# VizRec: A Two-Stage Recommender System for Personalized Visualizations

**Belgin Mutlu, Eduardo Veas, Christoph Trattner, Vedran Sabol**
Know Center
Inffeldgasse 13, 8010 Graz, Austria
{bmutlu, eveas, ctrattner, vsabol}@know-center.at

## ABSTRACT
Identifying and using the information from distributed and heterogeneous information sources is a challenging task in many application fields. Even with services that offer well-defined structured content, such as digital libraries, it becomes increasingly difficult for a user to find the desired information. To cope with an overloaded information space, we propose a novel approach – VizRec– combining recommender systems (RS) and visualizations. VizRec suggests personalized visual representations for recommended data. One important aspect of our contribution and a prerequisite for VizRec are user preferences that build a personalization model. We present a crowd based evaluation and show how such a model of preferences can be elicited.

### Author Keywords
visualization recommendation; recommender systems; collaborative filtering; crowd-based experiment

### ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces—*Interaction styles , Graphical user interfaces (GUI)*; H.3 Information Storage and Retrieval: Digital Libraries—*Collection , Dissemination, Standards*

## INTRODUCTION
Finding the intended or right piece of information in a huge and continuously growing information space is not only a tedious, time consuming task, but a pressing challenge. Recommender systems (RS) try to tackle this problem by suggesting relevant information to users in a personalized or non-personalized manner. The suggestion takes the form of a list of items. Yet, a recommendation list can become incomprehensible and tedious when dealing with multi-dimensional data, where a user needs to compare and relate information in different data dimensions, a task that requires the user to process the list of recommendations sequentially. Visualizations deal with the overload problem by encoding information visually; leveraging the parallel capabilities of the human visual system to facilitate exploration and discovery in large datasets. To present a comprehensible picture of an overloaded information space we present in this paper a novel recommender system – called VizRec – that takes advantage of both worlds: (i) recommending information in a personalized manner to users and (ii) encode the information space with visualizations that help explore large amounts of multidimensional data efficiently.

Personalization plays an important role in presenting people relevant choices, for example for items to buy in a virtual marketplace (see e.g., Amazon.com) or to search for information on the Web (e.g., Google's Web search). Personalization is possible after a model of the users preferences has been built. We build a visualization recommender system (RS), that for arbitrary data sets and a given user's profile recommends interactive visual representations of the data appropriate to the task and tailored to the habits of the user. Visualizations leverage perceptual abilities inherent to the human visual system to encode information visually [2]. Not surprisingly, previous work attempted to capture factual knowledge of the visualization domain to recommend visualizations. Voigt et al. built a knowledge base to recommend visualizations for semantic web data [5]. A lean approach presented by Mutlu et al. concentrated on pragmatic, simple visual encoding facts and a complementing mapping algorithm to match data attributes with visual channels [3]. Despite these efforts, creating a visualization for arbitrary data is still challenging, and often referred as a task for experts. It requires an understanding of the users needs and preferences as well as visual encoding and perception guidelines. Considering just visual encoding of data leads to all possible valid visualizations, and disregards that a user often concentrates on particular aspects of the dataset. Henceforth, the visualization recommender needs: 1) to comply with visual encoding guidelines when proposing valid visualizations and 2) to only propose visualizations that make sense for the user. The former insures that valid configurations will be generated, the latter that the generated visualizations correspond to the needs and preferences of the user. Our recommender system (see Figure 2) meets those requirements by (1) automatically identifying a set of appropriate visualizations by analyzing the compatibility between both them and input data using on a rule-based mapping algorithm, (2) filtering a subset of those selected vi-


*IUI'15 Companion*, March 29–April 1, 2015, Atlanta, GA, USA.
Copyright © 2015 ACM 978-1-4503-3308-5/15/03.
http://dx.doi.org/10.1145/2732158.2732190

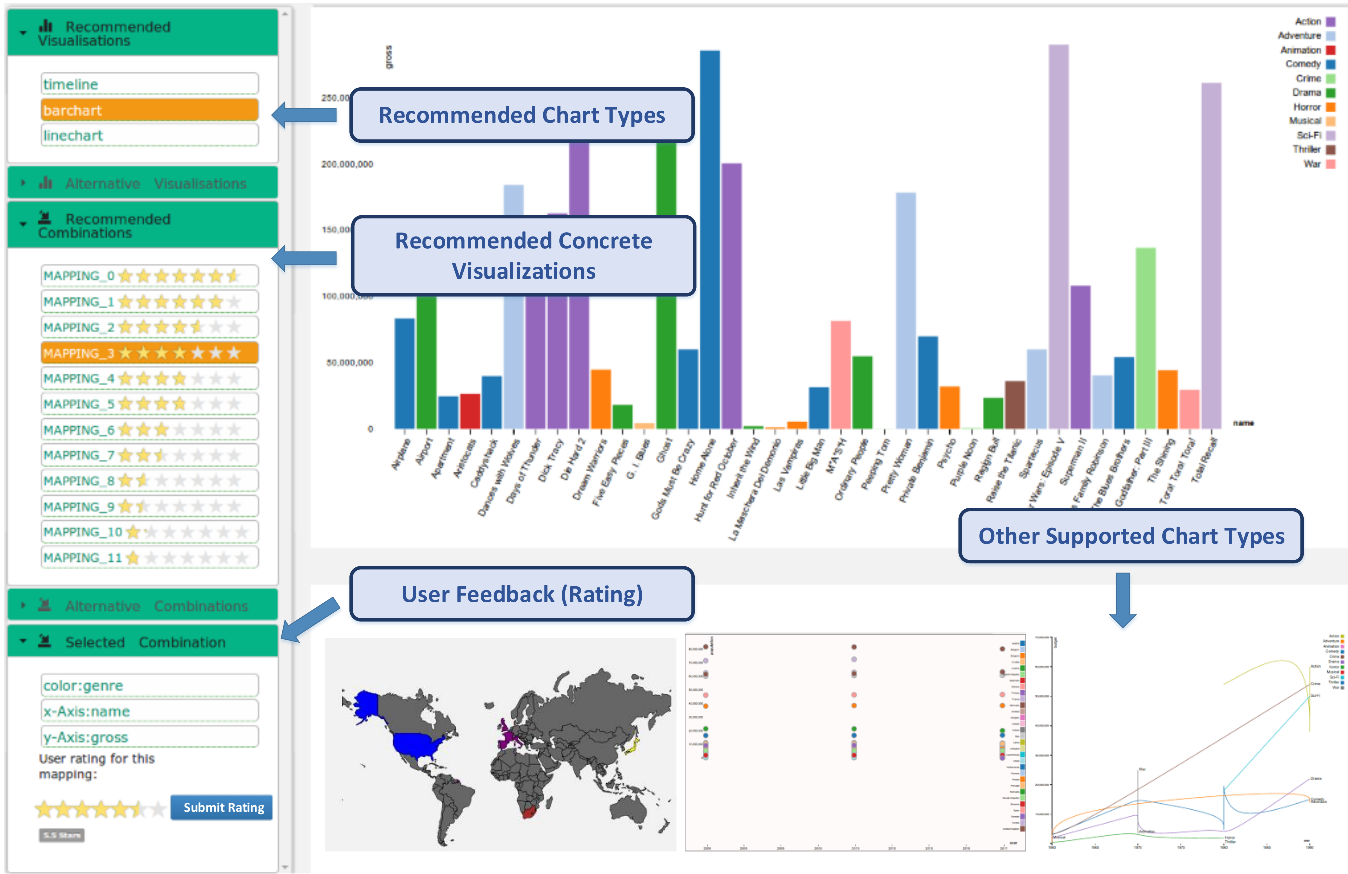


**Figure 1. Example of a visualization recommendation for a particular user with known preferences.**

sualizations based on user's preferences and (3) representing those visualizations as personalized recommendations.

The research question and focus of this paper is how to elicit user preferences pertaining the visualizations that better support the needs of each user. To obtain the user preferences, we executed a crowd-sourcing experiment using Amazon Mechanical Turk[1]. Using our mapping algorithm, we generated a set of all possible visualization configurations and mapping combinations, for an example data set from the cultural and educational domain. Subsequently, we ask participants to grade their quality. The results from our evaluation, i.e., the ratings obtained by the users helped to discriminate useful visualizations from a number of junk charts. Furthermore, they define a user model, which serves as an crucial input for our visualization recommender and enables us to measure its performance.

## THE APPROACH

Our approach intends to build personalized visual representations out of the recommendation list of a RS for cultural and scientific content. From a user query comprising the current visited page and extracted query terms, a Federated Recommender (FR) compiles recommendations from a number of associated service providers (e.g., Mendeley, Europeana and ZBW) and reverts with a list of relevant items. The list contains relevant items described by attributes common to all involved repositories. To cope with an overgrowing recommendation space of multidimensional data we propose a two-stage visualization RS, with a rule based stage to comply with visual encoding principles and a collaborative filtering stage for personalization of the results.

### Preprocessing

Before any recommendation is generated, a data preprocessor analyzes syntax and semantics of the data and generates a data description. In particular, the preprocessor performs (1) datatype identification tasks to determine the syntax and (2) data mining tasks (e.g., latent semantic analysis) to determine the semantics of the data. Syntax analysis serves to categorize data into standard data types such as categorical, temporal and numerical, – transformed into primitive data types string, date and number concretely – whereby the semantic analysis serves to define specific data types, such as geographical coordinates from a list of numerical fields. Syntax and semantic information of the data is used to identify the compatibility between visualizations and data in order to choose the appropriate visualizations.

[1] Amazon Mechanical Turk: http://aws.amazon.com/de/documentation/mturk/

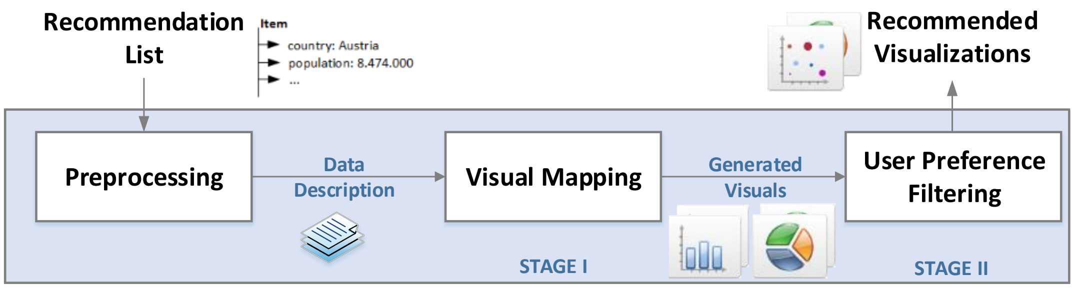


**Figure 2. VizRec workflow for recommending visualizations.**

### Stage One: Rule Based Recommendations

The mapping algorithm uses an ontology of chart patterns, each describing one possible combination of visual components of a chart and data types supported. For instance, the two possible patterns for the bar chart are, (1) $\{x-axis : string,\ y-axis : number\}$, and (2) $\{x-axis : date,\ y-axis : number\}$. Note, the patterns result from the fact that, depending on the properties of a chart, a visual component can support different data types in different combinations, such as *string* with *number* and *date* with *number* for the bar chart. The recommender counts four standard charts, and can be extended with additional charts by adding the description of instantiation patterns.

Following the approach of Mutlu et al. [3], a mapping operator, maps data points to visual channels (e.g., axes) of a visualization based on the following principles: (i) one attribute will be instantiated to one visual channel of a visualization, (ii) the datatype of the attributes should be compatible with those of the channels, and (iii) every mandatory visual channel of a visualization should be instantiated.

### Stage Two: Personalized Recommendations

To personalize the visualization suggestions we benefit from from user-based Collaborative Filtering (CF). The algorithm to do so requires ratings from users for the visualizations. Hence, we have designed a multidimensional rating based on scales pertaining the perceived usefulness and also visual organization of a visualization. The collected user feedback, will be applied by the CF to provide reasonable predictions of the active user's preferences, to predict which mapping combination(s) the active user might prefer, and to recommend her the top-N suitable combinations. Hence, the personalized list of recommendations is nothing else then a subset of all possible mapping combinations for the incoming data. Figure 1 shows a screenshot of VizRec with alternative visualizations highlighted and ratings for the current mapping combination.

The usefulness and rating of a visualization depends on the task and the information needs of the user. The challenge is to elicit preferences on these terms: which visualization is useful for particular information needs. To do so, we investigated in a crowd-sourced study, how people rate the out generated visual suggestions.

## EXPERIMENTAL SETUP

We conducted a crowd-based experiment using Amazon Mechanical Turk to obtain the user preferences. In particular, the experiment setup was the following: we used MovieLens[2] dataset as input for our visual recommender system and obtained 55 possible mapping combinations – concretely 32 combinations for bar chart, 9 for line chart, 13 for timeline and 1 combination for geo chart. We separated the suggested combinations in 18 HITs (Human Intelligent Task), each of which contained 3 randomly chosen mapping combinations and one additional HIT with only one combination. Once a user of Amazon Mechanical Turk (also called worker) accepted the HIT, she had to rate the quality of the generated mapping combination according to the following nine usability factors which we chose from a list of factors presented in [4] and [6]: (1) cluttered, (2) organized, (3) confusing, (4) easy to understand, (5) boring, (6) exciting, (7) useful, (8) effective, and (9) satisfying. The rating scale was between 1 and 7, where 1 meant not applicable and 7 very applicable.

## RESULTS AND DISCUSSION

Since, we wanted to obtain as much user feedback as possible per visualization by different types of people we asked overall 10 different workers per HIT, so that 36 workers in summary have participated in the experiment. Note that a worker could execute multiple HITs. The experiment started on 26th of November and ended on 3th of December 2014. The allotted working time per HIT was 900 sec. and the average working time of workers was 570 sec. per HIT. Figure 3 shows the results of the experiment.

The summary shows that some charts and their specific configurations were difficult to understand for workers. This is particularly the case with complex charts, such as timeline, line and geo chart. They mainly provide meaningful information for only specific kind of data and for few specific configurations of their visual channels (e.g., axes). On the other side, participants felt confident in identifying the information displayed by the bar chart. We believe that this is due to its simple data model, which is not that much flexible as in other charts in terms of possible mappings. Another reason may be that the displayed information is fixed to bars (e.g., in contrast to the line chart where different kinds of interpolation techniques between data points can be used).

[2] Movielens: https://movielens.org/

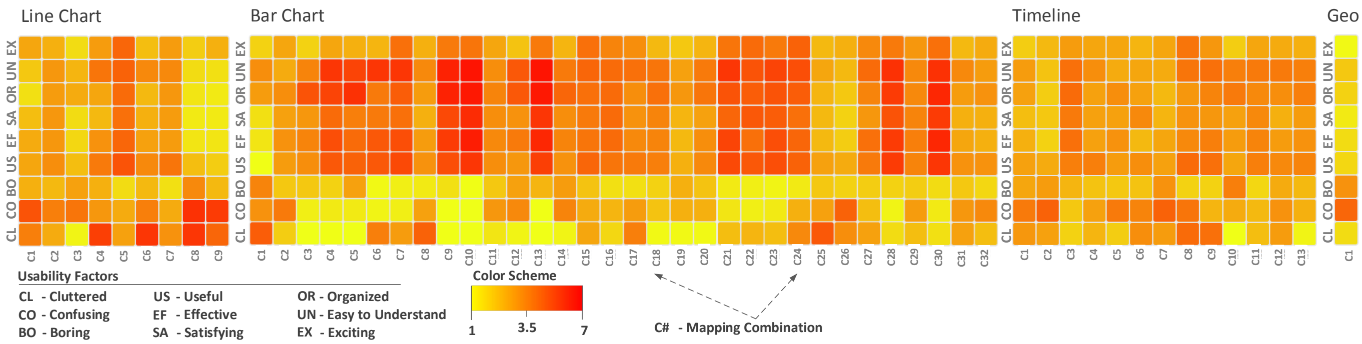


**Figure 3. Evaluation results: average of explicit ratings given to a particular mapping combination considering nine different usability factors, where 1 means not applicable and 7 very applicable.**

The main outcome of this experiment is the information about user preferences. The scores obtained can already be used as startup training for a CF-RS. Although just few charts have been considered here, we could identify that charts having rather complex data models and higher flexibility in terms of their configurations are likely to provide junk visualizations that user cannot understand and interpret. For such cases, and especially when mixing different visualizations for a single user query, the use of recommender systems becomes inevitable. In our ongoing work, we will extend the experiment to different datasets, additional visualizations, and in the end use the gathered information about user preferences to measure the improvements of visualization suggestions when applying RS.

## RELATED WORK

Recommending visualizations using collaborative filtering approaches is a relatively new research trend and up to now only a few related studies exist in that context. The most relevant one is a work conducted by Voigt et al. [5] who presents a knowledge-assisted, context-aware system to recommend visualizations for semantic Web data. Voigt et al. utilize an RDF-S/OWL vocabulary to annotate the data sources and visualization components, which contains factual knowledge of the visualization domain. However, our annotation of the visualization components strictly focuses on describing the visual encoding process, hence we represent visualizations in terms of their visual components. Instead of pursuing a through specification encompassing all known expert knowledge about visual perception, we concentrate on pragmatic, simple facts that will aid the sensible mapping (e.g., [1]), extending the description to four different types of visualizations. We differ from the presented work also regarding to the used RS approach, since we use a memory based user-based collaborative filtering algorithm instead of a model-based approach item-based approach.

## CONCLUSIONS

In this paper we presented a Visual Recommender, called VizRec, which (i) uses a ruled-based mapping algorithm in order to detect appropriate visualizations based on the syntax and semantics of the input data and (ii) to recommend only those visualizations which might make sense for the user regarding the current context and her preferences. To generate recommendations we benefit from user-based Collaborative filtering approaches, which uses a collection of user preferences to make predictions or generate recommendations. Hence, to obtain the user preferences, which are nothing else than ratings user gave for visualizations, we executed a crowd-sourcing experiment using Amazon Mechanical Turk. We presented the results of the experiment in a heat map and will reuse them in our future work to measure the performance of our recommender.

## ACKNOWLEDGMENTS

This work is funded by the EC 7th Framework project EEXCESS (grant 600601). The Know-Center GmbH is funded within the Austrian COMET Program – managed by the Austrian Research Promotion Agency (FFG). The study was carried out during the tenure of an ERCIM “Alain Bensoussan” fellowship program by the third author.

## REFERENCES

1. Mackinlay, J. Automating the design of graphical presentations of relational information. *ACM Trans. Graph. 5*, 2 (Apr. 1986), 110–141.
2. Munzner, T. Applying information visualization principles to biological network displays. vol. 7865 (2011), 78650D–78650D–13.
3. Mutlu, B., Hoefler, P., Tschinkel, G., Veas, E., Sabol, V., Stegmaier, F., and Granitzer, M. Suggesting visualizations for published data. In *Proc. of VISIGRAPP’14* (2014).
4. Seffah, A., Donyaee, M., Kline, R., and Padda, H. Usability measurement and metrics: A consolidated model. *Software Quality Journal 14*, 2 (2006), 159–178.
5. Voigt, M., Franke, M., and Klaus, M. Using expert and empirical knowledge for context-aware recommendation of visualization components. In *eKnow*, IARIA (2012).
6. Zheng, X. S., Lin, J. j. W., Zapf, S., and Knapheide, C. Visualizing user experience through ”perceptual maps”: Concurrent assessment of perceived usability and subjective appearance in car infotainment systems. In *Proc. of ICDHM’07*, Springer-Verlag (Berlin, Heidelberg, 2007), 536–545.