

## Faceted search of heterogeneous geographic information for dynamic map projection

Noemi Mauro<sup>a</sup>, Ardissono Liliana<sup>a,\*</sup>, Maurizio Lucenteforte<sup>a</sup>

<sup>a</sup>*Dipartimento di Informatica, Università degli Studi di Torino, Corso Svizzera 185, I-10149 Torino, Italy*

---

### Abstract

This paper proposes a faceted information exploration model that supports coarse-grained and fine-grained focusing of geographic maps by offering a graphical representation of data attributes within interactive widgets. The proposed approach enables (i) a multi-category projection of long-lasting geographic maps, based on the proposal of efficient facets for data exploration in sparse and noisy datasets, and (ii) an interactive representation of the search context based on widgets that support data visualization, faceted exploration, category-based information hiding and transparency of results at the same time. The integration of our model with a semantic representation of geographical knowledge supports the exploration of information retrieved from heterogeneous data sources, such as Public Open Data and OpenStreetMap.

We evaluated our model with users in the OnToMap collaborative Web GIS. The experimental results show that, when working on geographic maps populated with multiple data categories, it outperforms simple category-based map projection and traditional faceted search tools, such as checkboxes, in both user performance and experience.

*Keywords:* faceted information exploration, dynamic projection of geographic maps, GIS, geographic information search

---

\*This is to indicate the corresponding author.

*Email addresses:* [noemi.mauro@unito.it](mailto:noemi.mauro@unito.it) (Noemi Mauro), [liliana.ardissono@unito.it](mailto:liliana.ardissono@unito.it) (Ardissono Liliana), [maurizio.lucenteforte@unito.it](mailto:maurizio.lucenteforte@unito.it) (Maurizio Lucenteforte)

Declarations of interest: none.

## 1. Introduction

Several works promote the development of faceted search interfaces (Hearst, 2006) to reduce information overload and keep users in control of the search process in exploratory search (Marchionini, 2006). However, most of the existing approaches support an individual user who inspects a single data category, e.g., documents or movies, while pursuing a short-term information goal. Similarly, most research on geographic information search focuses on helping individual users retrieve relevant data for particular short-term goals; e.g., finding the available routes between two Points of Interest (Quercia et al., 2014), identifying the 2-star hotels in a specific area (Lionakis & Tzitzikas, 2017), or studying the relations among the items of an information category (Andrienko et al., 2007).

Indeed, map management can go farther than that in order to provide long-term representations of shared projects to users having different information interests. For instance, in participatory decision-making (Coulton et al., 2011; Brown & Weber, 2012), Hu et al. (2015) point out that 2D maps and 3D virtual environments can facilitate participants' learning and understanding, especially as far as spatial decision-making processes are concerned; see also (Al-Kodmany, 1999) and (Simpson, 2001). Moreover, maps can support information sharing and collaboration in simpler and less formal scenarios. For example, if somebody is planning a holiday, a custom map including selected Points of Interest, hotels, and so forth, would provide a personalized projection of the area to be visited that the user can consult and annotate before, during and after the trip, possibly in collaboration with the other people travelling with her/him to gain a common view of the vacation.

These scenarios suggest the development of custom maps that define Personal Information Spaces (Ardito et al., 2013, 2016) useful to organize individual and group activities. For this purpose, maps should be adapted to reflect temporary information goals while persistently storing data in order to facilitate a

30 quick projection and resumption of the collaboration context.

In this paper we present a faceted exploration model for the management of this type of map. Our model supports a flexible, map-based visualization of heterogeneous data and it enables map focusing to satisfy specific information needs by offering two graphical interactive exploration functions:

- 35 • The former enables coarse-grained map projection on data categories via opacity tuning, without taking the facets of items into account; all the items of a category are subject to the same visualization policy.
- The latter combines opacity tuning with fine-grained faceted search support to enable map projection at different granularity levels, by taking the  
40 properties of information items into account.

In both cases, the projection is only visual and the information stored in the map is preserved. The work we present has the following innovative aspects:

- 45 • *Efficient multi-category faceted projection of long-lasting custom maps to answer temporary information needs in sparse and noisy datasets.* Our model suggests information visualization constraints based on attributes of data which support an efficient exploration of the information items stored in the maps.
- 50 • *Representation of the search context by associating each data category to a compact graphical widget that supports interactive data visualization, faceted exploration, category-based information hiding and transparency of results.* The widgets of the categories searched by the user are located in a side bar of the user interface and play the role of breadcrumbs, representing the types of information that (s)he has explored during the interaction with the system and the applied visualization constraints.

55 Our model supports geographic information search within the OnToMap collaborative Web GIS (Ardissono et al., 2017, 2018). We tested the model in a user study to assess User Experience and performance in exploratory search. For the

experiments we compared different graphical widgets supporting faceted exploration, from traditional ones such as checkboxes, to space-filling ones based on treemaps (Shneiderman, 1992) and sunburst (Stasko et al., 2000) diagrams. The study showed that, when working on geographic maps populated with heterogeneous information, our model outperforms simple category-based map projection and traditional faceted search tools such as checkboxes. Specifically, we obtained the best user performance and experience results using the widget based on the sunburst diagram, which displays visualization criteria in a compact structure.

This article builds on the work described in (Ardissono et al., 2018), which presents our first opacity tuning model. With respect to that work, the present paper introduces graphical widgets that support fine-grained data management and a novel approach to the selection of efficient facets for information exploration in sparse and noisy datasets. The widgets extend the category hiding function provided by the previous model with faceted data exploration to enhance information search and visualization. The present paper also provides an extensive evaluation of the faceted exploration model.

In the following, Section 2 presents our research questions. Section 3 outlines the related research. Section 4 overviews OnToMap and Section 5 describes our model. Section 6 presents the experiments we carried out and Section 8 discusses the evaluation results. Section 9 summarizes limitations and future work and the Appendix reports a few tables of details.

## 2. Research questions

We designed the faceted search model presented in this paper after two preliminary experiments with users carried out in the urban planning domain; see (Voghera et al., 2016, 2018). In those experiments, the projection of long-lasting maps on specific types of information emerged as a useful feature to support data interpretation during project development. This feature was also requested in the final analysis phase, in which human planners analyzed complex maps

obtained by integrating the students' projects to identify the most recurring represented territorial elements.

The present work describes the faceted search support offered by the current  
90 version of OnToMap and investigates its usefulness to data search and interpretation in a project map. We pose the following research questions:

RQ1: *Does faceted exploration of map content help users in finding the needed information in a geographic map that visualizes different types of data?*

RQ2: *How does a compact, graphical view of the exploration options available  
95 to the user, which also shows the status of the information visualization constraints applied to a map, impact on her/his efficiency and experience in data exploration?*

RQ3: *How much does the user's familiarity with the widgets for faceted exploration impact on her/his efficiency in search and on her/his appreciation  
100 of the exploration model they offer?*

The experiments described in Section 6 are aimed at answering these questions.

### 3. Background and related work

Exploratory search of large information spaces challenges users in the specification of efficient queries because, as most people are hardly familiar with  
105 the search domain, their information goals are often ill-defined (Marchionini, 2006; White & Roth, 2006). In this paper we focus on faceted search as an alternative, or a complement, to query typing in order to use browsing-based navigation as a proactive guide to information exploration, given the structure of the information space.

110 Starting from the pioneer filtering model proposed by Ahlberg & Shneiderman (1994), both Sacco (2000)'s Dynamic Taxonomies and Hearst (2006)'s faceted search model propose to use dynamic filters extracted from items metadata as constraints that the system can suggest to help the user identify relevant terms for information filtering and visualization of results. Specifically, Hearst

Table 1: Classification of information search models.

| Citations                     | Model                        | Facets               | Filters  | Visualization of results                             |
|-------------------------------|------------------------------|----------------------|--|--|
| (Ahlberg & Shneiderman, 1994) | FilmFinder                   | static list of terms | buttons<br>sliders, ...                            | scatterplot  |
| (Cao et al., 2010)            | FacetAtlas                   | terms                | -  | 3D diagram with<br>multidimensional<br>relations     |
| (Cao et al., 2011)            | SolarMap                     | terms                | topic facets +<br>keyword facets                   | document clusters +<br>radial facet<br>visualization |
| (Chang et al., 2019)          | SearchLens                   | terms                | lenses   | ranked list +<br>match to lenses                     |
| (Dachselt & Frisch, 2007)     | Mambo                        | from<br>metadata     | attributes   | stack-based<br>hierarchical facets +<br>zoom on data |
| (Dachselt et al., 2008)       | FacetZoom                    | from<br>metadata     | attributes   | stack-based<br>hierarchical facets +<br>zoom on data |
| (Hearst et al., 2002)         | Flamenco                     | from metadata        | text queries<br>hyperlinks                         | matrix view + miniatures                             |
| (Hildebrand et al., 2006)     | /facet                       | from RDF<br>metadata | attributes +<br>hyperlinks                         | hierarchical +<br>facets + list                      |
| (Hoeber & Yang, 2006)         | HotMap                       | terms                | text queries                                       | list + term distribution<br>with colors              |
| (Hoeber & Yang, 2006)         | Concept<br>Highlighter       | concepts             | text queries                                       | list + concept distribution<br>with colors           |
| (Lee et al., 2009)            | FacetLens                    | attributes           | text query +<br>attributes                         | matrix-based<br>bubbles by filter                    |
| (Lionakis & Tzitzikas, 2017)  | PFSgeo                       | from RDF<br>metadata | attributes +<br>hyperlinks +<br>geo. dimension     | ranked list  |
| (Papadakos & Tzitzikas, 2014) | Hippalus                     | from RDF<br>metadata | attributes +<br>hyperlinks                         | ranked list  |
| (Peltonen et al., 2017)       | Topic-Relevance<br>Map       | terms                | text query   | radial distance +<br>relative distance               |
| (Petrelli et al., 2009)       | -                            | from RDF<br>metadata | attributes +<br>hyperlinks                         | multiple<br>visualizations                           |
| (Stadler et al., 2014)        | Facete                       | from RDF<br>metadata | attributes +<br>sem. relations +<br>geo. dimension | ranked list +<br>map-based<br>visualization          |
| (Sutcliffe et al., 2000)      | Thesaurus-<br>Result Browser | terms                | Hierarchical<br>Thesaurus                          | bullseye<br>browser                                  |

115 et al. (2002) present the Flamenco framework in which facet-based filtering is  
based on the exposure of hierarchical faceted metadata that describes the items  
of the search domain, i.e., apartments, or images (Yee et al., 2003).

120 Researchers also investigate ways to support the specification of the facets to  
filter results, as well as the access to Semantic Web information and Linked Data  
(W3C, 2018). As far as facet specification is concerned, new types of elements  
are proposed to filter the set of results; e.g., keywords or terms extracted from  
textual queries, as in HotMap (Hoeber & Yang, 2006), concepts extracted from  
a document pool, as in Concept Highlighter (Hoeber & Yang, 2006), or terms

extracted from a thesaurus as in Thesaurus-Results Browser (Sutcliffe et al.,  
125 2000). FacetLens (Lee et al., 2009) visualizes clickable facets in matrix-based  
bubbles, each one associated with a different search filter. Moreover, FacetZoom  
(Dachselt et al., 2008) proposes a stack-based visualization of hierarchical facets,  
also applied in Mambo (Dachselt & Frisch, 2007) as a model to combine faceted  
browsing with zoomable user interfaces. SearchLens (Chang et al., 2019) enables  
130 users to define long-lasting composite facet specifications (denoted as lenses) to  
support information filtering on multiple search sessions. In SearchLens, the  
user can specify the importance of the selected facets; thus, filtering is based on  
soft constraints used to rank search results.

In the faceted exploration of semantic data (Tzitzikas et al., 2017), search  
135 interfaces expose rich metadata that support browsing the information space  
through semantic relations. For instance, in the /facet browser, Hildebrand  
et al. (2006) propose to combine hierarchical faceted exploration with keyword-  
based search. Moreover, Petrelli et al. (2009) enables the user to search for  
heterogeneous types of information about items (e.g., texts and images) linked  
140 according to semantic relations, by extracting facets to guide exploration. Hip-  
palus (Papadakos & Tzitzikas, 2014) introduces the Faceted and Dynamic Tax-  
onomies to manage both hard and soft constraints in faceted filtering of semantic  
data and PFSgeo (Lionakis & Tzitzikas, 2017) extends Hippalus to geographic  
information management. Finally, focusing on geographic information, Stadler  
145 et al. (2014) propose a semantic navigation method for SPARQL-accessible data  
(W3C, 2017a; OCG, 2017) in the Facete browser.

Similar to the cited works, our model exposes metadata derived from se-  
mantic knowledge representation. However, it enables users to work on maps  
populated with multiple data categories, i.e., with heterogeneous information,  
150 as well as to focus the maps on temporary interests without losing the overall set  
of data they contain. This is useful to answer information needs in long-lasting  
user activities. Notice also that OnToMap does not assume to work on RDF data  
in order to comply with more general data sources, like public crowdmapping  
platforms, thanks to the mediation of its domain ontology. Moreover, it sup-

ports: (i) a browsing-based exploration guided by the structure of the domain ontology, which makes it possible to search for information following both IS-A and semantic relations; (ii) the semantic interpretation of free text queries to identify the data categories (ontology concepts) of interest by abstracting from the specific words occurring in the queries, via Natural Language Processing (Ardissono et al., 2016; Mauro et al., 2017). More generally, OnToMap enables search support over a configurable set of data categories; in this way, it enables complex map development on different information domains. In contrast, most of the previous systems work on a single data type or on a pre-defined set of data categories, as in (Petrelli et al., 2009).

The dynamic extraction of facets can challenge the user with a large number of browsing options. Oren et al. (2006) focus on the efficiency of exploration and they promote the facets that enable the user to split the set of results in balanced subsets in order to minimize navigation steps. In comparison, we propose a facet selection policy suitable for sparse and highly unbalanced result sets, such as those typically retrieved from crowdsourced data sources, in which very few properties of items split results in subsets having similar cardinality.

Some works propose interactive graphical presentations of keywords to support sensemaking in the exploration of document sets. For instance, Peltonen et al. (2017) propose the Topic-Relevance Map to summarize on a radial basis the keywords (filters) characterizing the result set, using distance from the center to represent relevance to the search query and angle between keywords to denote their topical similarity. Moreover, FacetAtlas (Cao et al., 2010) relates topics in a 3D diagram supporting the representation of multi-dimensional relations among them, and SolarMap (Cao et al., 2011) combines topic-based document clustering with a radial representation of facets to support a two-level, topic-based document filtering. While these models are appropriate to the representation of topics in datasets of unstructured information, they are less relevant to OnToMap, which is fed with structured data and benefits from its domain ontology to organize the presentation of information.



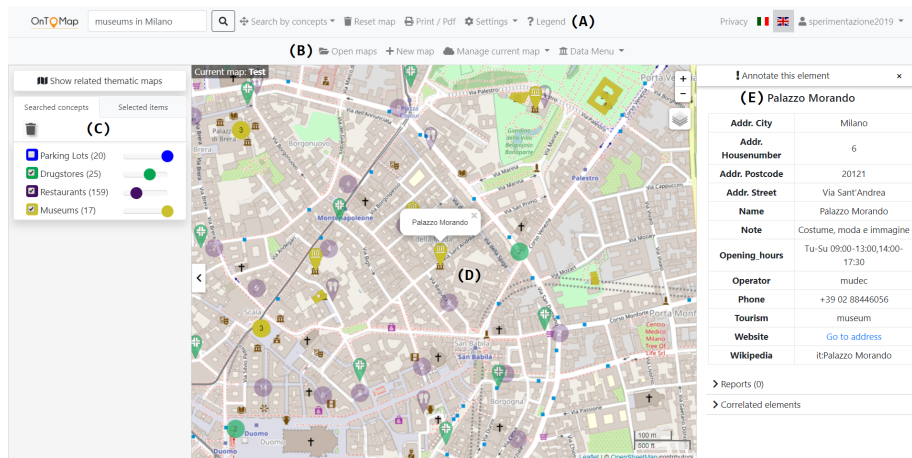


Figure 1: OnToMap user interface showing the widgets based on transparency sliders.

#### 185 4. OnToMap overview

The OnToMap Web collaborative GIS supports the management of interactive maps for information sharing and participatory decision-making (Ardissono et al., 2017). A semantic representation of domain knowledge based on an OWL ontology (W3C, 2017b) defines the structure of the information space and enables data retrieval from heterogeneous sources by applying ontology mappings; see (Mauro et al., 2019). This ontology currently makes it possible to query a dataset of Public Open Data about Piedmont area in Italy and the OpenStreetMap (OSM) server (OpenStreetMap Contributors, 2017); in this paper, we focus on OSM data because it is a more general case than the other one. The ontology also provides graphical details for map visualization, such as the color and icon associated to each data category; e.g., drugstores are depicted in light green and they are represented as icons marked by a cross.

OnToMap supports the creation of public and private custom geographic maps to help project design and group collaboration. Search support is based on free text queries and on category browsing. Textual queries are semantically interpreted using Natural Language Processing techniques with Word Sense Disambiguation (Moro et al., 2014); see (Ardissono et al., 2016; Mauro et al., 2017).

Data categories can be browsed by means of a simple alphabetical menu with auto-completion or by navigating a graphical representation of the taxonomy defined in the domain ontology.

Figure 1 shows the user interface of OnToMap. The top bar includes the control panel that supports (A) free text search (“Search...”) and category browsing (“Browse by concepts”), basic map management and user authentication; (B) map management tools available to authenticated users. The left side bar (C) displays the widgets of the data categories that the user has searched for during the interaction: a different widget is associated to each category in order to regulate the visualization of its items in the map.

The main portion of the interface (D) contains the geographic map, which displays information items as pointers with category-specific icons or as geometries, depending on the input data. Color coding (Hoeber & Yang, 2006) visually connects the widgets representing data categories in the side bar with the corresponding items in the map. Moreover, when several items of a category are located in a restricted area, a cluster colored as the category is displayed to avoid cluttering and, at the same time, provide visual information about the grouped items. If the user clicks on the visual representation of an item, the system generates a table (E) describing its details.

## 5. Information exploration model

Our information exploration model is integrated in the OnToMap system to support information search and it includes two main types of functions, implemented as interactive graphical widgets.

### 5.1. *Exploration function 1: coarse-grained map projection by means of transparency sliders*

This function, introduced in (Ardissono et al., 2018), supports map projection via opacity tuning: for each searched category, a transparency slider enables the user to assign different levels of opacity to its items; the widget also has a

checkbox to temporarily hide information by means of a click, without changing the degree of opacity selected for the category.

The side bar of Figure 1 shows the widgets based on transparency sliders. In the map, museums are visualized in full color because the slider of the “Museums” category is selected and tuned to maximum opacity. Differently, drugstores and restaurants are semi-transparent and the map hides the items of the “Parking Lots” category because its slider is de-selected.

The transparency slider does not enable the specification of constraints on facet values; i.e., it works at the granularity level of the represented category and it uniformly tunes the opacity of items. Nevertheless, this widget supports visual simplification by enabling the user to temporarily hide information by type. Basically, opacity tuning enables her/him to highlight the information in focus while maintaining an overview of what has been searched in the map. This model is inspired by Colby & Sholl (1991)’s work on layers visualization but it separately handles the opacity of items belonging to different categories; moreover, it supports the visualization of multiple layers, as a generalization of Translucent Overlay (Lobo et al., 2015).

### *5.2. Exploration function 2: faceted approach*

This function combines coarse-grained and fine-grained specification of visualization constraints by integrating transparency sliders with faceted information exploration. The widgets implementing this function include a transparency slider and an internal component showing the facets of the represented category. The internal component can be a set of checkboxes, a treemap or a sunburst diagram, depending on the layout selected for the user interface, and it enables the user to specify visualization constraints based on facet values. The transparency slider works in combination with facet selection and tunes the opacity of the visualized items. The widgets are interactive and they can be opened or closed by clicking on them; a closed widget only shows its own transparency slider; e.g., see “Drugstores” in Figure 2, which shows the layout based on checkboxes.

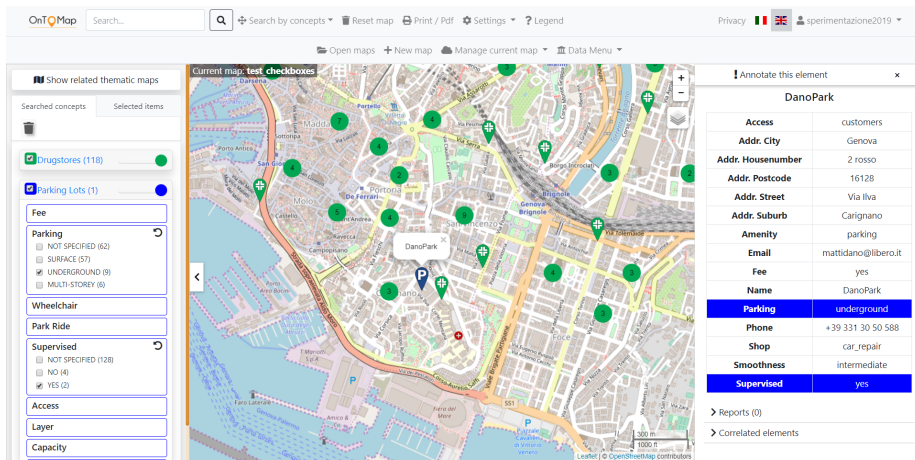


Figure 2: User interface showing the widgets based on checkboxes.

Let us consider a facet  $f$  of a category  $C$  and the set of retrieved items that belong to  $C$ , henceforth denoted as  $E_C$ , i.e., extension of  $C$ . The visualization of the values of  $f$  in the widget is aimed at providing the user with a preview of the corresponding items in the map. For this purpose we adopt a standard approach to facet suggestion (Oren et al., 2006; Hearst, 2006):

- The widget only displays the values  $\{v_1, \dots, v_m\}$  of  $f$  that have at least one item in  $E_C$  to prevent the user from following links to zero solutions.
- The values of  $f$  are sorted from the most frequent to the least frequent ones in  $E_C$ . Moreover the widget shows, or makes available on mouse over, the number of items corresponding to each value. Notice that the widget may also show a “NOT SPECIFIED” value to represent the subset of items in which  $f$  is not defined. This is aimed at providing the user with a visual representation of the coverage of the facet in the results.
- In order to limit visual complexity, long lists of values are dropped, making their tails available on demand by providing a “More...” link or a “+” symbol, depending on the layout of the widget.

By default, none of the facets in the widget of a category  $C$  is selected. If

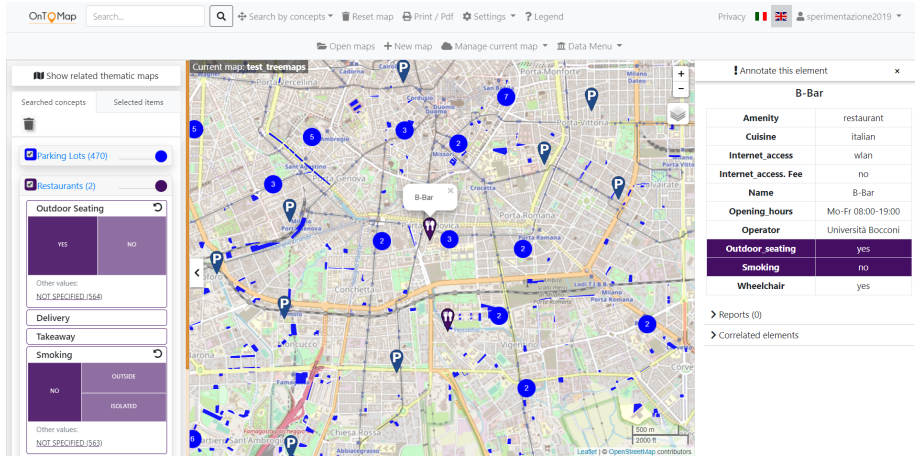


Figure 3: User interface showing the treemaps as faceted exploration widgets.

the user picks one or more values of the same facet, this is interpreted as an OR constraint because (s)he has specified that all those values are eligible for visualization. Conversely, the selection of values that belong to distinct facets of  $C$  generates an AND constraint because it identifies the items having more than one property restricted to specific values. For instance, if the user chooses  $f_i = v_{i1}$ ,  $f_i = v_{i2}$  and  $f_j = v_{j1}$ , items  $\{x \in E_C \mid f_i(x) \in \{v_{i1}, v_{i2}\} \wedge f_j(x) = v_{j1}\}$  are shown and the other items are hidden.

We use color coding to link visualization constraints to map content: the tables showing the details of items highlight the facets corresponding to the selected visualization constraints in the color associated to the category. In this way, the user can quickly identify the characteristics that make items eligible for being displayed. For instance, the table of “DanoPark” in Figure 2 has the “Parking” and “Supervised” facets highlighted in blue because they correspond to the visualization constraints imposed on the “Parking Lots” widget.

### 5.2.1. Layouts of the widgets for faceted information exploration

Before providing details about how we select the facets to be included in the widgets we present the layouts we developed.

- The **widgets based on checkboxes** contain a rimmed rectangle for each

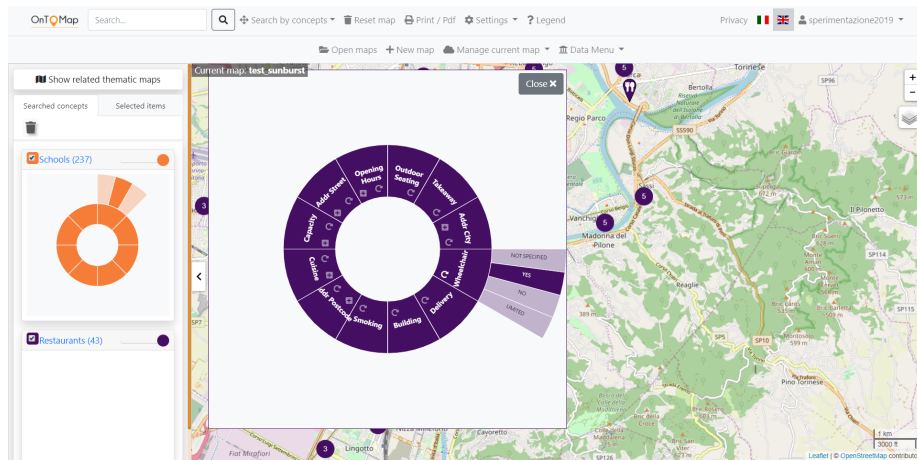


Figure 4: User interface showing the sunburst as faceted exploration widgets.

facet to be shown. By clicking on a rectangle, the user expands (or closes) the corresponding facet. An expanded facet shows values as checkboxes and offers a “More...” link to show the hidden ones. For instance, Figure 2 shows eight facets of widget “Parking Lots”: “Fee”, . . . , “Capacity”. Users can select the checkboxes to impose visualization constraints in the map. In the figure the user has expanded the “Parking” and “Supervised” facets and (s)he has selected values “UNDERGROUND” and “YES”.

- **The widgets based on treemaps** include facets in rimmed rectangles as well. However, when a facet is expanded, its values are displayed as components of a treemap whose size depends on the cardinality of the corresponding set of items (larger size means larger cardinality); see Figure 3. Long values are shortened but they can be visualized, together with the cardinality of the corresponding sets of items, on mouse over. Only the most frequent values are included in the treemap; the other ones are available below it or on demand (“More...”) and the user can add them to the treemap by means of a click. The user can (de)select values by clicking on them. The selected values take the color of the category (e.g., “YES” in “Outdoor Seating” and “NO” in “Smoking”); the other ones

have a pale tone of the same color.

- 315 • The **widgets based on the sunburst diagram** show the facets of the represented category  $C$  in a ring having the color associated to  $C$ . The diagram is visualized in a pop-up window that the user can open or close by means of a click, and the side bar of the user interface only displays the thumbnails of the sunbursts; see Figure 4. The user can expand each  
320 facet by clicking on the portion of ring representing it: values are shown in a second level, sorted clockwise by decreasing frequency in the extension of  $C$ . Only the most frequent values are shown but the user can view and add the other ones by clicking on the “+” button located in each portion of the internal ring; the sunburst is extended by adapting the size of the  
325 displayed values. The user can (de)select values by clicking on them and color coding is applied to link visualization constraints to map content.

### 5.3. Selection of facets to be included in the information exploration widgets

The dynamic generation of widgets for the exploration of search results retrieved from open data sources is challenged by the amount and variability of  
330 the information items to be managed. Facets have thus to be analyzed in order to identify the most convenient ones for map content analysis.

#### 5.3.1. Navigation quality in semantic data repositories

Oren et al. (2006) introduce the *navigation quality* of a facet  $f$  to describe its efficiency in supporting information browsing of RDF data repositories. This  
335 measure takes values in  $[0, 1]$  (where 1 is the best value) and is based on the product of three metrics, which take values in  $[0, 1]$  as well:

1. The *balance* of  $f$ , i.e., its capability to split results in subsets having similar cardinality; equally distributed facets have maximum balance.
2. An inverse measure of the *number of distinct values* of  $f$  occurring in the  
340 results, denoted as “*object cardinality*”. The authors consider as acceptable the facets that have between 2 and 20 values because they can be displayed in a search interface without overloading the user.

Table 2: Value distribution of facets “Cuisine”, “Outdoor Seating” and “Takeaway” (OSM tag: “amenity=restaurant”) in Torino city bounding box; retrieved using Overpass Turbo (OpenStreetMap Wiki contributors, 2019) on Sept. 20th, 2019. The results include 719 items, out of which 432 specify the value of “Cuisine”, 92 specify the value of “Outdoor Seating” and 76 specify the value of “Takeaway”.

| Cuisine                             | Count | Outdoor Seating | Count | Takeaway | Count |
|-------------------------------------|-------|-----------------|-------|----------|-------|
| PIZZA                               | 111   | NO              | 59    | YES      | 62    |
| ITALIAN                             | 74    | YES             | 33    | NO       | 10    |
| REGIONAL                            | 37    |                 |       | ONLY     | 4     |
| CHINESE                             | 28    |                 |       |          |       |
| JAPANESE                            | 17    |                 |       |          |       |
| ITALIAN; PIZZA                      | 15    |                 |       |          |       |
| SUSHI                               | 11    |                 |       |          |       |
| ITALIAN; REGIONAL                   | 10    |                 |       |          |       |
| PIZZA; ITALIAN                      | 10    |                 |       |          |       |
| MEXICAN                             | 9     |                 |       |          |       |
| KEBAB                               | 7     |                 |       |          |       |
| INDIAN                              | 4     |                 |       |          |       |
| ASIAN                               | 3     |                 |       |          |       |
| CHINESE; JAPANESE                   | 3     |                 |       |          |       |
| FISH                                | 3     |                 |       |          |       |
| INTERNATIONAL                       | 3     |                 |       |          |       |
| ITALIAN; PIZZA; REGIONAL            | 3     |                 |       |          |       |
| ITALIAN_PIZZA                       | 3     |                 |       |          |       |
| PERUVIAN                            | 3     |                 |       |          |       |
| STEAK_HOUSE                         | 3     |                 |       |          |       |
| AMERICAN                            | 2     |                 |       |          |       |
| CHINESE; PIZZA                      | 2     |                 |       |          |       |
| GREEK                               | 2     |                 |       |          |       |
| ITALIAN_PIZZA; PIZZA                | 2     |                 |       |          |       |
| KEBAB; PIZZA                        | 2     |                 |       |          |       |
| LOCAL                               | 2     |                 |       |          |       |
| MEDITERRANEAN                       | 2     |                 |       |          |       |
| PIZZA; KEBAB                        | 2     |                 |       |          |       |
| REGIONAL; ITALIAN                   | 2     |                 |       |          |       |
| AFRICAN                             | 1     |                 |       |          |       |
| ...                                 |       |                 |       |          |       |
| <i>56 more values with Count=1.</i> |       |                 |       |          |       |

3. The *frequency* of  $f$  in the results, i.e., the percentage of retrieved items in which the value of  $f$  is specified.

345 Navigation quality cannot be applied in OnToMap because of its assumptions: firstly, statistics about OSM data provided by TagInfo (OSM Contributors, 2019) show that most of the tags are hardly used.<sup>1</sup> This can be explained

<sup>1</sup>For instance, by invoking <https://taginfo.openstreetmap.org/tags/amenity=restaurant#combinations> it is possible to learn that “amenity=restaurant” has 178



because crowd-mappers tend to underspecify the items they map; moreover, they sometimes define new tags instead of using the existing ones, thus generating  
350 a plethora of synonyms which increase in an uncontrolled way the number of distinct facets and values. This phenomenon is so widespread that several efforts try to systematize OpenStreetMap through semantic knowledge representation; e.g., see (Codescu et al., 2011; Ballatore et al., 2013). We also notice that several results retrieved from OSM are unbalanced and can be split into (i) a large set  
355 of items in which the facet is not available, (ii) a few values identifying sets of items with reasonable cardinality, and (iii) a long tail of values represented by one or two items. For instance, Table 2 shows the distribution of three facets retrieved from OSM by searching for “amenity=restaurant” (which corresponds to the “Restaurants” category of the OnToMap domain ontology) on Torino city  
360 bounding box. The facets have fairly poor coverage and they are unbalanced: “Cuisine” is specified in 432 items out of 719 and it exhibits a distribution with a long tail; “Outdoor Seating” and “Takeaway” are specified in 92 and 76 items respectively; “Name”, not shown, is balanced but it only occurs in 675 items.

We thus define a novel approach to the computation of facet efficiency that  
365 suits these types of distribution and is robust towards information lack. The idea is that (i) coverage has to be taken into account as a separate factor to select useful facets, and (ii) balance and number of values have to be controlled by the cardinality of the sets of items identified by the facet.

### 5.3.2. *Our approach: evaluating exploration cost in sparse, unbalanced datasets*

370 When searching for information in crowdsourced data sources, the suggestion of the most representative facet values in a result set is a primary goal because it enables the system to provide the user with a relevant number of items to choose from. Moreover, it can be complemented by free text queries that let the user

---

different tags, only 36 of which occur in more than 2% of the items mapped in OSM worldwide. Moreover, the most frequent tag is “name”, which is only defined in 90.92% of items, in spite of its importance as a POI identifier.

express specific information needs; e.g., in OnToMap free text queries support  
 375 the retrieval of very specific items, such as “*Pediatric* hospitals in Torino”. It  
 thus makes sense to propose facets that, regardless of balance, identify some  
 fairly large subsets of items, possibly leaving the long tail apart or making it  
 available on demand; e.g., consider the first values of “Cuisine” in Table 2.  
 Given these premises, we propose a two-step evaluation of facets efficiency to  
 380 exploration support.

1) In the first step, we consider *frequency* as a pre-filtering metric to exclude  
 from any further computation the facets that appear very rarely in the results.  
 Having sampled a set of queries to OSM and taking Taginfo statistics as a  
 baseline, we empirically set to 3% the minimal frequency threshold under which  
 385 a facet is considered as useless. Only the facets over this threshold are considered  
 for the evaluation of their efficiency.

2) In the second step, given the highly variable distribution of facets, we  
 consider balance and number of values in combination. We are interested in  
 facets that split  $E_C$  in at least some portions having significant cardinality  
 because they identify homogeneous, relatively large sets of items to be analyzed.  
 These facets enable the system to propose visualization criteria that significantly  
 reduce the search space by showing the most representative values, leaving the  
 other ones on demand. Differently, in a small result set, as those typically  
 retrieved when the selected bounding box is strict, there are few items; therefore,  
 the efficiency in splitting results is less important because the user can easily  
 analyze items one by one. In order to capture this intuition, we compute the  
*cost of exploring the extension  $E_C$*  of a category  $C$  by means of a facet  $f$  that  
 takes values in  $\{v_1, \dots, v_m\}$  as follows:

$$explorationCost(f) = \frac{-\sum_{j=1}^m p(v_j) \log_2 p(v_j)}{meanCard(f)} \quad (1)$$

$explorationCost(f)$  takes values in  $\mathbb{R}^+$ . In the formula,  $p(v_j)$  is the probability  
 of  $v_j$  in  $E_C$ , computed by considering the values  $v_j \neq$  “NOT SPECIFIED”.  
 Moreover,  $meanCard(f)$  is the mean frequency of the values of  $f$  in  $E_C$ , i.e.,

the mean cardinality of the subsets of results identified by  $f$ :

$$\text{meanCard}(f) = \frac{|E_C|}{m} \quad (2)$$

The components of Equation 1 have the following roles:

- The numerator represents the (not normalized) entropy of  $f$ , which takes values in  $\mathbb{R}^+$ . The entropy of an information source is an average measure of the amount of uncertainty of its own  $m$  symbols; it is positively  
390 influenced by both the number of values that the source can take and by the balance of the corresponding subsets of items. For instance, given two balanced facets  $f_1$  and  $f_2$ , if  $f_2$  has more values than  $f_1$ ,  $f_2$  also has higher entropy than  $f_1$ . Moreover, if two facets have the same number of values,  
395 the most balanced one has the highest entropy. Finally, if all the items of  $E_C$  have the same value of  $f$  (e.g., all the schools located in the bounding box are primary ones), the entropy is 0, meaning that the facet does not help discriminate among the items of  $E_C$ .
- The denominator of Equation 1 captures our interest in the facets that  
400 split results in fairly large subsets: even though a facet  $f$  has high entropy (e.g., because it has several values), its cost is smoothed if the subsets of items it identifies have high mean cardinality, because  $f$  enables the user to browse a large portion of results in few steps.

Figure 5 graphically compares the exploration cost of Equation 1 with Oren  
405 et al. (2006)’s navigation quality on a few facets; see Tables 8 and 9 in the Appendix for details. We consider “Cuisine”, “Takeaway”, “Outdoor Seating” and “Name”, based on the data described in Table 2, and two toy examples:

- “Ex1”, specified in 160 items, has 8 distinct balanced values, with  $\text{meanCard} = 20$ .
- “Ex2”, specified in 24 items, has 8 distinct balanced values, with  $\text{meanCard} = 3$ .

|                 | Exploration cost | (1 - navigation quality) |
|-----------------|------------------|--------------------------|
| Outdoor Seating | 0.0205           | 0.8924                   |
| Takeaway        | 0.0335           | 0.9497                   |
| Ex1             | 0.1500           | 0.8949                   |
| Cuisine         | 0.8738           | 1.0000                   |
| Ex2             | 1.0000           | 0.9842                   |
| Name            | 9.3987           | 1.0000                   |

Figure 5: Exploration cost and complement of navigation quality of a set of facets. The color scale varies from the lowest cost values, depicted in green, to the highest ones, in red. Notice that colors are tuned to the values observed in this example; i.e.,  $[0, 10]$  for exploration cost and  $[0, 1]$  for the complement of navigation quality.

Notice that Oren and colleagues compute a quality measure, i.e., the highest values are the preferred ones; conversely, we compute a cost function that has the opposite interpretation. In order to facilitate the comparison, Figure 5 graphically shows the *complement of navigation quality* in the  $[0, 1]$  interval and it tunes the color scale to the values observed in this example; i.e.,  $[0, 10]$  for exploration cost and  $[0, 1]$  for the complement of navigation quality.<sup>2</sup>

- In both approaches “Name” has very high cost, which is desirable because this facet identifies hundreds of subsets of items to be browsed one by one.
- According to (Oren et al., 2006), “Outdoor Seating”, “Ex1” and “Takeaway” are moderately inefficient, and “Takeaway” has higher cost than the other ones; the reason is the low coverage of these facets and, with the exception of “Ex1”, their lack of balance. Differently, our model attributes low cost to these facets because they have few values which represent non elementary sets of solutions to be inspected.

<sup>2</sup>Oren et al. (2006)’s model introduces the  $\sigma$  and  $\mu$  parameters for the computation of balance and object cardinality metrics but we could not find the exact values that they applied in their experiments. We reproduced the expected behavior, following the indications given in the paper, by setting  $\mu = 2$  and  $\sigma = 4.9$ .

- The main disagreement is in the evaluation of “Cuisine” and “Ex2”. According to (Oren et al., 2006), “Cuisine” is totally inefficient because of its partial coverage of items, lack of balance and high number of values. Moreover, “Ex2” is penalized by the lack of coverage of results. In our approach “Cuisine” has moderate cost, in spite of the many values it can take, because it identifies a few large subsets that deserve attention when browsing results, and the long tail of the facet can be ignored. “Ex2” has higher cost than “Cuisine” because it identifies very small sets of items.

In summary, our approach supports the identification of facets which are not “perfect” from the *divide et impera* viewpoint because they only occur in a subset of results and/or they split data in an unbalanced way. However, it works on realistic cases in which balanced, frequent facets are extremely rare. Moreover, it promotes facets that split results in subsets having a significant cardinality because they are valuable for browsing results.

#### 5.4. Selection of facets to be included in the widgets

In order to select the facets to be shown in the widgets, we first exclude those having  $cost(f) = 0$  because this means that they have a single value in  $E_C$ . Then, we sort facets by increasing cost and we include them in the widget up to a maximum number of 12 to avoid cluttering the user interface.

By applying Equation 1 to the results of query “amenity=restaurant” on Torino city bounding box, we obtain the following sorted list of facets: “Outdoor Seating”, “Takeaway”, “Wheelchair”, “Delivery”, “Addr city”, “Smoking”, “Building”, “Addr postcode”, “Cuisine”, “Capacity”, “Addr street” and “Opening hours”; see Figure 4. Almost all these facets correspond to semantically relevant dimensions. Only “Addr city” seems useless because the query is bounded in Torino city; however, according to the geocoder we use, the area of the map that is considered includes Torino and a few small cities in its boundary. Other facets, such as “Name”, “Phone” and “URL”, are excluded from the sunburst because they have very high cost (they are identifiers) and thus take the final positions in the ranked list. Facets such as “Cuisine 1”, which

is redundant with respect to “Cuisine”, are excluded because they are below the minimum coverage threshold. Indeed, “Cuisine 1” is the typical tag that somebody has duplicated instead of using the main “Cuisine” one.

## 6. Validation of our faceted exploration model

### 460 6.1. Study design

We conducted a user study by exploiting OnToMap to evaluate the four types of information exploration widgets described in Section 5, as far as data interpretation in a geographic map is concerned. Specifically, we were interested in comparing:

- 465 • The exploration model based on transparency sliders (which supports information hiding at the granularity level of the data category) to the more expressive one that also supports faceted exploration.
- The alternative graphical models we defined for faceted information exploration in order to understand which ones are more effective to help users  
470 in the exploration of an information space via map projection.

For the experiment we defined a simple project planning scenario concerning the preparation of a tourist trip in Torino city. We instructed each participant that (s)he should imagine to plan a tour with some friends in different areas of the town. We also explained that, for each area, (s)he would find the information  
475 about Points of Interest that might be visited (e.g., urban parks, monuments, etc.), as well as travel facilities (e.g., parking lots), by exploring a custom geographic map focused on the specific area and previously prepared by her/his friends. We aimed at separately evaluating the four widget types we defined but we wanted to minimize the learning effect on participants. Therefore, we  
480 prepared four maps, each one focused on a different geographic area of Torino city. Each map was populated with multiple data categories representing Points of Interest and travel facilities. We investigated participants’ performance and

User Experience in four map learning tasks, each one using a different type of widget and map:

- 485 • *Task1*: question answering using checkboxes in combination with transparency sliders.
- *Task2*: question answering using treemaps in combination with transparency sliders.
- *Task3*: question answering using sunburst in combination with transparency sliders.
- 490 • *Task4*: question answering using transparency sliders.

The study was a within-subjects design one. We considered each treatment condition as an independent variable and every participant received the 4 treatments. We counterbalanced the order of tasks to minimize the impact of result biases and the effects of practice and fatigue. People participated in the user study on a voluntary basis, without any compensation, and they signed a consent to the treatment of personal data. The participation to the user study took place live, i.e., we did not perform any online interviews.

## 6.2. The experiment

500 One person at a time performed the study which lasted about 30 minutes. Before starting the user study, the participant watched a video describing the widgets and showing how they work. After that, (s)he interacted with OnToMap on a sample map to get acquainted with the user interface of the system. We did not impose any restrictions on this activity and we allowed the participant to take as much time as (s)he needed in order to comply with diverse backgrounds and levels of confidence with technology. Then, we asked her/him to answer a pre-test questionnaire designed to assess demographic information, cultural background, as well as familiarity with map-based online applications.

510 During the study, we asked the participant to use OnToMap in the context of the organization of the trip. For each task (s)he had to look at the associated

map and (s)he had to answer two questions which required counting elements that have certain properties, or identifying items given their descriptions. For each specific map, all participants answered the same two questions. As far as counting is concerned, we forced the participant to analyze the map by asking  
515 her/him to answer the questions in a geographic area delimited by an orange border. In this way, (s)he could not simply read the cardinality information provided by the faceted exploration widgets, which work by taking the bounding box of the map as a reference to specify how many items satisfy the selected visualization constraints. The questions proposed to the participants had the  
520 following templates:

- How many *category name* having *characteristic<sub>1</sub>* and/or ... and/or *characteristic<sub>n</sub>* are visualized within the area delimited by the orange line in the map?

For instance, “How many Christian churches accessible to wheelchairs are  
525 visualized within the area delimited by the orange line in the map?”. In the question, “Christian” is a value of facet “Religion” and wheelchair accessibility corresponds to value “YES” of facet “Wheelchair”.

- Find *category name* having *characteristic<sub>1</sub>* and/or ... and/or *characteristic<sub>n</sub>* within the orange line in the map, and list them.

530 E.g., find restaurants serving pizza or Italian food (values of “Cuisine”).

In *Task1*, *Task2* and *Task3*, we proposed selective questions because we wanted to understand whether the widgets helped participants satisfy specific information needs by exploring the metadata of the searched categories and by projecting the maps accordingly. Differently, the questions of *Task4* did not require the  
535 imposition of any visualization constraints because participants only used the transparency sliders; in this task we assessed the general usefulness of category-based map projection in reducing the visual complexity of a map that includes diverse types of information. This function was appreciated by users in a previous experiment (Ardissono et al., 2018) but we wanted to evaluate it extensively.



Table 3: Post-task questionnaire (translated from the Italian language).

| # | Question  |
|---|---|
| 1 | How familiar are you with the widget that you just used?                                    |
| 2 | How much did the widget help you find the information that you were looking for in the map? |
| 3 | How much did the widget help you save effort in answering the questions we asked you?       |
| 4 | Please, rate the ease of use of the widget you just used.                                   |
| 5 | Please, rate the novelty of the widget you just used.                                       |
| 6 | Did you encounter any difficulties in finding the information that you were looking for?    |
| 7 | Is there any aspect of the widget you used that you particularly appreciated?               |

540 While the participant carried out a task, the experimenter took notes about how much time (s)he used to answer the questions, sitting at some distance from her/him. We did not put any time restrictions on question answering and we allowed checking the answers multiple times.

As objective performance indicators, we measured task completion time and  
 545 the percentage of correctly answered questions. As a subjective measure, we analyzed User Experience: after the completion of each task, the participant filled in a post-task questionnaire to evaluate the type of widget (s)he had just used. We were interested in evaluating the following traits of the facet-based widgets: familiarity, helpfulness in finding information, effort saving in solving  
 550 the task to be performed, ease of use and novelty. We defined the questions to be posed by taking inspiration from NASA TLX questionnaire (NASA, 2019); however, for simplicity, we kept a 5-points Likert scale for the expression of ratings. Table 3 shows our questionnaire: for questions 1-5 the participant had to provide values from 1, the worst value, to 5, the best one; questions 6 and 7  
 555 were open to free text comments.

After the completion of the four tasks the participant filled in a post-test

questionnaire to compare the widgets.<sup>3</sup> We also asked her/him to provide feedback to improve the User Experience in OnToMap. For the experiments we used a set of laptops with 15.6" display and 1920x1080 resolution.

## 560 7. Results

### 7.1. Demographic data and background

For the user study, we recruited 62 participants (32.3% women, 66.1% men and 1.6% not declared). Their age is between 20 to 70 years, with a mean value of 33.45. They are part of the University staff (researchers, professors and secretaries) and students, as well as people working in the industry or retired. In the pre-test questionnaire we analyzed their background and familiarity with technology: 41.9% of participants have a scientific background, 29% a technical one, 21% humanities and linguistics, 6.5% economics and law, 1.6% arts. Regarding the education level, 46.8% of them attended the high school, 45.2% the university, 6.5% have a Ph.D and 1.6% attended the middle school. 41.9% of people declared that they use e-commerce platforms or online booking services monthly, 38.7% said one or two times per year and 19.4% weekly. Moreover, 56.9% declared that they often use online services based on geographic maps, 17.7% sometimes and 25.8% every day.

### 575 7.2. User performance

Table 4 shows the results concerning participants' execution time and percentage of correct answers for each task. A Friedman test on execution times among the four tasks showed that there is a statistically significant difference between them:  $\chi^2(3) = 207.57$ ,  $p < 0.001$ , Kendall's  $W = 0.56$ . The percentages of correct answers is statistically significant, too:  $\chi^2(3) = 14.14$ ,  $p < 0.002$ , Kendall's  $W = 0.04$ .

---

<sup>3</sup>Also in this case, we took inspiration from User Experience Questionnaire and NASA TLX.

Table 4: Participants’ performance during the execution of individual tasks. Time is expressed in seconds and the best values are in boldface. Significance is encoded as (\*\*)  $p < 0.001$  and (\*)  $p < 0.002$ .

| Widget type             | Min time  | Max time   | Mean time      | Correct answers |
|-------------------------|-----------|------------|----------------|-----------------|
| 1: Checkboxes           | 33        | 184        | 94.26          | <b>100.00%*</b> |
| 2: Treemaps             | 33        | 180        | 77.39          | 98.39%          |
| 3: Sunburst             | <b>20</b> | 149        | <b>55.94**</b> | <b>100.00%*</b> |
| 4: Transparency sliders | 23        | <b>146</b> | 57.05          | 95.16%          |

As shown in the table, people achieved the lowest mean execution time and they correctly answered 100% of the questions when they used the widget based on the sunburst diagram. In comparison, when they used the checkboxes, they  
585 correctly answered all the questions but they spent the longest time to complete the task. By using the treemaps, participants spent a long time to perform the tasks (almost as long as with checkboxes) but they correctly answered 98.39% of the questions. Finally, they spent relatively little time with transparency sliders but they provided 95.16% correct answers. The high number of correct  
590 answers should not surprise because people could check them more than once.

We observed that, in *Task1*, *Task2* and *Task3*, almost all the participants removed some irrelevant data categories using the transparency sliders to reduce map cluttering; then, they used faceted exploration to analyze data. However, they leaned to use the checkboxes embedded in the transparency sliders instead  
595 of using the sliders to tune the opacity of items.

### 7.3. User Experience - post-task questionnaire

Table 5 shows the results of questions 1-5 of the post-task questionnaire and Table 6 shows the results of a Friedman significance test applied to these results.

- **Question 1 (familiarity):** participants were most familiar with the widgets based on checkboxes and, in second position, with the transparency  
600 sliders. They were less familiar with the treemaps and much less with the

Table 5: Results of the post-task questionnaire. The best values are shown in boldface. Significance is encoded as (\*\*)  $p < 0.001$  and ( $\diamond$ )  $p < 0.03$ .

| Question #                  | 1             | 2           | 3           | 4                                | 5             |
|-----------------------------|---------------|-------------|-------------|----------------------------------|---------------|
| Task1: Checkboxes           |               |             |             |                                  |               |
| Mean                        | <b>3.90**</b> | 4.03        | 3.77        | 4.02                             | 2.94          |
| Variance                    | 1.40          | 1.08        | 1.39        | 0.84                             | 1.31          |
| St. Dev.                    | 1.18          | 1.04        | 1.18        | 0.91                             | 1.14          |
| Task2: Treemap              |               |             |             |                                  |               |
| Mean                        | 3.32          | 4.00        | <b>3.95</b> | 3.98                             | 3.48          |
| Variance                    | 1.21          | 0.66        | 0.87        | 0.84                             | 0.84          |
| St. Dev.                    | 1.10          | 0.81        | 0.93        | 0.91                             | 0.92          |
| Task3: Sunburst             |               |             |             |                                  |               |
| Mean                        | 2.95          | <b>4.11</b> | 3.84        | 3.87                             | <b>4.10**</b> |
| Variance                    | 1.62          | 0.72        | 1.22        | 0.84                             | 0.97          |
| St. Dev.                    | 1.27          | 0.85        | 1.10        | 0.91                             | 0.99          |
| Task4: Transparency sliders |               |             |             |                                  |               |
| Mean                        | 3.79          | 3.85        | 3.69        | <b>4.31<math>\diamond</math></b> | 3.02          |
| Variance                    | 1.28          | 1.21        | 1.20        | 0.87                             | 1.52          |
| St. Dev.                    | 1.13          | 1.10        | 1.10        | 0.93                             | 1.23          |

sunburst diagrams ( $p < 0.001$ ).

- **Question 2 (helpfulness):** the results are not statistically significant but the generally high ratings prove that participants perceived all the widgets as helpful to find information items in the maps. The transparency sliders received the lowest ratings.
- **Question 3 (effort saving):** the results are not statistically significant; however, similar to Question 2, the transparency sliders are evaluated worse than the other widgets. In this case, ratings show that participants

605

Table 6: Statistical significance of the post-task questionnaire results.

| Question # | Friedman's $\chi^2$ | df | $p$ -value | Kendall's $W$ |
|------------|---------------------|----|------------|---------------|
| 1          | 25.038              | 3  | < 0.001    | 0.1346        |
| 2          | 4.6779              | 3  | = 0.197    | 0.0251        |
| 3          | 1.4063              | 3  | = 0.704    | 0.0076        |
| 4          | 9.4442              | 3  | < 0.03     | 0.0508        |
| 5          | 43.611              | 3  | < 0.001    | 0.2345        |

610 felt that the widgets helped them to save efforts during task execution but  
values are a bit lower than those of Question 1.

- **Question 4 (ease of use):** participants perceived transparency sliders as the easiest tool, followed by the checkboxes, treemaps and sunburst diagram ( $p < 0.03$ ). This finding is in line with the results of Question 1:  
615 even though sliders and checkboxes are in a different preference order, the Pearson Correlation between the answers to Question 1 and Question 4 shows that they are positively correlated both on checkboxes ( $\rho = 0.5015$ ) and on transparency sliders ( $\rho = 0.4802$ ).

- **Question 5 (novelty):** participants perceived the widgets based on the  
620 treemaps and sunburst diagrams as more innovative than the other ones; they also evaluated the checkboxes as the least innovative one ( $p < 0.001$ ). The Kendall's  $W$  value (0.2345) is the best one across the five questions. This demonstrates that there is more agreement among participants about the perception of novelty of widgets with respect to the other evaluation  
625 dimensions, i.e., ease of use, and so forth.

About a quarter of the participants answered the free text questions; the percentages reported below refer to this set of people.

- **Question 6 (difficulties):** 50% of the participants who answered this question declared that, due to the amount of textual information displayed  
630 in the checkboxes, they had difficulties in the identification of the widgets

Table 7: Post-test questionnaire (translated from the Italian language).

| #  | Question/statement   |
|----|--|
| 1  | The widget was familiar to me.   |
| 2  | The widget helped me find the information I needed.  |
| 3  | The widget helped me to save effort in answering the questions.  |
| 4  | The widget was easy to use.  |
| 5  | The widget is novel.   |
| 6  | Do you think that using transparency sliders in combination with checkboxes, treemaps or sunburst diagram is useful? |
| 7  | Which information exploration widget would you use again in the future?  |
| 8  | Why?   |
| 9  | Which information exploration widget did you like the least?   |
| 10 | Why?   |

representing the categories of interest in the side bar. Some people pointed out that the treemap and the sunburst were new visualization models; thus, they initially had some difficulties in understanding how they worked. A few participants complained about the shortening of facet values in the treemaps because they had to move the mouse over their components to read the information. The only observed limitation of the sunburst was that it is visualized in a separate window, partially covering the map.

- **Question 7 (appreciations):** some participants liked the graphics of the treemaps and declared that the size of the components representing facet values provides an intuitive visualization of the cardinality of the corresponding sets of items. About 25% of people perceived the sunburst as good to compactly visualize all the facets and values of a data category. They also appreciated the fact that the sunburst reduces the vertical expansion of the side bar; thus, it limits the scrolling to reach the widgets of

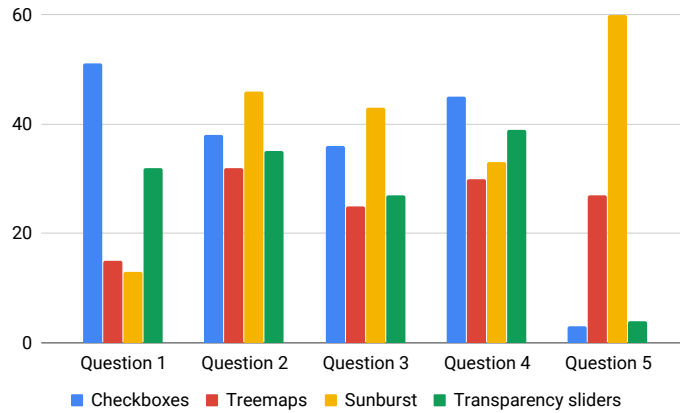


Figure 6: Post-test: evaluations of the questions listed in Table 3.

645 interest. Some participants specified that they liked the correspondence of colors between sliders and items in the map; i.e., color coding. In general, the transparency slider was perceived as useful to reduce information overload by imposing visualization constraints on whole data categories.

#### 7.4. User Experience - post-test questionnaire

650 After participants completed the four tasks, we asked them to fill in a post-test questionnaire to capture their overall experience with the widgets.

In the first part of this test we asked them to select the widgets which better matched familiarity, helpfulness, effort saving, easy of use and novelty; see Table 7. People could check multiple options in case more than one widget satisfied them; therefore, the percentages reported below may be over 100%. The results 655 of this part of the test, shown in Figure 6, are consistent with those of the post-task questionnaires. Specifically, they confirm that:

- **Question 1 (familiarity):** people considered the checkboxes as the most familiar widget and they placed transparency sliders in second position.
- 660 • **Question 2 (helpfulness):** the sunburst was perceived as more helpful than the other widgets as far as information finding is concerned.

- **Question 3 (effort saving):** the sunburst, followed by the checkboxes, was the preferred widget from the viewpoint of effort saving. This is different from the results of the post-task questionnaires but it should be noticed that those results are not statistically significant.
- **Question 4 (ease of use):** the checkboxes were evaluated as the easiest widget to use, slightly easier than transparency sliders. In the post-task results were reversed but the two widgets were anyway the best rated ones.
- **Question 5 (novelty):** the sunburst was perceived as the most novel widget.

Regarding the second part of the post-test questionnaire (see Table 7):

- **Question 6 (transparency sliders with facet-based widget):** 53.2% of participants declared that the joint usage of transparency sliders with checkboxes, treemaps or sunburst efficiently helps information exploration. They found it convenient to organize search in two steps: (i) visual simplification of maps by hiding the data categories irrelevant to the questions, using transparency sliders; (ii) identification of the items of the category of interest on the basis of their properties, using faceted exploration widgets.
- **Questions 7 and 8 (future usage of widgets and why):** 60% of people stated that they would use the widget based on the sunburst again because it offers a complete view of each data category. Moreover, 56% declared that they would use the checkboxes again because this is a widespread way to search for information.
- **Questions 9 and 10 (least preferred widget and why):** 34% of participants evaluated the treemaps as the least preferred widget because they are not intuitive and they are difficult to use; 37% did not like the transparency sliders either because they poorly help solving complex search tasks. 21% of people did not like the sunburst, mostly because it covers part of the map instead of being displayed within the side bar. Finally,



690 15% declared that they would not use the checkboxes in the future because they carry a large amount of textual information and it's difficult to identify the relevant values.

## 8. Discussion

The user performance and experience results consistently suggest that the  
695 sunburst is the best widget for faceted information exploration. Specifically, User Experience results can be explained as follows:

- Regarding the familiarity with the types of widget (Question 1), we expected that people would be more familiar with checkboxes and transparency sliders because they are used to support faceted search in several  
700 e-commerce and booking applications, while treemaps and sunburst are rarely used outside scientific contexts.
- Question 2 provides some evidence that participants perceived the widget based on the sunburst as the most helpful one (post-test), while transparency sliders were suitable to solve simple search problems because they  
705 do not support facet-based map projection (post-task and post-test). People also considered the transparency slider as useful within a facet-based widget (Question 6 - post-test).
- As far as effort saving is concerned (Question 3), the moderate appreciation and the mixed ratings given by participants might be explained  
710 by considering that, even though all the widgets support map projection, they require some interaction, which could be perceived as an effort.
- Participants' familiarity with the widgets can explain the fact that they evaluated transparency sliders and checkboxes as the top easy-to-use tools (Question 4), and treemaps and sunburst as the most novel ones (Question 5). Moreover, the moderate ease of use attributed to treemaps and  
715 sunburst can partially depend on the fact that people had to learn how to use them (Question 6 - post-task).

Interestingly, in the answers to the free text questions (Question 6 - post-task and Questions 9 and 10 - post-test) a relevant number of participants criticized  
720 the amount of textual information visualized in the checkboxes, complaining that it challenges the identification of the relevant widgets or values in the side bar. Actually, all the faceted widgets include the same information, generated as described in Section 5. Therefore, this comment can be interpreted in a different way, in relation to the lack of compactness of the layout provided by  
725 the checkboxes (and presumably also by the treemaps, even though they save a bit more space than the former).

The widget based on the treemaps was the least preferred one because it was not particularly intuitive and it was difficult to use (Questions 9 and 10 - post-test). Despite the appeal of its graphics (Question 7 - post-task), this widget  
730 challenges the user with readability issues. Moreover, similar to the checkboxes, it occupies a fairly relevant amount of vertical space in the side bar (Question 6 - post-task), thus increasing the amount of scrolling needed to inspect the other treemaps.

We conclude that the experimental results help us answer our research ques-  
735 tions, which we repeat here for the reader's convenience:

RQ1: *Does faceted exploration of map content help users in finding the needed information in a geographic map that visualizes different types of data?*

We compared participants' performance and experience using different widgets for faceted information exploration, with respect to transparency  
740 sliders alone. By using some of these widgets, people could complete a set of map learning tasks more quickly and/or precisely than by only using the sliders. However, participants appreciated the combination of faceted exploration with basic category hiding because the latter enables the user to quickly hide irrelevant data, thus reducing visual complexity in the  
745 map, and the former supports detailed exploration of relevant items. This is different from traditional faceted exploration, in which users search for information within a single category. It also indicates that, when maps

show multiple types of information, faceted exploration is effective but can be strengthened by adopting a model that jointly supports coarse-grained and fine-grained data projection.

RQ2: *How does a compact, graphical view of the exploration options available to the user, which also shows the status of the information visualization constraints applied to a map, impact on her/his efficiency and experience in data exploration?*

The results of the experiment show that not all the facet-based widgets equally helped participants while executing the tasks of the experiment. The reason for this difference is in the capability of the widgets to clearly and compactly describe the search context.

Specifically, the widget based on the sunburst was considered as particularly useful and effective, and it supported the best user performance. This finding is in line with previous experiments; e.g., those by Stasko et al. (2000). It can be explained by the fact that the sunburst provides a compact representation of the facets of a category, supporting the readability of their values. The compactness of this widget also enhances the conciseness of the side bar; in turn, this reduces scrolling during faceted search. Conversely, the treemaps challenged participants because their graphical layout hampers readability and their vertical extension excessively increases the length of the side bar.

Participants appreciated both sunburst and checkboxes; however, when using the latter they completed the tasks slower than with the former. As the main difference between these two widgets is in their vertical extension (much more compact in the sunburst), we can say that this is the main dimension determining the difference in performance.

The transparency sliders achieved the lowest user performance results because they fail to support the specification of fine-grained visualization constraints. However, using the sliders in combination with the other widgets was perceived as a very convenient approach because it enables to

first focus the map on the categories of interest, and then further project it by imposing detailed visualization constraints on the remaining items.

780 RQ3: *How much does the user's familiarity with the widgets for faceted exploration impact on her/his efficiency in search and on her/his appreciation of the exploration model they offer?*

The results of the experiment suggest that the familiarity with the widgets does not influence users' efficiency in search: the best performance in task  
785 execution was achieved by using the sunburst, which most participants considered as moderately ease to use and they did not know before interacting with OnToMap. Moreover, familiarity positively influences people's disposition towards the faceted exploration widgets and their perception of ease of use; see the case of the checkboxes. However, participants ap-  
790 preciated the sunburst as well because it efficiently supports exploration at the expense of some initial learning effort. We can thus conclude that, if the widget is not too difficult to use, the functionality it provides can override the effect of its familiarity on user appreciation.

## 9. Conclusions and future work

795 We presented a faceted information exploration approach supporting a flexible visualization of heterogeneous geographic data. Our model provides a multi-category faceted projection of long-lasting geographic maps to answer temporary information needs; this is based on the proposal of efficient facets for data exploration in sparse and noisy datasets. Moreover, the model provides a graphical  
800 representation of the search context by means of alternative types of widget that support interactive data visualization, faceted exploration, category-based information hiding and transparency of results at the same time.

We carried out a user study involving 62 people who have diverse familiarity with technology and with map-based online systems. The results of this study  
805 show that, when working on maps populated with multiple data categories, our model outperforms simple category-based map projection and traditional

faceted search tools such as checkboxes. Moreover, the layout that uses the sunburst diagram as a graphical widget supports the best user performance and experience, thanks to its clarity and visual compactness. We thus conclude that  
810 this implementation is promising for flexible faceted exploration in Geographic Information Search. The described work has limitations that we plan to address:

- Our model only supports the specification of hard visualization constraints on facet values; i.e., the items having a certain value of a facet are either shown, or hidden. However, the user might want to specify preferences.  
815 Therefore, similar to what has been done in some related works (see Section 3), we plan to manage soft visualization constraints.
- So far, we present search results in geographic maps and we provide item details in dynamically generated tables showing their properties. In order to enhance data interpretation and sensemaking, we plan to develop  
820 additional visualization models supporting visual analytics; e.g., see (Andrienko et al., 2007; Tsai & Brusilovsky, 2019; Cardoso et al., 2019).
- We designed the questionnaires of our user study by taking inspiration from existing sources (NASA TLX and User Experience Questionnaire) but we personalized the questions in order to test the specific aspects which  
825 are the focus of the present paper. We plan a credibility/validity analysis to verify that our questionnaires are strictly related to these sources.
- Further experiments are needed to validate the proposed model with a larger set of people and on mobile phones (the OnToMap user interface scales well to the screens of tablets).
- Currently, our model supports a “one size fits all” type of faceted search  
830 that exploits general efficiency criteria to guide the user in data exploration. However, some researchers propose to adapt facet suggestion to the user’s preferences in order to personalize the navigation of the information space; e.g., see (Tvarožek et al., 2008; Tvarožek & Bieliková,  
835 2010; Koren et al., 2008; Abel et al., 2011). In our future work, we plan

to offer multiple data exploration strategies which the user can choose from, including a user-adaptive facet suggestion that depends on her/his preferences and on the search context.

- Depending on their roles, in some scenarios users might need to access different, long-lasting custom views of a shared information space (Rasmussen & Hertzum, 2013). We plan to extend our model by introducing permanent, user-dependent views on map content.

## 10. Acknowledgements

This work was supported by the University of Torino which funded the conduct of the research and preparation of the article.

## References

- Abel, F., Celik, I., Houben, G.-J., & Siehndel, P. (2011). Leveraging the semantics of tweets for adaptive faceted search on twitter. In L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, & E. Blomqvist (Eds.), *The Semantic Web – ISWC 2011* (pp. 1–17). Berlin, Heidelberg: Springer Berlin Heidelberg. URL: [https://doi.org/10.1007/978-3-642-25073-6\\_1](https://doi.org/10.1007/978-3-642-25073-6_1). doi:10.1007/978-3-642-25073-6\_1.
- Ahlberg, C., & Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '94* (pp. 313–317). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/191666.191775>. doi:10.1145/191666.191775.
- Al-Kodmany, K. (1999). Using visualization techniques for enhancing public participation in planning and design: process, implementation, and evaluation. *Landscape and urban planning*, *45*, 37–45. URL: [https://10.1016/S0169-2046\(99\)00024-9](https://10.1016/S0169-2046(99)00024-9). doi:10.1016/S0169-2046(99)00024-9.

- Andrienko, G. L., Andrienko, N. V., Jankowski, P., Keim, D. A., Kraak, M.-J., MacEachren, A., & Wrobel, S. (2007). Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, *21*, 839–857. URL: <https://doi.org/10.1080/13658810701349011>. doi:10.1080/13658810701349011.
- 865
- Ardissono, L., Delsanto, M., Lucenteforte, M., Mauro, N., Savoca, A., & Scanu, D. (2018). Map-based visualization of 2D/3D spatial data via stylization and tuning of information emphasis. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces AVI '18* (pp. 38:1–38:5). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/3206505.3206516>. doi:10.1145/3206505.3206516.
- 870
- Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., Voghera, A., & La Riccia, L. (2016). Exploration of cultural heritage information via textual search queries. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct MobileHCI '16* (pp. 992–1001). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/2957265.2962648>. doi:10.1145/2957265.2962648.
- 875
- Ardissono, L., Lucenteforte, M., Mauro, N., Savoca, A., Voghera, A., & La Riccia, L. (2017). OnToMap: Semantic community maps for knowledge sharing. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media HT '17* (p. 317–318). New York, NY, USA: Association for Computing Machinery. URL: <https://doi.org/10.1145/3078714.3078747>. doi:10.1145/3078714.3078747.
- 880
- 885 Ardito, C., Costabile, M. F., Desolda, G., Lanzilotti, R., Matera, M., Piccinno, A., & Picozzi, M. (2013). Personal information spaces in the context of visits to archaeological parks. In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI CHIItaly '13*. New York, NY, USA: Association for Computing Machinery. URL: <https://doi.org/10.1145/2499149.2499173>. doi:10.1145/2499149.2499173.
- 890

- 895 Ardito, C., Costabile, M. F., Desolda, G., & Matera, M. (2016). Supporting professional guides to create personalized visit experiences. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct MobileHCI '16* (pp. 1010–1015). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/2957265.2962650>. doi:10.1145/2957265.2962650.
- 900 Ballatore, A., Bertolotto, M., & Wilson, D. C. (2013). Geographic knowledge extraction and semantic similarity in OpenStreetMap. *Knowledge and Information Systems*, *37*, 68–81. URL: <https://doi.org/10.1007/s10115-012-0571-0>. doi:10.1007/s10115-012-0571-0.
- 905 Brown, G., & Weber, D. (2012). Measuring change in place values using public participation GIS (PPGIS). *Applied Geography*, *34*, 316 – 324. URL: <http://www.sciencedirect.com/science/article/pii/S0143622811002438>. doi:<https://doi.org/10.1016/j.apgeog.2011.12.007>.
- 910 Cao, N., Gotz, D., Sun, J., Lin, Y., & Qu, H. (2011). SolarMap: multifaceted visual analytics for topic exploration. In *Proceedings of the 11th IEEE International Conference on Data Mining KDD '08* (pp. 101–110). Vancouver, BC, Canada: IEEE. URL: <https://doi.org/10.1109/ICDM.2011.135>. doi:10.1109/ICDM.2011.135.
- Cao, N., Sun, J., Lin, Y.-R., Gotz, D., Liu, S., & Qu, H. (2010). FacetAtlas: multifaceted visualization for rich text corpora. *IEEE Transactions on visualization and Computer Graphics*, *16*, 1172 – 1181. URL: <https://doi.org/10.1109/TVCG.2010.154>. doi:10.1109/TVCG.2010.154.
- 915 Cardoso, B., Sedrakyan, G., Gutiérrez, F., Parra, D., Brusilovsky, P., & Verbert, K. (2019). IntersectionExplorer, a multi-perspective approach for exploring recommendations. *International Journal of Human-Computer Studies*, *121*, 73 – 92. URL: <http://www.sciencedirect.com/science/article/pii/S1071581918301903>. doi:10.1016/j.ijhcs.2018.04.008.



- 920 Chang, J. C., Hahn, N., Perer, A., & Kittur, A. (2019). SearchLens: Composing and capturing complex user interests for exploratory search. In *Proceedings of the 24th International Conference on Intelligent User Interfaces IUI '19* (pp. 498–509). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/3301275.3302321>. doi:10.1145/3301275.3302321.
- 925 Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., & Rau, R. (2011). OSMonto – an ontology of openstreetmap tags. In *Proceedings of State of the map Europe (SOTM-EU) 2011* (pp. 55–64). Vienna, Austria.
- Colby, G., & Sholl, L. (1991). Transparency and blur as selective cues for complex visual information. In *Proceedings of Image Handling and Reproduction Systems Integration* (pp. 114–125). SPIE Digital Library volume 1460. URL: <https://doi.org/10.1117/12.44415>. doi:10.1117/12.44415.
- 930 Coulton, C., Chan, T., & Mikelbank, K. (2011). Finding place in community change initiatives: Using GIS to uncover resident perceptions of their neighborhoods. *Journal of Community Practice*, 19, 10–28. URL: <https://www.tandfonline.com/doi/abs/10.1080/10705422.2011.550258>. doi:10.1080/10705422.2011.550258.
- 935 Dachsel, R., & Frisch, M. (2007). Mambo: A facet-based zoomable music browser. In *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia MUM '07* (pp. 110–117). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/1329469.1329484>. doi:10.1145/1329469.1329484.
- 940 Dachsel, R., Frisch, M., & Weiland, M. (2008). FacetZoom: A continuous multi-scale widget for navigating hierarchical metadata. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '08* (pp. 1353–1356). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/1357054.1357265>. doi:10.1145/1357054.1357265.
- 945 Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., & Yee, K.-P. (2002). Finding the flow in web site search. *Communications of the ACM*, 45,

- 42–49. URL: <http://doi.acm.org/10.1145/567498.567525>. doi:10.1145/  
950 567498.567525.
- Hearst, M. A. (2006). Design recommendations for hierarchical faceted search interfaces. In *Proceedings of SIGIR 2006, Workshop on Faceted Search* (pp. 26–30).
- Hildebrand, M., van Ossenbruggen, J., & Hardman, L. (2006). /facet: A  
955 browser for heterogeneous semantic web repositories. In I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, & L. M. Aroyo (Eds.), *The Semantic Web - ISWC 2006* (pp. 272–285). Berlin, Heidelberg: Springer Berlin Heidelberg. URL: [https://doi.org/10.1007/11926078\\_20](https://doi.org/10.1007/11926078_20). doi:10.1007/11926078\_20.
- 960 Hoerber, O., & Yang, X. D. (2006). A comparative user study of web search interfaces: HotMap, concept highlighter, and google. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence WI '06* (pp. 866–874). Washington, DC, USA: IEEE Computer Society. URL: <https://doi.org/10.1109/WI.2006.6>. doi:10.1109/WI.2006.6.
- 965 Hu, Y., Lv, Z., Wu, J., Janowicz, K., Zhao, X., & Yu, B. (2015). A multistage collaborative 3D GIS to support public participation. *International Journal of Digital Earth*, 8, 212–234. URL: <https://doi.org/10.1080/17538947.2013.866172>. doi:10.1080/17538947.2013.866172.
- Koren, J., Zhang, Y., & Liu, X. (2008). Personalized interactive faceted search.  
970 In *Proceedings of the 17th International Conference on World Wide Web WWW '08* (pp. 477–486). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/1367497.1367562>. doi:10.1145/1367497.1367562.
- Lee, B., Smith, G., Robertson, G. G., Czerwinski, M., & Tan, D. S. (2009). FacetLens: Exposing trends and relationships to support sensemaking within  
975 faceted datasets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '09* (pp. 1293–1302). New York,

NY, USA: ACM. URL: <http://doi.acm.org/10.1145/1518701.1518896>.  
doi:10.1145/1518701.1518896.

Lionakis, P., & Tzitzikas, Y. (2017). PFSgeo: Preference-enriched faceted search  
980 for geographical data. In H. Panetto, C. Debruyne, W. Gaaloul, M. Papazoglou, A. Paschke, C. A. Ardagna, & R. Meersman (Eds.), *On the Move to Meaningful Internet Systems. OTM 2017 Conferences* (pp. 125–143). Cham: Springer International Publishing.

Lobo, M.-J., Pietriga, E., & Appert, C. (2015). An evaluation of interactive  
985 map comparison techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems CHI '15* (pp. 3573–3582). New York, NY, USA: ACM. URL: <http://doi.acm.org/10.1145/2702123.2702130>. doi:10.1145/2702123.2702130.

Marchionini, G. (2006). Exploratory search: From finding to understanding.  
990 *Communications of the ACM*, 49, 41–46. URL: <https://doi.org/10.1145/1121949.1121979>. doi:10.1145/1121949.1121979.

Mauro, N., Ardissono, L., Di Rocco, L., Bertolotto, M., & Guerrini, G. (2019).  
Impact of semantic granularity on geographic information search support.  
In *Proceedings of 2018 IEEE/WIC/ACM International Conference on Web  
995 Intelligence (WI)* (pp. 323–328). Santiago, Chile: IEEE. URL: <https://ieeexplore.ieee.org/document/8609610>. doi:10.1109/WI.2018.00-73.

Mauro, N., Ardissono, L., & Savoca, A. (2017). Concept-aware geographic in-  
formation retrieval. In *Proceedings of the International Conference on Web  
Intelligence WI '17* (p. 34–41). New York, NY, USA: Association for Comput-  
1000 ing Machinery. URL: <https://doi.org/10.1145/3106426.3106498>. doi:10.1145/3106426.3106498.

Moro, A., Raganato, A., & Navigli, R. (2014). Entity Linking meets Word Sense  
Disambiguation: a Unified Approach. *Transactions of the Association for  
Computational Linguistics (ACL)*, 2, 231–244. URL: [https://www.aclweb.  
1005 org/anthology/Q14-1019](https://www.aclweb.org/anthology/Q14-1019). doi:10.1162/tac1\_a\_00179.

- NASA (2019). NASA TLX - Task Load Index. <https://humansystems.arc.nasa.gov/groups/tlx/>.
- OCG (2017). GeoSPARQL - a geographic query language for RDF data. <http://www.opengeospatial.org/standards/geosparql>.
- 1010 OpenStreetMap Contributors (2017). Openstreetmap. <https://www.openstreetmap.org>.
- OpenStreetMap Wiki contributors (2019). OverpassTurbo. <https://overpass-turbo.eu/>: OpenStreetMap Wiki.
- Oren, E., Delbru, R., & Decker, S. (2006). Extending faceted navigation for RDF  
1015 data. In I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, & L. M. Aroyo (Eds.), *The Semantic Web - ISWC 2006* (pp. 559–572). Berlin, Heidelberg: Springer Berlin Heidelberg. URL: [https://doi.org/10.1007/11926078\\_40](https://doi.org/10.1007/11926078_40). doi:10.1007/11926078\_40.
- OSM Contributors (2019). taginfo. <https://taginfo.openstreetmap.org/>.
- 1020 Papadakos, P., & Tzitzikas, Y. (2014). Hippalus: preference-enriched faceted exploration. In *Proceedings of the workshops of EDBT/ICDT Joint Conference* (pp. 167–172). Athens, Greece: CEUR volume 1133.
- Peltonen, J., Belorustceva, K., & Ruotsalo, T. (2017). Topic-relevance map: Visualization for improving search result comprehension. In *Proceedings  
1025 of the 22nd International Conference on Intelligent User Interfaces IUI '17* (p. 611–622). New York, NY, USA: Association for Computing Machinery. URL: <https://doi.org/10.1145/3025171.3025223>. doi:10.1145/3025171.3025223.
- Petrelli, D., Mazumdar, S., Dadzie, A.-S., & Ciravegna, F. (2009). Multi vi-  
1030 sualization and dynamic query for effective exploration of semantic data. In A. Bernstein, D. R. Karger, T. Heath, L. Feigenbaum, D. Maynard, E. Motta, & K. Thirunarayan (Eds.), *The Semantic Web - ISWC 2009* (pp. 505–520).

Berlin, Heidelberg: Springer Berlin Heidelberg. URL: [https://doi.org/10.1007/978-3-642-04930-9\\_32](https://doi.org/10.1007/978-3-642-04930-9_32). doi:10.1007/978-3-642-04930-9\_32.

1035 Quercia, D., Schifanella, R., & Aiello, L. M. (2014). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media HT '14* (pp. 116–125). New York, NY, USA: ACM. URL: <https://dl.acm.org/doi/10.1145/2631775.2631799>. doi:<https://doi.org/10.1145/2631775.2631799>.

1040 Rasmussen, R., & Hertzum, M. (2013). Visualizing the application of filters: A comparison of blocking, blurring, and colour-coding whiteboard information. *International Journal of Human-Computer Studies*, 71, 946 – 957. URL: <http://www.sciencedirect.com/science/article/pii/S1071581913000839>. doi:<https://doi.org/10.1016/j.ijhcs.2013.06.002>.

Sacco, G. M. (2000). Dynamic taxonomies: A model for large information bases. *IEEE Transactions on Knowledge and Data Engineering*, 12, 468–479. URL: <https://doi.org/10.1109/69.846296>. doi:10.1109/69.846296.

1050 Shneiderman, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11, 92–99. URL: <https://doi.org/10.1145/102377.115768>. doi:10.1145/102377.115768.

1055 Simpson, D. (2001). Virtual reality and urban simulation in planning: a literature review and topical bibliography. *Journal of Planning Literature*, 15, 359–376. URL: <https://journals.sagepub.com/doi/10.1177/08854120122093078>. doi:10.1177/08854120122093078.

1060 Stadler, C., Martin, M., & Auer, S. (2014). Exploring the web of spatial data with Facete. In *Proceedings of the 23rd International Conference on World Wide Web WWW '14 Companion* (p. 175–178). New York, NY, USA: Association for Computing Machinery. URL: <https://doi.org/10.1145/2567948.2577022>. doi:10.1145/2567948.2577022.

- 1065 Stasko, J., Catrambone, R., Guzdial, M., & McDonald, K. (2000). An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies*, 53, 663 – 694. URL: <http://www.sciencedirect.com/science/article/pii/S1071581900904208>. doi:<https://doi.org/10.1006/ijhc.2000.0420>.
- 1070 Sutcliffe, A., Ennis, M., & Hu, J. (2000). Evaluating the effectiveness of visual user interfaces for information retrieval. *International Journal of Human-Computer Studies*, 53, 741–763. URL: <http://dx.doi.org/10.1006/ijhc.2000.0416>. doi:10.1006/ijhc.2000.0416.
- Tsai, C.-H., & Brusilovsky, P. (2019). Exploring social recommendations with visual diversity-promoting interfaces. *ACM Transactions on Interactive Intelligent Systems*, 10, 5:1–5:34. URL: <http://doi.acm.org/10.1145/3231465>. doi:10.1145/3231465.
- 1075 Tvarožek, M., Barla, M., Frivolt, G., Tomša, M., & Bieliková, M. (2008). Improving semantic search via integrated personalized faceted and visual graph navigation. In *Proceedings of the 34th Conference on Current Trends in Theory and Practice of Computer Science SOFSEM'08* (pp. 778–789). Berlin, Heidelberg: Springer-Verlag. URL: <http://dl.acm.org/citation.cfm?id=1785934.1786006>. doi:10.1007/978-3-540-77566-9\_67.
- 1080 Tvarožek, M., & Bieliková, M. (2010). Generating exploratory search interfaces for the semantic web. In P. Forbrig, F. Paternó, & A. Mark Pejtersen (Eds.), *Human-Computer Interaction* (pp. 175–186). Berlin, Heidelberg: Springer Berlin Heidelberg. URL: [https://doi.org/10.1007/978-3-642-15231-3\\_18](https://doi.org/10.1007/978-3-642-15231-3_18). doi:10.1007/978-3-642-15231-3\_18.
- 1085 Tzitzikas, Y., Manolis, N., & Papadakos, P. (2017). Faceted exploration of RDF/S datasets: a survey. *Journal of Intelligent Information Systems*, 48, 329–364. URL: <https://doi.org/10.1007/s10844-016-0413-8>. doi:10.1007/s10844-016-0413-8.

- 1090 Voghera, A., Crivello, R., L.Ardissono, Lucenteforte, M., Savoca, A., & L. La  
Riccia (2016). Production of spatial representations through collaborative  
mapping. An experiment. In *Proc. of 9th Int. Conf. on Innovation in Urban  
and Regional Planning (INPUT 2016)* (pp. 356–361). URL: [http://www.  
input2016.it/conference\\_2016/conference\\_proceedings](http://www.input2016.it/conference_2016/conference_proceedings).
- 1095 Voghera, A., La Riccia, L., Lucenteforte, M., & Savoca, A. (2018). Rappre-  
sentare il territorio attraverso il collaborative mapping. il progetto OnToMap.  
In *Smartness e healthiness per la transizione verso la resilienza. Orizzonti di  
ricerca interdisciplinare sulla città e il territorio*. (pp. 239–252). Milano, Italy:  
Franco Angeli volume 86.1 of *BETHA - Built Environment Technologies and*  
1100 *Healthy Architectures*.
- W3C (2017a). SPARQL query language for RDF. [https://www.w3.org/TR/  
rdf-sparql-query/](https://www.w3.org/TR/rdf-sparql-query/).
- W3C (2017b). Web ontology language (OWL). <https://www.w3.org/OWL/>.
- W3C (2018). Linked Data. [https://www.w3.org/standards/semanticweb/  
1105 data](https://www.w3.org/standards/semanticweb/data).
- White, R., & Roth, R. (2006). Supporting exploratory search - introduction.  
*Communications of the ACM*, 49, 36–39. URL: [https://doi.org/10.1145/  
1121949.1121978](https://doi.org/10.1145/1121949.1121978). doi:10.1145/1121949.1121978.
- Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata  
1110 for image search and browsing. In *Proceedings of the SIGCHI Conference on  
Human Factors in Computing Systems CHI '03* (pp. 401–408). New York, NY,  
USA: ACM. URL: <http://doi.acm.org/10.1145/642611.642681>. doi:10.  
1145/642611.642681.

## Appendix

Table 8: Entropy, mean cardinality and exploration cost of the facets displayed in Figure 5.

|                 | Entropy | Mean Cardinality | Exploration cost |
|-----------------|---------|------------------|------------------|
| Outdoor Seating | 0.9416  | 46.0000          | 0.0205           |
| Takeaway        | 0.8482  | 25.3333          | 0.0335           |
| Ex1             | 3.0000  | 20.0000          | 0.1500           |
| Cuisine         | 4.3895  | 5.0233           | 0.8738           |
| Ex2             | 3.0000  | 3.0000           | 1.0000           |
| Name            | 9.3987  | 1.0000           | 9.3987           |

Table 9: Balance, object cardinality, frequency and navigation quality of the facets shown in Figure 5, according to (Oren et al., 2006) with  $\mu = 2$  and  $\sigma = 4.9$ . We remind that the colors of facets in Figure 5 correspond to the complement of the values reported in the present table.

|                 | Balance | Object cardinality | Frequency | Navigation quality |
|-----------------|---------|--------------------|-----------|--------------------|
| Outdoor Seating | 0.8587  | 0.9794             | 0.1280    | 0.1076             |
| Takeaway        | 0.5175  | 0.9201             | 0.1057    | 0.0503             |
| Ex1             | 1.0000  | 0.4725             | 0.2225    | 0.1051             |
| Cuisine         | 0.3663  | 4.54E-66           | 0.6008    | 9.99E-67           |
| Ex2             | 1.0000  | 0.4725             | 0.0334    | 0.0157             |
| Name            | 1.0000  | 0.0000             | 0.9388    | 0.0000             |