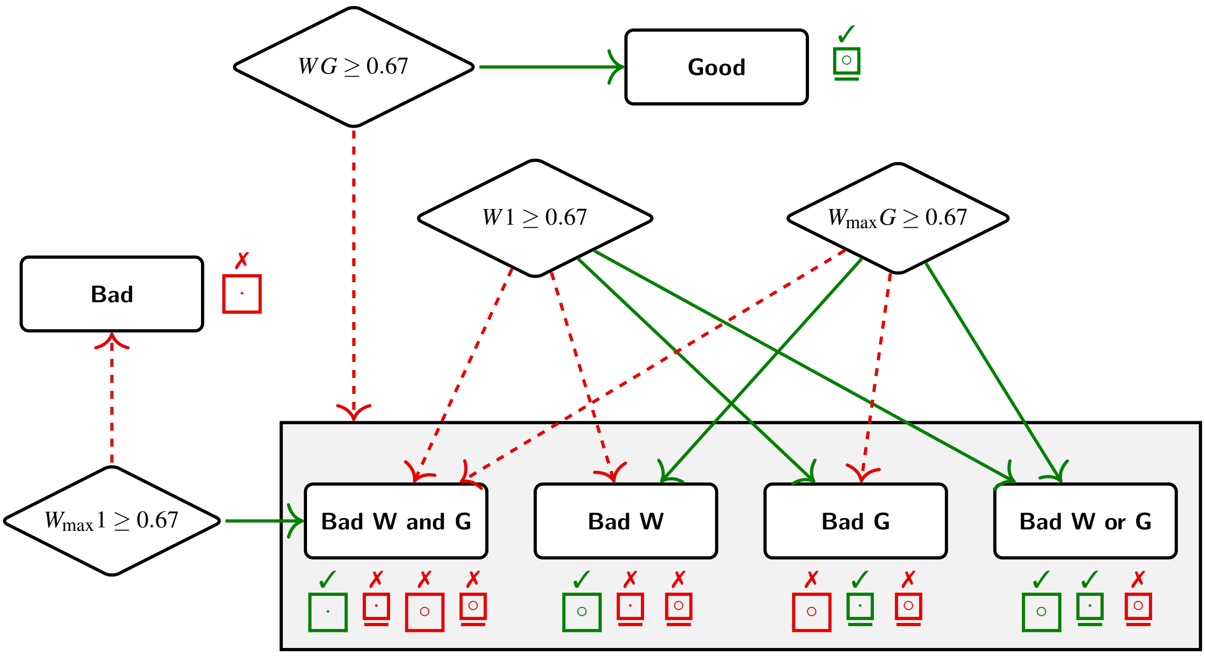

Figure 3: There are six combinations of the cWG, $cW_{max}1$, $cW_{max}G$ and $cW1$ values that lead to different interpretations. In this graph, a uniform line means that the condition was true, while a dashed one means that the condition was false. True and false correspond to the index being greater than or equal to 0.67 and smaller than 0.67, respectively. These conditions are also reinforce with colors green and red, both in the lines and the icons.

Figure 2 presents the iconography used in Figure 3 to summarize the six interpretations of the index.

In all these scenarios, changing the dimensions of the glyph implies reducing the glyph size and changing the dimensions of the window implies increasing the window size. Usually, the interpretation of the visibility index is not an easy task for a novice user. It requires advance knowledge not only on the index but also on the usage of scatterplots.

3.3 An ontology for metric-based decision making

An ontology defines a common vocabulary for those who need to share information in a specific domain. It includes machine-interpretable definitions of basic concepts in the domain and the relationships among them. An ontology is an explicit formal description of concepts (called classes) in a domain of discourse; for each concept it includes not only the properties that describe the concept's features and attributes, but also the restrictions on these properties. An ontology, in conjunction with a set of individual instances of classes, constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base begins [36, 37].

In this work, the ontology (created using Racer Pro [38]) describes the interpretation of the visibility index as defined in the previous section (see Figure 4). Six atomic concepts represent the six possible interpretations, *badWorG*, *badW*, *badG*, *badWandG*, *bad* and *good*. The other atomic values are for temporal values. The four objects cWG, $cW_{max}1$, $cW_{max}G$ and $cW1$ are the input values.

The knowledge base is formed by this ontology and the instance values of cWG, $cW_{max}1$, $cW_{max}G$ and $cW1$ corresponding to a particular visualization (see, for example, Figure 5). From this knowledge base, the semantic reasoner can derive a conclusion from one of the possible *badWorG*, *badW*, *badG*, *badWandG*, *bad* and *good* interpretations.

Both the semantic reasoner and the ontology can be tested in a development environment. However, we designed a case study to test these elements in a more realistic scenario that is easily accessible to the visualization community.

3.4 Implementation & Use Cases

The application created from this research is publicly available through a web page[1], which in turn uses a web service[2] and a semantic reasoner running in our server. Moreover, the web page and the web service are the connection between the user and the reasoner. The reasoner not only uses the ontology to answer the user questions, but also makes it transparent to the user.

In the web page, the user sets the amount of data and the desired parameters for the visualization (size of the window and size of the glyph). Derived values from this information are sent to the reasoner via a web

[1] http://www.cs.uns.edu.ar/~dku/scatterd3/
[2] http://cs.uns.edu.ar/vizws?wsdl

```
(in-knowledge-base viz test-viz)
(signature
:atomic-concepts
(badWorG badWandG badW badG bad good
cond1 cond2 cond3 cond4 cond5 cond6 cond7
cond8 cond9)
:attributes
( (real WG) (real Wmax1)) (real WmaxG) (real
W1) )
:individuals( p )
:objects( cWG cWmax1 cWmaxG cW1)
)
(equivalent bad (< Wmax1) 0.67 ) )

(equivalent good (> WG 0.67 ) )

(equivalent cond1 (< WG 0.67 ) )
(equivalent cond3 (> WmaxG 0.67 ) )
(equivalent cond4 (and cond3 (> W1 0.67 ) ) )
(equivalent badWorG (and cond4 cond1 ) )

(equivalent cond5 (< WmaxG 0.67 ) )
(equivalent cond6 (< W1 0.67 ) )
(equivalent cond7 (and cond5 cond6 ) )
(equivalent badWandG (and cond2 cond7 ) )

(equivalent cond8 (and cond1 cond6 ) )
(equivalent badW (and cond8 cond3 ) )

(equivalent cond9 (and cond1 cond5 ) )
(equivalent badG (and cond9 (> W1 0.67 ) ) )

(constrained p cWG WG)
(constrained p cWmax1 Wmax1)
(constrained p cWmaxG WmaxG)
(constrained p cW1 W1)
```

Figure 4: Ontology that contains the interpretation of the visibility metric values. This ontology was created using RacerPro [38].

```
(constrained p cWG 0.9801111264686102)
(constrained p cWmax1 0.9999748556415142)
(constrained p cWmaxG 0.9972653458047721)
(constrained p cW1 0.999813956777575)
```

Figure 5: Instances corresponding to a dataset with 508 items, a 400-pixel window, 5-pixel glyphs and 740 pixels of available visible space to draw the scatterplot.

service. The reasoner needs to know the metric values cWG, $cW_{max}1$, $cW_{max}G$ and $cW1$. From the ontology and these values, the reasoner derives a conclusion that represents the metric interpretation. This response is given back to the user through the web page interface. Figure 6 shows the architecture of the web page and Figure 7 shows its state chart diagram.

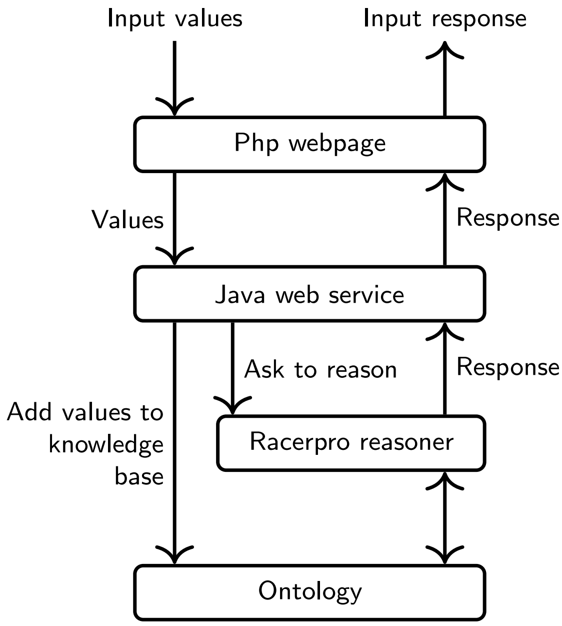

Figure 6: Main architecture of our application.

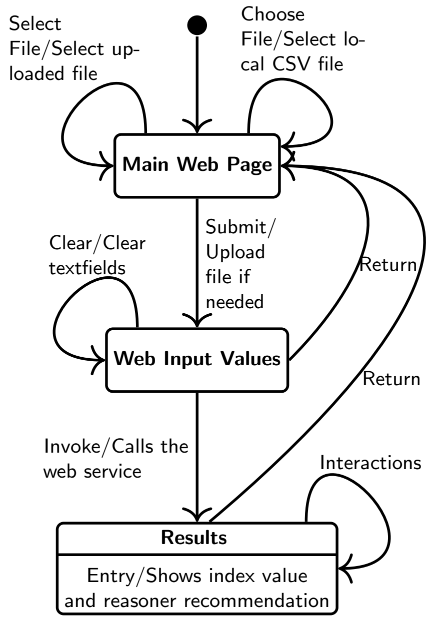

Figure 7: State chart diagram of the web page used in this case study.

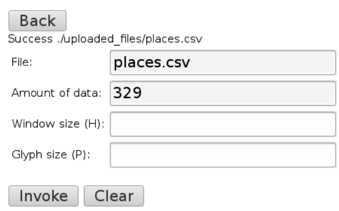

Figure 8: After selecting the file (e.g. places.csv [39]), the user selects the desired window's size and glyph's size.

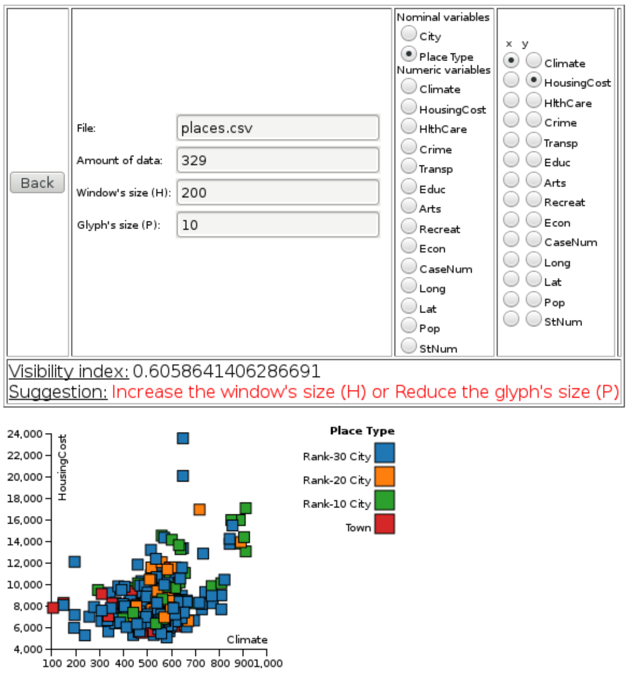

Figure 9: If the user selects to visualize the data in a 200-pixel window with a 10-pixel glyph, the system recommends to increase the former or decrease the latter.

Our research work can also be accessible through our web service. The service takes four integers as input parameters which correspond to the values calculated for cWG, $cW_{max}1$, $cW_{max}G$ and $cW1$ as described earlier. The reasoner response is returned as a string.

Once the CSV file to visualize is selected, the user can choose the size of the scatterplot and the size of the glyph to use. Figure 8 shows and example where the 329 items of the places.csv dataset were loaded. From this information the system shows the resultant visualization and a conclusion about it. This conclusion can be that the visualization is fair enough, that it is improvable or that it cannot be improved. In case the visualization is improvable, the system gives a suggestion on how to accomplish it. Figure 9 and Figure 10 show the same dataset with different configurations, and therefore different suggestions. Figure 9 shows an example where a user selects to visualize the data in a 200-pixel window with a 10-pixel glyph, therefore the system recommends to increase the former or decrease the latter. Following the system suggestion, the user selects a bigger window (a 500-pixel window) and then the system concludes that the visualization is fair enough (see Figure 9). For both cases, Figure 11 and Figure 12 show the respective values that are necessary to include in the knowledge base.

4 Conclusions and future work

A successful visualization allows the user to gain insight into the data in an effective way. Even with today's visualization systems that provide the user a considerable control over the visualization process, it can be difficult to produce a fair visualization. In most cases, the only way to evaluate the quality of a visualization is to create it.

This work presents a semantic-based prediction of data visibility in scatterplots that also suggests how to improve the representation. The suggestion is automatically achieved, without user intervention. We used a metric to calculate data visibility, and its interpretation and further suggestion were achieved by a knowledge base (ontology + individual instances) and a semantic reasoner.

This work is a contribution in the construction of a unified semantics for the visualization process, hence in the creation of a visualization system that automatically assist the user in the configuration and design of visualizations. This visualization system should ensure that, even if the user is not an expert in the field, the generated visualization is a fair one.

As future work, we plan to improve the reasoner response adding more instances of acceptance and/or rejection. We also aim to expand the reasoner application domain by including additional visualization techniques and more metrics into the ontology.

Acknowledgements

This work was partially funded by PGI 24/N050, PGI 24/ZN33 and PGI 24/N037, Secretaría General de Ciencia y Tecnología, Universidad Nacional del Sur, Bahía Blanca, Argentina.

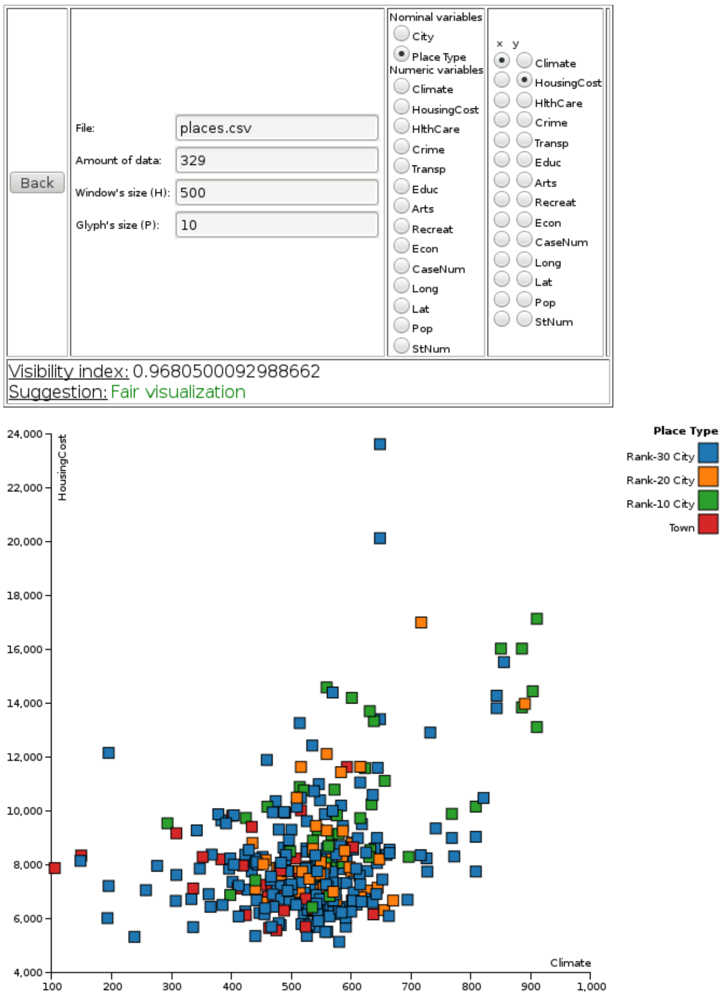

Figure 10: Following the system suggestion presented in Figure 9, the user selects a bigger window (a 500-pixel window) and then the system concludes that the visualization is fair enough.

(constrained p cWG 0.6058641406286691)
(constrained p $cW_{max}1$ 0.99999540124056270)
(constrained p $cW_{max}G$ 0.9962311670015127)
(constrained p $cW1$ 0.9992098249422849)

Figure 11: Necessary instances in the knowledge base to evaluate the visualization of a scatterplot of 329 data items, in a 200-pixel window with a 10-pixel glyph. The maximum window's size is the available visible space in the browser below the table, in this particular case, this value was 973 pixels. These instances correspond to the configuration from Figure 9.

(constrained p cWG 0.9680500092988662)
(constrained p $cW_{max}1$ 0.99999540124056270)
(constrained p $cW_{max}G$ 0.9962311670015127)
(constrained p $cW1$ 0.999959880876266)

Figure 12: Necessary instance in the knowledge base to evaluate the visualization of a scatterplot of the same 329 data items with a 10-pixel glyph but in a 500-pixel window. The maximum window's size was also 973 pixels. These instances correspond to the configuration from Figure 10.

Competing interests

The authors have declared that no competing interests exist.

References

[1] W. J. Schroeder, B. Lorensen, and K. Martin, *The visualization toolkit: an object-oriented approach to 3D graphics*. Kitware, 2004.

[2] S. Escarza, M. L. Larrea, D. K. Urribarri, S. M. Castro, and S. R. Martig, "Integrating semantics into the visualization process," in

Scientific Visualization: Interactions, Features, Metaphors (H. Hagen, ed.), vol. 2 of *Dagstuhl Follow-Ups*, pp. 92–102, Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2011.

[3] E. Bertini, D. Keim, and A. Tatu, "Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization," *IEEE Transactions on Visualization & Computer Graphics*, vol. 17, pp. 2203–2212, 09 2011.

[4] H. J. Schulz, M. Angelini, G. Santucci, and H. Schumann, "An enhanced visualization process model for incremental visualization," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 7, pp. 1830–1842, 2016.

[5] D. K. Urribarri and S. M. Castro, "Prediction of data visibility in two-dimensional scatterplots," *Information Visualization*, vol. 16, no. 2, pp. 113–125, 2017.

[6] D. J. Duke, K. W. Brodlie, D. A. Duce, and I. Herman, "Do you see what i mean?," *IEEE Computer Graphics and Applications*, vol. 25, no. 3, pp. 6–9, 2005.

[7] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pp. 336–343, IEEE, 1996.

[8] S. K. Card and J. Mackinlay, "The structure of the information visualization design space," in *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pp. 92–99, IEEE, 1997.

[9] E. H. Chi, "A taxonomy of visualization techniques using the data state reference model," in *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pp. 69–75, IEEE, 2000.

[10] M. Tory and T. Möller, "A model-based visualization taxonomy," *Computing Science Department, Simon Fraser University, Technical Report No. TR*, vol. 6, 2002.

[11] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry, "Task taxonomy for graph visualization," in *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1–5, ACM, 2006.

[12] K. W. Brodlie and N. Mohd Noor, "Visualization notations, models and taxonomies," in *Theory and Practice of Computer Graphics 2007, Eurographics UK Chapter Proceedings*, pp. 207–212, Eurographics Association, 2007.

[13] Z. Xu, H. Chen, and Z. Wu, "Applying semantic web technologies for geodata integration and visualization," in *International Conference on Conceptual Modeling*, pp. 320–329, Springer, 2005.

[14] T. Nguyen, T. Ho, and D. Nguyen, "Data and knowledge visualization in knowledge discovery process," *Recent Advances in Visual Information Systems*, pp. 135–159, 2002.

[15] K. Thellmann, M. Galkin, F. Orlandi, and S. Auer, "Linkdaviz – automatic binding of linked data to visualizations," in *The Semantic Web - ISWC 2015* (M. Arenas, O. Corcho, E. Simperl, M. Strohmaier, M. d'Aquin, K. Srinivas, P. Groth, M. Dumontier, J. Heflin, K. Thirunarayan, K. Thirunarayan, and S. Staab, eds.), (Cham), pp. 147–162, Springer International Publishing, 2015.

[16] R. Minu and K. Thyagharajan, "Semantic rule based image visual feature ontology creation," *International Journal of Automation and Computing*, vol. 11, no. 5, pp. 489–499, 2014.

[17] C. G. Healey, R. S. Amant, and J. Chang, "Assisted visualization of e-commerce auction agents," in *Graphics Interface*, vol. 1, pp. 201–208, 2001.

[18] M. Golemati, C. Halatsis, C. Vassilakis, A. Katifori, and U. Peloponnese, "A context-based adaptive visualization environment," in *Information Visualization, 2006. IV 2006. Tenth International Conference on*, pp. 62–67, IEEE, 2006.

[19] D. Koop, C. E. Scheidegger, S. P. Callahan, J. Freire, and C. T. Silva, "Viscomplete: Automating suggestions for visualization pipelines," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1691–1698, 2008.

[20] O. Gilson, N. Silva, P. Grant, and M. Chen, "From web data to visualization via ontology mapping," *Computer Graphics Forum*, vol. 27, no. 3, pp. 959–966, 2008.

[21] K. Brodlie, D. Duce, D. Duke, *et al.*, "Visualization ontologies: Report of a workshop held at the national e-science centre," tech. rep., e-Science Institute (April 2004), 2004.

[22] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. Van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver, "Data, information, and knowledge in visualization," *IEEE Computer Graphics and Applications*, vol. 29, no. 1, 2009.

[23] E. Kalogerakis, S. Christodoulakis, and N. Moumoutzis, "Coupling ontologies with

graphics content for knowledge driven visualization," in *Virtual Reality Conference, 2006*, pp. 43–50, IEEE, 2006.

[24] A. Unwin, M. Theus, and H. Hofmann, *Graphics of large datasets: visualizing a million*. Springer Science & Business Media, 2006.

[25] E. J. Wegman, "Huge data sets and the frontiers of computational feasibility," *Journal of Computational and Graphical Statistics*, vol. 4, no. 4, pp. 281–295, 1995.

[26] M. Tory and T. Möller, "Rethinking visualization: A high-level taxonomy," in *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pp. 151–158, IEEE, 2004.

[27] M. S. T. Carpendale, "Considering visual variables as a basis for information visualisation," tech. rep., University of Calgary, Calgary, AB, 2003.

[28] P. Rautek, S. Bruckner, and E. Gröller, "Semantic layers for illustrative volume rendering," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1336–1343, 2007.

[29] M. Gerl, P. Rautek, T. Isenberg, and E. Gröller, "Semantics by analogy for illustrative volume visualization," *Computers & Graphics*, vol. 36, no. 3, pp. 201–213, 2012.

[30] C. R. Salama, M. Keller, and P. Kohlmann, "High-level user interfaces for transfer function design with semantics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 5, 2006.

[31] M. L. Larrea, S. R. Martig, and S. M. Castro, "Semantic based visualization: a first approach," in *X Workshop de Investigadores en Ciencias de la Computación*, pp. 273–277, 2008.

[32] M. Larrea, S. Martig, and S. Castro, "Semantics-based color assignment in visualization," *Journal of Computer Science & Technology*, vol. 10, no. 1, p. 5, 2009.

[33] M. Larrea, S. Martig, and S. Castro, "Semantics-based visualization building process," in *Computer Science & Technology Series. XVI Argentine Congress of Computer Science Selected Papers*, pp. 145–155, Editorial de la Universidad de La Plata, 2011.

[34] M. Larrea, S. Escarza, D. Urribarri, S. Castro, A. Herrera, C. Escobal Blanco, and G. Cerúsico, "Ontologías y semántica en el proceso de visualización," in *XVI Workshop de Investigadores en Ciencias de la Computación*, pp. 290–295, 2014.

[35] S. Castro, M. Larrea, D. Urribarri, L. Ganuza, and S. Escarza, "Métricas, técnicas y semántica para la visualización de datos," in *XX Workshop de Investigadores en Ciencias de la Computación*, pp. 361–365, 2018.

[36] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, Stanford Knowledge Systems Laboratory, march 2001.

[37] M. A. Abbas, W. F. W. Ahmad, and K. S. Kalid, "Semantic web technologies for pre-school cognitive skills tutoring system.," *J. Inf. Sci. Eng.*, vol. 30, no. 3, pp. 835–851, 2014.

[38] V. Haarslev, K. Hidde, R. Möller, and M. Wessel, "The racerpro knowledge representation and reasoning system," *Semantic Web*, vol. 3, no. 3, pp. 267–277, 2012.

[39] R. Boyer and D. Savageau, *Places rated almanac: Your guide to finding the best places to live in America*. Rand McNally & Co., 1985.

Citation: D. Urribarri, M. Larrea and S. Castro. *Automatic suggestions to improve the quality of scatterplots during its creation: A case study of ontology and semantic reasoning applied to visualization*. Journal of Computer Science & Technology, vol. 19, no. 2, pp. 100–109, 2019.

DOI: 10.24215/16666038.19.e10

Received: December 07, 2018. **Accepted:** June 12, 2019.