

How to model a customized visualization

Mohamed Mouine and Guy Lapalme

RALI-DIRO

University of Montreal

Montreal, Qc Canada

{mouinemo,lapalme}@iro.umontreal.ca

Abstract—Modelling a customized view is a daunting task that takes into account several parameters, the most important being the user profile. In this paper we study many facets of modelisation and the issues that must be taken into account. We propose an approach in four steps to model a personalised visualization: extract data, guess user's needs and preferences, generate a visualization and finally improve it by user evaluation. It has been applied to the context of weather information in Canada.

Keywords—visualization; personalized; evaluation; weather; user profile; clustering.

Keywords—

I. INTRODUCTION

In recent years, there has been an explosion in the volume of data generated in all fields of knowledge, complicating their analysis and exploration. Data mining helps locate the necessary information. Information visualization helps better infer the information. Information display seeks to solve the problem of representing different types of data, so that it can be easily transmitted and interpreted.

Information visualization has become a necessity and not a luxury. Several fields use it to simplify display of the web search results [1], [2]

There are two types of problems in modelling and visualization. The first is finding a method to display all the information; the second is selecting and displaying only the information needed. In this paper we propose a method to solve this second type of problem. In the context of weather information. Environment Canada (EC) produces 26 MB of data (twice a day) describing (numerically) weather conditions throughout Canada. From this, EC prepares, also twice a day, about 1000 reports. Each report displays a summary of the meteorological information about a region of Canada. Given the large amount of information produced, these summaries contain only part of the information that users could have if they could access all the generated data by EC.

Our approach can be applied to the generation of customized visualization from a large mass of information. The first step of our approach is extracting interesting facts from the mass of raw data that we have as input. This issue is explained in section II. The main element for the generation of a customized visualization is a user profile. The needs and

preferences of the user are discussed in Section III. Section IV deals with the creation of the visualization. In section V, we propose improving the result, by evaluating the result from users' point of view and recomputing the needs and preferences of this profile according to the users' interaction with the visualization. Figure 1 gives the overall architecture of our approach.

II. WHAT IS THE INPUT?

The main goal of any visualization is to help a user analyse a large amount of information. Information may have several forms (one-dimensional data, two-dimensional data, multidimensional data, hierarchies and graphs, algorithms and software, etc.) and visualization can be in different forms using several techniques (standard 2D/3D displays, geometrically transformed displays, icon-based displays, dense pixel displays, stacked displays, etc.) [3]

In our application (weather) we can extract information from two file formats (XML and CSV) from Environment Canada's website¹. As shown in step ① in figure 1, we have created a separate module responsible for searching, cleaning, transforming and modelling data to extract the information needed for our purpose. This module is queried by our system to get up to date information from files produced by EC. For the rest of this paper, we will not deal with data extraction. We consider that our system, using the data extraction module, has the information the users need.

III. WHO IS THE RECEIVER?

The generation of a visualization takes into account the users' profile, i.e their needs and preferences in order to generate a satisfactory visualization. This section corresponds to step ② of fig.1. If a user profile has been given, the task is easier, otherwise the system must guess their needs and preferences based on other available information. Automatic profiling is common in the field of e-commerce and advertising, for guessing what the user thinks in order to offer products of interest. To achieve this, they consider all (legal) methods including cookies and the historical results of search engines.

In our project, the task is more difficult since we cannot rely on cookies. Our visualization system is intended to be

¹<http://dd.weatheroffice.ec.gc.ca/meteocode/que/>

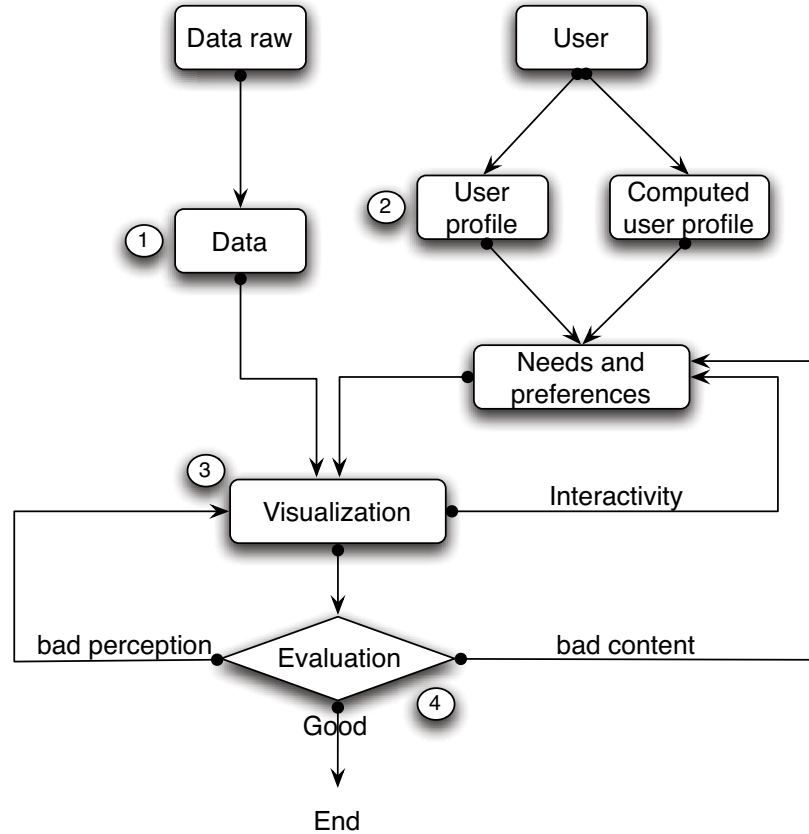


Figure 1. Our approach to creating a good personalized visualization in four steps; ① preparing data from row data; ② determining needs and preferences of the user using a user profile and/or a computed user profile; ③ creating the visualization with which the user can interact according to her needs and preferences; ④ evaluating to improve visualization

installed on the website of Environment Canada, owned by the Government of Canada for which the privacy of users is a priority. This precludes our use of cookies.

We gather as much information about the user (location, date, time, language preference in the browser...) without touching her private life. Using this information we search our database of former users who used the system to group users with similar information. We weight thereafter the degree of similarity according to the distance in the cluster between the current user and users in the same cluster. The degree of profile similarity gives hints about preferences of the current user.

The goal of a visualization is to convey to users a wealth of data that they will be able to translate (understand) and analyse. Perception [4], the way in which humans interpret an object, realized in two steps [5]:

- **Pre-attentive treatment (unintentional):** For many years, researchers have studied how the human visual system analyses images. An important result was the initial discovery of a limited set of visual properties that are detected very quickly and accurately by the low

level visual system. These properties were originally called pre-attentive because their detection seemed to precede attention. We now know that attention plays a vital role in what we see, even at this early stage of vision.

- **Perception:** it is at this level that visual effects captured in the previous steps will be transformed into structured objects. This step is slower than the previous one. It is in this step that the user will find all the details of the visualization.

In [6] we proposed a method to personalize a visualization by clustering user's profiles. The visualizations are generated dynamically by our system that saves each final configuration chosen by a user. The idea of the approach is to aggregate similar former users with the current one. Clustering is used for determining the settings of users. The distance between the current user and others in a cluster will be used to compute the similarity. The similarity rate gives us an idea of what could be the needs and preferences of the current user taking into account similarity with others. The similarity rate is used to weight needs and preferences.

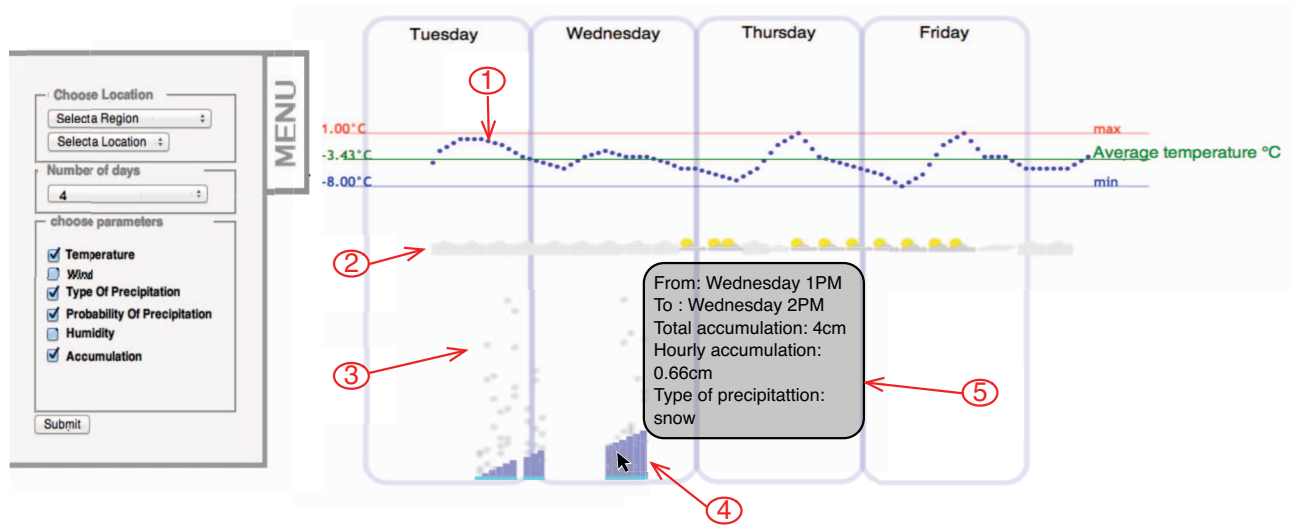


Figure 2. Generated visualization showing some types of information that can be selected by the user. Circled numbers are added here for reference purposes. ① Temperature: how temperature is displayed gives an idea of the trend. Maximum and minimum line show limits of the temperature for the chosen period. ② Cloud cover. ③ TOP (Type Of Precipitation): rain or snow. Quantity displayed is proportional to the quantity provided. ④ Accumulation: shows the accumulation zone and total rainfall (mm) or snow (cm). ⑤ The user may have more details about aspect by putting the cursor over it (tooltip). Using the preference menu (in left), the user can select a region (Province) and location (city), modify number of the days to display and which parameter to display.

We set a minimum threshold for each criterion (needs and preferences). To decide which needs and preferences will be considered in visualization (step ③ in figure 1), we take into account only results greater than the threshold.

IV. HOW TO CREATE A VISUALIZATION?

We've prepared data (section II) and user's needs and preferences (section III). This section explain the step ③ in figure 1 which has two input: data ① and user's needs and preferences ②.

To create a visualization, Lam et al. [7] propose to start by understanding the environment and work practices and give methods and examples: field observation, interviews and laboratory observation. Agrawala et al. [8] suggest in the first step of their approach for visualization creation to explore existing visualizations in the same field to identify design principles. For the Environment Canada website, we analysed several competitor websites to become familiar with the concepts and principles used in this domain.

In this paper, we focus on the user. Therefore, exploring existing visualizations must take into account user's point of view. Chen [9] studied 10 unsolved problems in this area, the first three being problems from a user-centric perspective:

- **usability** the information contained in the visualization should meet the needs of users; thus the number of

usability studies and evaluations of visualization [10], [11], [12], [13];

- **perception** the principles of perception were incorporated into rendering algorithms in order to optimize rendering computation and produce an ideal visualization from a human point of view and not from a machine standpoint [14];
- **prior knowledge of the user** can be considered a parameter of the second problem. In general, users need two types of knowledge before understanding the message conveyed in the information displayed: how to use a system for displaying information and how to interpret its content.

Based on studying visualizations from competitors' web sites to EC, principles and concepts inferred in this field, using perception principles [4] and interaction techniques [15], our system produces a weather report at location initially inferred from the IP address (see fig.2) with user's needs and preferences computed from the clustering of similar users (explained in section III).

In contrast to reports, containing static summaries, prepared by EC, this visualization contains more detailed of information for the user. This is possible because we extract all the information from in the 26MB of data related to the user's profile. In the visualization shown in fig.2, a user can see ① the temperature: each point is the temperature

of corresponding hour and maximum and minimum line show limits of the temperature for the chosen period, ② cloud cover, ③ the type of precipitation (rain or snow), the probability of precipitation, humidity, the wind (not displayed in this figure) and/or ④ accumulation of a specific location in current and next days. At first glance, the user can see and have an idea of the general trend of the temperature. Maximum and minimum lines show limits of the temperature for the chosen period. The user understands if it will be snowing (or raining), when, how much, how long...If wind is displayed (not shown in fig.2), speed and wind direction are shown by an arrow's direction and size.

We can enrich the visualization by adding computed information (average, maximum, minimum...) to simplify the perception by the user.

V. HOW TO IMPROVE VISUALIZATION?

Generation of a customized visualization is a mixture of art and science. The scientific side is interested primarily in identifying the content of the visualization. On the other hand the artistic side is more interested in "how to display this content". The user perception of good content visualization is the main goal. As our visualization is customized, the contents of the display depends on the user profile.

Although we computed user's needs and preferences, this is a prediction that we hope to be near the real needs and preferences, but sometimes users differ in preferences and in perception. We propose two ways to take into account feedback from users.

A. Interaction

Besides visualization techniques for effective data mining, it is necessary to use some interaction. Interaction techniques [5] allow the data analyst to interact directly with the visualizations and dynamically change them based on exploration targets. Several techniques of interaction [15] can be used: select, explore, reconfigure, encode, abstract/elaborate, filter and connect. For improving visualization we regenerate a new one according to the user choices in the menu in fig.2. We save the final state of the display with the user profile in order to use it in the future to compute needs and preferences of another similar user.

Preference menu (left in fig.2) lets the user modify the visualization, select the region (Province) and location (city), modify number of days and which parameter to display. User may get more details by hovering the cursor over the part of interest. A tooltip will then display more details.

B. Evaluation

We can measure qualitatively user feedback with interviews and quantitatively with usage statistics. We can also perform more formal studies to determine to what extent our visualizations improve the interpretation of information.

Evaluation criteria must quantify the effectiveness of some aspects of the visualization. Tory et al. [16] and Zuk et al. [17] propose an heuristic evaluation. Although this way can provide very good qualitative evaluation, finding appropriate heuristics is hard task. Some examples of evaluation of heuristics are given in [18], [19], [20] and [21].

Our visualization should be expressive and therefore have all the information the user needs and only this information. It must also be efficient to be interpreted quickly and accurately. Generally user feedback is focussed on two aspects: the rate of perception and how much content meets user expectations.

If the feedback shows that users have a difficulty with perception, we have to modify concepts, principles and techniques of visualization generation. If we have a problem with a visualization content then we have to re-compute user's needs and preferences.

To evaluate our application, we will use two techniques: automatic evaluation made by all system's users and professional evaluation made by EC's employers.

1) *Users evaluation*: Users interact with the visualization in order to synchronize it with their needs and preferences. We evaluate visualization content based on user's interaction. To improve future visualization, We take into account this evaluation in clustering result and the step of computing needs and preferences.

- We consider that the last result of the interaction is the best visualization responding needs and preferences of this user. We save all features with user's profile to use it in computing user's needs and preferences.
- If a user does not interact with the visualization, we consider that it contains the appropriate information.

A score is used to decide the influence to apply in the display. The computation of this score is based on:

- If many users change the display and adds that new parameter, that means this parameter was not included in the user's preferences. The score increases the weight of this feature proportionally with the number of users who choose it until it will be displayed for this type of profile.
- If many users modify the visualization by removing one or more parameters, it means that those users having similar profile prefer that this parameter not be included in visualization. The score is decreased proportionally with the number of users who did not like it until it is not displayed for this type of profile.

To evaluate user's perception, we intend also to use a qualitative evaluation by asking users their opinion about visualization.

2) *Professional evaluation*: Evaluation by Users is limited to the ones who participated. Sometimes, this technique does not permit to communicate all the needs, preferences and suggestions. We are collaborating with EC's profes-

sionals to provide more suggestion and evaluation feedback about results.

VI. CONCLUSION

Information visualization helps analyse data in order to understand its contents. In this paper, we propose an approach to model a personalized visualization in four steps. The first one consists understanding, cleaning and extracting visualization's input data. Given that the visualization might be personalized according to the user, the second step is determining user's needs and preferences. Using data from the first step and taking into account needs and preferences guessed, the third step is creating a visualization which is evaluated and improved as a last step. We applied this approach to generate a personalised weather report for EC website users. The resulting visualization takes into account the current user's profile whose needs and preferences are based on former users having a similar profile.

REFERENCES

- [1] T. N. Nguyen and J. Zhang, "A novel visualization model for web search results," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 12, no. 5, pp. 981–988, 2006.
- [2] W. Wiza, K. Walczak, and W. Cellary, "Periscope: a system for adaptive 3d visualization of search results," in *Proceedings of the ninth international conference on 3D Web technology*. ACM, 2004, pp. 29–40.
- [3] D. A. Keim *et al.*, "Information visualization and visual data mining," *IEEE transactions on Visualization and Computer Graphics*, vol. 8, no. 1, pp. 1–8, 2002.
- [4] C. Ware, *Information visualization: perception for design*. Morgan Kaufmann, 2012.
- [5] M. Ward, G. Grinstein, and D. Keim, *Interactive data visualization: foundations, techniques, and applications*. AK Peters, Ltd., 2010.
- [6] M. Mouine and G. Lapalme, "Using clustering to personalize visualization," in *Information Visualisation (IV), 2012 16th International Conference on*. IEEE, 2012, pp. 258–263.
- [7] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 9, pp. 1520–1536, 2012.
- [8] M. Agrawala, W. Li, and F. Berthouzoz, "Design principles for visual communication," *Communications of the ACM*, vol. 54, no. 4, pp. 60–69, 2011.
- [9] C. Chen, "Top 10 unsolved information visualization problems," *Computer Graphics and Applications, IEEE*, vol. 25, no. 4, pp. 12–16, 2005.
- [10] N. Elmqvist and J. S. Yi, "Patterns for visualization evaluation," 2012.
- [11] C. Plaisant, "The challenge of information visualization evaluation," in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2004, pp. 109–116.
- [12] J. Goldberg and J. Helfman, "Eye tracking for visualization evaluation: Reading values on linear versus radial graphs," *Information Visualization*, vol. 10, no. 3, pp. 182–195, 2011.
- [13] C. Weaver, "Look before you link: Eye tracking in multiple coordinated view visualization," *BELIV 10: BEyond time and errors: novel evaluation methods for Information Visualization*, p. 2, 2010.
- [14] A. McNamara, K. Mania, and D. Gutierrez, "Perception in graphics, visualization, virtual environments and animation," in *SIGGRAPH Asia 2011 Courses*. ACM, 2011, p. 17.
- [15] J. S. Yi, Y. ah Kang, J. T. Stasko, and J. A. Jacko, "Toward a deeper understanding of the role of interaction in information visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, no. 6, pp. 1224–1231, 2007.
- [16] M. Tory and T. Moller, "Evaluating visualizations: do expert reviews work?" *Computer Graphics and Applications, IEEE*, vol. 25, no. 5, pp. 8–11, 2005.
- [17] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock, and S. Carpendale, "Heuristics for information visualization evaluation," in *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. ACM, 2006, pp. 1–6.
- [18] K. Baker, S. Greenberg, and C. Gutwin, "Heuristic evaluation of groupware based on the mechanics of collaboration," *Engineering for human-computer interaction*, pp. 123–139, 2001.
- [19] C. Gutwin and S. Greenberg, "The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces," in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2000.(WET ICE 2000). Proceedings. IEEE 9th International Workshops on*. IEEE, 2000, pp. 98–103.
- [20] S. Greenberg, G. Fitzpatrick, C. Gutwin, and S. Kaplan, "Adapting the locales framework for heuristic evaluation of groupware," *Australasian Journal of Information Systems*, vol. 7, no. 2, 2007.
- [21] J. Drury and M. G. Williams, "A framework for role-based specification and evaluation of awareness support in synchronous collaborative applications," in *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2002. WET ICE 2002. Proceedings. Eleventh IEEE International Workshops on*. IEEE, 2002, pp. 12–17.