Techniques for Improved Speech-Based Access to Diagrammatic Representations

Tomas Murillo-Morales^(✉) and Klaus Miesenberger

Johannes Kepler University, Linz, Austria

{Tomas.Murillo_Morales, Klaus.Miesenberger}@jku.at

Abstract. Natural language interfaces (NLIs) are a novel approach to non-visual accessibility of diagrammatic representations, such as statistical charts. This paper introduces a number of methods that aim to compensate for the lack of sight when accessing semantically-enhanced diagrams through a Web-based NLI.

Keywords: Diagram semantics · Accessibility · Natural language interface

1 Introduction

Diagrams help humans with problem solving and understanding. Graphic representations of a problem exploit the natural perceptual, cognitive, and memorial capacities of the human visual system [1]; the brain finds it easier to process information if it is presented as an image rather than as words or numbers [2]. Therefore, properly graphically displaying information simplifies its understanding, assisting the sighted reader in building their mental model of the problem and easing the problem solving process. For instance, whereas looking through a large numerical table takes a considerable amount of mental effort, the same information properly displayed in a visual manner (e.g. in a scatterplot) may be grasped in a matter of seconds.

Diagrams are thus commonplace in science, journalism, finance, and many aspects of daily life. However, blind persons are generally excluded from accessing them. Visually impaired persons need to make use of assistive technologies and alternative representations of visually perceivable information, ranging from simple textual descriptions of the graphic to practically equivalent tactile transcriptions. Nevertheless, current non-visual methods of accessing diagrams are either too simplistic (linear text descriptions, audification approaches, surface haptic displays), too expensive or cumbersome to produce (tactile shape displays, tactile hard-copies, multimodal approaches, force-feedback devices), or still in their early infancy (vibrotactile displays, interactive systems, natural-language-based systems). This lack of comprehensive, inexpensive, and easy-to-use approaches has resulted on graphical information being described as the “last frontier in accessibility” [3].

Communicative images are a novel approach to accessibility of graphics that may help solve many of these drawbacks for a sizable number of graphic domains. They can be defined as two-dimensional objects (graphics) integrated with a dialogue interface which are equipped with an associated knowledge database [4]. Image semantics are

described in a structured way by means of ontologies using semantic categories, their properties and relationships. A dialogue interface (natural language interface, NLI) then lets users efficiently retrieve the underlying knowledge by means of natural language queries. NLIs do not require any specific software or hardware, and therefore communicative images can be accessed online the same way blind people are used to navigate the Web in a usual manner. Therefore, communicative graphics have the potential of emerging as a holistic approach to the accessibility of many conventional graphics, freeing users from having to employ cumbersome, domain-dependent methods.

The serial nature of speech has an obvious disadvantage over tactile or some sonification approaches which do not impose such a cognitive load on the reader's processing capabilities. Interestingly enough, in spite of this fact blind readers much prefer speech feedback when navigating e.g. link diagrams, and other means are preferred only rarely [5]. However, being a novel approach, communicative graphics have not been thoroughly evaluated, and so far their domain of applicability has been limited to simple photographic content accessed by means of controlled natural language queries.

We have previously introduced broadening efforts to include domains other than photography in communicative images by defining a hierarchical formal semantic knowledge base in the form of ontologies underpinning visualization domains i.e. diagrams along with an accessible Web user interface to semantically-enhanced graphics based on natural language (Fig. 1) [6]. In addition, an authoring tool for seamless creation of semantic formal markup given a vector graphic has been developed, allowing authors to associate ontological instances and property occurrences to any number of a given vector graphic's constituent elements such as paths, shapes, or text elements [7].

Fig. 1. Block diagram of our current prototype of a framework for speech-based access to graphics

The hierarchical knowledge base is made up of four main layers [8]: (1) an upper ontology describing syntactic aspects of visualization common across domains, (2) several domain ontologies underpinning broad visualization domains, such as statistical charts and maps, (3) task ontologies describing analytical low-level tasks that can be performed on a given domain which may then be combined to form high-level activities, and (4) user/system ontologies defining user-made and system annotations on individual graphic elements that ease non-visual access to a diagram, allowing users to customize diagram navigation to their specific needs.

A server-side controller, written in Python, is in charge of inferring the task to be performed according to which keywords are present in a user's subquery after a Natural Language Processing module has tagged each part-of-speech component of the original query. The controller then tries to select the most appropriate task among those low-level tasks that can be performed on the current domain according to the semantic information stored in the task ontology level. If relevant, the operands and operators of the end SPARQL query are also fetched by the controller from the provided list of keywords. At the moment, different input text fields, one per available low-level task, have been implemented in the Web user interface in order to simplify the natural language processing [6]. In the end prototype we are planning to combine them into a single text field with more profound back end language understanding capabilities.

In this paper, we outline some of the components of the semantic hierarchy that aim to improve the efficiency and user experience of navigating and performing analytical tasks on diagrams in a non-visual manner, and discuss their implementation into a Web-based accessible interface.

2 Cognitive Benefits of Diagrammatic Representations

The initial phase before building the supporting knowledge base and its accompanying user interface consisted on an exhaustive study on the cognitive benefits that diagrammatic representations offer sighted persons in problem solving. Only after understanding these cognitive benefits it could be attempted to implement non-visual methods that compensate for the lack thereof.

Some of the particular reasons why sighted persons prefer diagrammatic representations to linear descriptions for problem solving are [9, 10]:

1. **Resemblance preservation:** Diagrams possess the ability of directly resembling what they represent, as opposed to linear descriptions, in which the reader has to imagine the entities that are being described. This resemblance may take place in two different ways:
 - a. **Literal resemblance:** many diagrams literally preserve the topology and geo-metric relationships of what is being represented, including the relative size and position between elements. This is the case in pictures and most maps.
 - b. **Homomorphic resemblance:** diagrams using this type of resemblance use topology and geometric relationships metaphorically to represent abstract relationships with similar abstract properties.

2. **Indexing:** diagrams avoid large amounts of search for the elements needed to make a problem-solving inference by grouping all information that is used together, which avoids the need to use and match symbolic elements. Besides locational indexing, they also allow two-dimensional indexing, this is, the ability of locating an object by its coordinates on the two-dimensional plane. For example, two-dimensional indexing allows a sighted reader to find the values of a given point in a scatterplot by matching its location to horizontal and vertical positions on the plot's axes.
3. **Breadth-then-depth search:** when visualizing data, lots of information can be acquired in parallel from a wide area. The perceptual abilities of the human sight allow persons to scan, recognize, and recall images rapidly. This leads to the ability of intentionally focusing the attention into a narrower, more detailed part of the whole image that the reader finds especially interesting. The capacity to gain a quick overview first, followed by the possibility of effortlessly filtering out uninteresting elements and obtaining details of the relevant ones makes visual data display a much more efficient means of retrieving information than linear representations. This cognitive aspect of visualization is very well known in data visualization research, from where the so-called "visual information-seeking mantra", that aims to replicate it, stems: overview first, zoom and filter, then details on demand [11].
4. **Computational off-loading:** Diagrams automatically support a large number of perceptual inferences by means of sight, which are extremely easy for sighted humans. Perceptual inferencing depends on the visual and spatial properties of the diagram encoding the underlying information in a way that capitalizes on automatic processing to convey the intended meaning. For example, the relationship between the two variables in a line chart (i.e. the trend) can be inferred immediately by means of sight. Diagrams also allow mental animation, i.e. mentally activating components in the representation of the system in a serial manner [12].

These cognitive advantages make proper data display a critical aspect to data analysis. Properly displaying information can make a difference by enabling people to understand complex matters and find creative solutions [2]. For example, it has been proven that students learn better when textual materials are enhanced with the inclusion of didactical graphics [13]. This applies not only to information retention, but also to deep comprehension, which finally turns into better solving skills [14]. It is thus paramount for technologies that aim to enable non-visual access to graphics to provide their users with alternative methods making up, at least to a certain extent, for the loss of these critical benefits.

3 Dialogue-Based Access to Diagrams: The User/System Level

The first three levels of our hierarchy depicting the semantics of diagrams were previously discussed in previous publications by the authors: [6, 8]. Here, we focus on semantic markup added to the supporting ontologies at the domain and user/system levels in order to make up for the loss of cognitive benefits presented in the previous section.

Literal Resemblance Preservation: Most visual attributes with the ability of carrying information in diagrams, such as shape, size, and color, cannot be accurately represented by aural means. Therefore, we have resorted to describing them by means of textual descriptions. A number of visual attributes have been included as datatype properties of the upper ontology in our hierarchy, and thus any graphic object may be enhanced with semantics depicting their visual properties. A user can then inquire information about objects sharing a certain visual attribute (e.g. “blue bars” in a bar chart) or ask about the visual properties of objects (e.g. “color of current bar”). When a user performs a filtering query according to visual attributes, the system controller detects it and chooses a standard low-level filtering task with the visual attribute as the operand and “equals” as its implicit operator (unless a label with the same value as the attribute exists, in which case the labelled elements are given priority). The filtering task itself is then performed by the task ontology handler, which is further delegated to the domain and upper level ontology handlers until the relevant semantic instances are finally obtained from the knowledge base via a SPARQL query, sent back to the controller and then output to the user via a simple natural language processing module.

Moreover, any graphical object might be given a user- or author- defined label with extra information (e.g. a certain object in a map might be pre-tagged as “Germany”, or users may choose to add a custom tag themselves). More on this can be read later. Special tasks may be also performed on objects carrying certain attributes. For example, of the Y coordinate value of Graphic_Object A is smaller than that of Graphic_Object B, A is said to be below B.

Homomorphic Resemblance Preservation: Metaphorical relationships between objects are included at the domain ontology level as object or datatype properties. For instance, a marriage between two persons, depicted in a family tree link diagram by a horizontal link connecting them, may be represented by the Family Tree ontology with an object property occurrence: (A, married_to, B). User tasks based on these properties can then be added at the Task ontology level. For example, the query “who is married to A?” would be interpreted as a Filtering low-level task that returns all subjects X of the triples in the domain ontology having the form (X, married_to, A).

Indexing: Graphic elements may be assigned ordinal values when they are meant to be navigated in a given order. For example, points in a line chart can be annotated from left to right (by adding occurrences of the has_order datatype property) so the user can navigate the entire line and perform tasks on each point consecutively. Navigation can be performed in a hierarchical fashion as well, since graphic objects (elementary or composite) belonging to composite objects are clearly identified in the supporting ontology in a recursive manner. Users may then issue “go to next/previous element” or “go one level up/down” queries in order to sequentially navigate through the graphic. Clusters of elements may be labelled using a common label for all elements in the cluster. These indexing features are implemented as object and datatype properties that support analytical and navigational tasks on the graphic. For instance, the following query performed on a bar chart: “average sales in January” can be performed as the system is able to first filter those bars labelled by the “January” label.

When a user moves from one graphic element to another, the “is_current_node” datatype property of the system ontology level has its subject updated to the current

element by the system's controller. The user may at any time issue a "where am I?" query, and every relevant property of the current element will be output, such as labels and user annotations, in order to help blind users with orientation while navigating complex graphics.

Breadth-then-depth search: A summary of the whole graphic can be requested by the user for certain domains, upon which the most salient features of the graphic are given to the user in order to obtain an initial, high-level overview of its content. For example, a bar chart overview includes its title, what it axis represent, the contents of its legend, a list of its labels, its extreme values, and, if bars are grouped, the average value of each group. Once the user is familiar with the general aspects of the graphic, low-level tasks can be performed [6], or the user may choose to navigate the graphic in a sequential, object-by-object manner, as previously described. Summaries are automatically computed by the system's controller on a domain-dependent basis.

Computational off-loading: The author's intention when designing the graphic, such as the general trend of a line chart, is a vital piece of information that sighted users can generally infer from observing the graphic. This implicit information can be explicitly marked in the supporting ontology so blind readers acquire quick insight of the graphic's semantics, as previously described in [6].

Another powerful method for offloading the user's working memory consists on user-defined home nodes. A home node is a distinctive element of the graph, analogous to a landmark, that is used as a base for exploration and to which users may return when lost [15]. The user may select any graphic object of the diagram of special interest as the current 'home node', and the system controller will update the subject of the "is_home_node" property occurrence to the current individual. The output to any further query performed by the user will try to relate to the home node. For instance, in bar charts, if a bar is chosen as the current home node every further query that outputs a numerical value will be compared to the home node's value. In weighted graphs, the sum of weights of the shortest path from the home node to the current node may be given, and so on. This way, relationships between distant elements of interest can be quickly inferred in a non-visual manner without the user having to remember every intermediate step.

In addition, users may add a number of personal textual annotations to the current graphical instance during exploration of the diagram. These custom tags are then added by the controller to their graphical object instance via the "has_user_label" datatype property. When this element's labels are output to the user during navigation or after performing an analytical task, the user-defined labels are given as well. This helps the user with remembering previously visited elements and understanding the general structure of the graphic.

Finally, a number of navigational shortcuts are provided to the user to provide some computational unload while navigating a complex diagram. Besides the "where am I?" command, quick jumps to the first and last node, as well as the home-node, are also provided by the system. Moreover, the user may jump to the nodes with the highest and lowest values of the diagram e.g. the tallest and shortest bars in a bar chart. In the future, we would also like to let the user ask about the evolution of a section of a graph e.g. whether a line chart has an increasing/decreasing tendency within two selected points.

4 Conclusions and Further Work

This paper has introduced a number of techniques for improved natural-language-based access to diagrams, supported by a semantic knowledge base in the form of hierarchical ontologies, that aim to make up for a number of known cognitive benefits that visualizing information provides to sighted users. Thus far we have applied most of these techniques to bar charts, and have implemented a first prototype of an accessible Web interface that groups common tasks that may be performed on them. We are now working on evaluating the current prototype with sighted and blind users. Evaluation will likely reveal which of these techniques can be improved and whether users find them useful in supporting analytical tasks performed on bar charts.

Besides adding more supported domains to the current prototype (line charts, scatterplots, link diagrams, etc.) we are also working on reinforcing it with other non-visual methods to improve navigation. For instance, we would like to find out whether non-speech sounds, such as short beeps or musical notes, can be useful in recognizing different landmarks, previously visited nodes, or some other salient features of a diagram by visually impaired users.

References

1. Lewandowsky, S., Spence, I.: The perception of statistical graphs. *Sociol. Methods Res.* **18** (2–3), 200–242 (1989)
2. Cukier, K.: Show me. New ways of visualising data. *The Economist*, February 2010. <http://www.economist.com/node/15557455>. Accessed 5 July 2017
3. Gardner, J.: Can mainstream graphics be accessible? *Inf. Technol. Disabil. E-J.* **14**(1) (2014)
4. Kopecek, I., Oslejsek, R.: Communicative Images. In: Dickmann, L., Volkmann, G., Malaka, R., Boll, S., Krüger, A., Olivier, P. (eds.) *SG 2011. LNCS*, vol. 6815, pp. 163–173. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22571-0_19
5. Ferres, L., Lindgaard, G., Sumegi, L., Tsuji, B.: Evaluating a tool for improving accessibility to charts and graphs. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **20**(5), 28 (2013)
6. Murillo-Morales, T., Miesenberger, K.: Non-visually performing analytical tasks on statistical charts. In: Cudd, P., de Witte, L. (eds.) *Studies in Health Technology and Informatics*, vol. 242, pp. 339–346. IOS Press (2017)
7. Murillo-Morales, T., Plhák, J., Miesenberger, K.: Authoring semantic annotations for non-visual access to graphics. *J. Technol. Persons Disabil.* **6**, 398–413 (2018)
8. Murillo-Morales, T., Miesenberger, K.: Ontology-based semantic support to improve accessibility of graphics. In: Sik-Lányi, C., Hoogerwerf, E.J., Miesenberger, K. (eds.) *Studies in Health Technology and Informatics*, vol. 217, pp. 255–260. IOS Press (2015)
9. Larkin, J.H., Simon, H.A.: Why a diagram is (sometimes) worth ten thousand words. *Cogn. Sci.* (1987). <https://doi.org/10.1111/j.1551-6708.1987.tb00863.x>
10. Goncu, C., Marriott, K., Aldrich, F.: Tactile Diagrams: Worth Ten Thousand Words? In: Goel, A.K., Jamnik, M., Narayanan, N.H. (eds.) *Diagrams 2010. LNCS (LNAI)*, vol. 6170, pp. 257–263. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14600-8_25
11. Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: *IEEE Symposium on Visual Languages*, Boulder, 3–6 September, pp. 336–343. IEEE Computer Society Press (1996). <https://doi.org/10.1109/vl.1996.545307>

12. Hegarty, M.: Mental animation: inferring motion from static displays of mechanical systems. *J. Exp. Psychol.: Learn. Mem. Cognit.* **18**(5), 1084–1102 (1992)
13. Levin, J., Anglin, G., Carney, R.: On empirically validating functions of pictures in prose. In: Willows, D., Houghton, H.A. (eds.) *The Psychology of Illustration: Basic Research*, vol. 1, pp. 51–85. Springer, New York, NY, USA (1987). https://doi.org/10.1007/978-1-4612-4674-9_2
14. Hochpöchler, U., Schnotz, W., Rasch, T., Ullrich, M., Horz, H., McElvany, N., Baumert, J.: Dynamics of mental model construction from text and graphics. *Eur. J. Psychol. Educ.* (2013). <https://doi.org/10.1007/s10212-012-0156-z>
15. Brown, A.J.: *Non-Visual Interaction with Graphs*. Ph.D. thesis, University of Manchester (2008)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.