Visualization of Multi-domain Ranked Data

Alessandro Bozzon[1], Marco Brambilla[1], Tiziana Catarci[2],
Stefano Ceri[1], Piero Fraternali[1], and Maristella Matera[1]

[1] Dipartimento di Elettronica e Informazione – Politecnico di Milano
{bozzon,mbrambil,ceri,fraterna,matera}@elet.polimi.it
[2] Dipartimento di Informatica e Sistemistica – Università "La Sapienza", Roma
catarci@dis.uniroma1.it

Abstract. This chapter focuses on the visualization of multi-domain search results. We start by positioning the problem in the recent line of evolution of search engine interfaces, which more and more are capable of mining semantic concepts and associations from text data and presenting them in sophisticated ways that depend on the type of the extracted data. The approach to visualization proposed in search computing extends current practices in several ways: the data to visualize are N-dimensional combinations of objects, with ranking criteria associated both to individual objects and to sets of combinations; object's properties can be classified in several types, for which optimized visualization families are preferred (e.g., timelines for temporal data, maps for geo-located information); combinations may exhibit any number of relevant properties to be displayed, which need to fit to the bi-dimensional presentation space, by emphasizing the most important attributes and de-emphasizing or hiding the less important ones. The visualization problem therefore amounts to deciding the best mapping between the data of the result set and the visualization space.

1 Introduction

Information visualization exploits presentation metaphors and interaction strategies for supporting perceptual inferences [1][15]. It is crucial for the success of any modern information-intensive application, particularly in multi-domain search, which produces articulated results comprising different types of objects extracted from multiple data sources and subject to various ranking criteria. The relevant issues in search computing interfaces are a mix of traditional search interface optimization concerns (e.g., how to ensure simplicity and high performance) and of new problems, such as how to exploit domain- and type-specific knowledge about data to achieve a "natural" ranking and display (e.g., using timelines, maps, or different kinds of charts), how to represent relationships among objects coming from different domains, how to convey the data provenance.

The design of a search interface must take into account the conservative nature of the visualization solutions so far adopted by search engines, which have scarcely evolved over the last decade and tend to exhibit a strong stability in the features offered to users, to preserve ease of use and efficacy for a wide audience of much

S. Ceri and M. Brambilla (Eds.): Search Computing II, LNCS 6585, pp. 53–69, 2011.

differentiated adopters. As the tasks supported by search engines increase in complexity, e.g., by providing concept and topic search [14][25], new features are introduced in the search interface of mainstream search engines cautiously, by progressively expanding the basic paradigm of keyword entry and vertical result list presentation. However, various factors still differentiate multi-domain search systems from traditional search engines.

- **Results are assembled from objects,** not necessarily corresponding to documents. For example, an object may derive from a deep Web record, for which there is no surface Web browsable page. This impacts the way in which feedback can be given to the user, because not always a document summary is the best way to provide provenance clues. For instance, some objects may be best presented showing a subset of their attributes and a link to the original information source, if this is available.
- **Results are not individual documents but combinations of objects.** This challenges the flat vertical list presentation approach, because not only provenance and result ranking must be conveyed, but also the relationships that connect objects in a result combination. Feedback on these relationships should be added to the search interface, e.g., to show why a hotel and a conference venue belong to the same combination.
- **Results are typed.** Unlike documents, which are unstructured, multi domain search retrieves typed objects. Knowing the object type is both a challenge and an opportunity: it is a challenge because exploiting the object type for differentiating the representations can make the interface complex and unstable, breaking the interaction style continuity that search engine users appreciate so much. It is also an opportunity, because type information can suggest optimized ways to present the results. For instance, if the interface recognizes geographical or temporal data it can offer the user the option to display results in a map or timeline.
- **Relationships have semantics.** objects forming a combination may be associated for different reasons: they may be close in space or time; they may share attribute values; they may be related in a hierarchy; they may be linked by a domain-specific relationship. In those cases when the interface is aware of the meaning of a relationship, it should represent it in an intuitive way.
- **The result space can be highly dimensional.** answers to multi-dimension queries may require the display of a high number of data attributes and associations; the interface should allow a compact visualization of the “most relevant” aspects of the result set and the easy appraisal of the features that have been hidden in the initial presentation. Given that the page layout is bi-dimensional, suitable summarization mechanisms are necessary to convey more than two data dimensions.

The initial design of data visualizations proposed for search computing were reported in [3] and focused on exploring the primitives for the manipulation of the result set and the refinement of the initial query, based on a very simple tabular representation of combinations. In this chapter, we instead concentrate on result visualization, and specifically on how to convey the relationships that link the objects in a combination, how to exploit knowledge about the type of an object forming a combination, and how to convey the ranking at both the individual data source and at the object combination level. We will in particular introduce one possible approach that capitalizes on well-known principles for data visualizations, trying to maximize their effectiveness for the specific characteristics of the search computing result sets.

This Chapter is organized as follows. Section 2 surveys a gallery of visualization examples from mainstream search engines and discusses the contributions from data visualization literature that can support the envisioned result presentation approach. Section 3 introduces the Search computing visualization problem, which consists of three parts: the data model of the result set, the existence of "expressive" types and dependencies within such model, and the heuristic approach for mapping the result set onto a visualization layout. Section 4 shows, without attempting an exhaustive approach, some of the visualizations that could be generated by applying heuristics. Section 5 presents the conclusions pointing to our future work.

2 Related Work in Search Engine Interfaces and Visualization Techniques

In this section we overview current search engine interfaces and the visualization paradigms for specific data (including categorical, geographical and temporal data).

2.1 Evolution of Search Engine Result Presentation

In recent years, search results visualization has evolved to include multiple features, which depend on the specific object being visualized, enriching the traditional result sets made of ordered list of objects (document surrogates), as the one shown in Fig. 1.

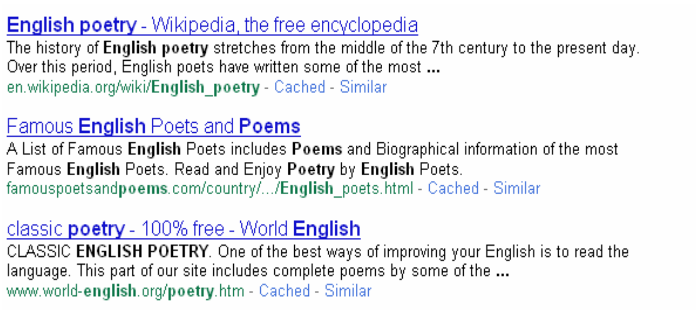


Fig. 1. Basic search engine result set as an ordered list of documents

The result set can be structured in sub-lists, e.g., to highlight entries of different pages within the same Web site, or result items of different nature (e.g., images related to a given Web documents), as in Fig. 2. The information content of result elements can be enriched if the type of the displayed information is known. For example, Fig. 3 shows the use of a timeline for the visualization of temporal data.

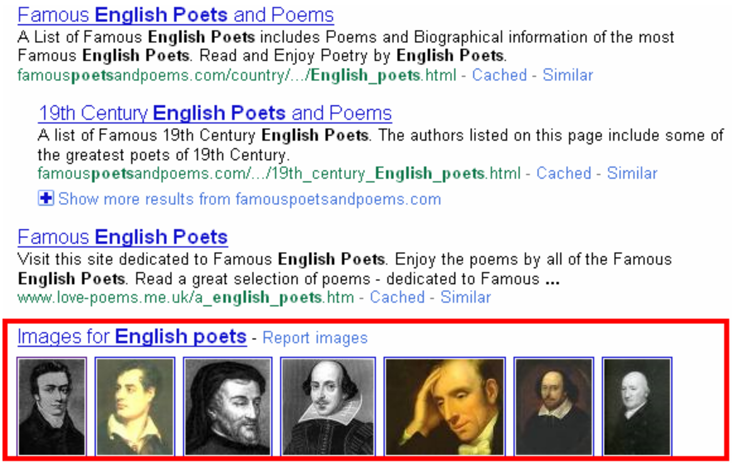


Fig. 2. A result set structured into a list of documents and a list of images

Fig. 3. Timeline visualization for a result set comprising temporally located documents

A special class of typed information is constituted by geo-referential data, which permits the extraction of geographical concepts; a geographical concept is an object associated with geographical coordinates and possibly categorized according to a geographic ontology (e.g., GeoNames ontology classes [13]). Geographical concepts are highlighted in the result set and typically positioned on a map. The identified concept can also be associated with semantically correlated concepts; for example, a town can be related to hotels, restaurants, and places of interest; a touristic place, a hotel or a restaurant can be associated with reviews.

The amount of structure extracted from documental data and the employed visualization primitives depends on the capacity of recognizing domain specific entities. The examples shown so far revolve around the presentation of instances of a single entity. New generation search engines are moving towards to the collection and integration of heterogeneous data sources. Kosmix [25] is an example of general-purpose topic discovery engine, which offers one-page information summaries about a topic. The presented information is retrieved through calls to Web services that extract information from deep Web data sources. The schema of each topic is a complex record type, which not only comprises typed properties of the entity but also associations to other entities representing related topics. For instance, the query “William Shakespeare” returns an instance of the entity *Writer*, as shown in Fig. 4, characterized by information deriving from multiple services (e.g., Biography.com, and Google) and also related to other related writer instances.

Fig. 4. An instance of the entity *Writer* in the Kosmix topic search engine

2.2 Related Work in Data Visualization

Data visualization has a long-standing tradition, which initially focused on the analysis of alternative visualization techniques for various categories of data [1][4]. Classic works like [27][23] offered guidance in selecting the most appropriate visualization techniques for different types of data (e.g., 1-, 2-, 3-dimensional data, temporal and multi-dimensional data, and tree and network data in [27]). Later works (e.g., [6]) explored the underlying conceptual structure of data-oriented visualization, highlighting a common framework of data visualization strategies (e.g., the data stages/transformation model in [6] and the classification based on data spatialization and visual perception in [26]), giving a deeper rationale to the taxonomies of visualization techniques. For example, in the field of relational data visualization, the work described in [29] proposes three families of 2D graphics that depend on the type of data to be represented on the axes, i.e., o*rdinal–ordinal*, when no dependency exists among the different attributes (e.g., a table), *ordinal–quantitative*, where a quantitative variable is dependent on the ordinal variable (e.g., a bar chart), or the ordinal and quantitative data can be independent (e.g., a Gantt chart), and *quantitative-quantitative*, used to represent the distribution of data as a function of one or both quantitative variables, also highlighting causal relationships between the two quantitative variables (e.g., a map). Each family then contains variants depending on the combination of the selected mark type (e.g., rectangle, circle, glyph, text, Gantt bar, line, polygon, and image) and its visual and retinal properties that best suit the characteristic properties of the single objects.

In [31] the author proposes a comprehensive taxonomy of glyph placement strategies with respect to both data types and user's task. The author in particular distinguishes between *data-driven* and *structure-driven* approaches: the former exploit data properties to determine the object location in the visualization space; the latter exploit some (implicit or explicit) relationships between data points, such as a temporal ordering or hierarchical relationships.

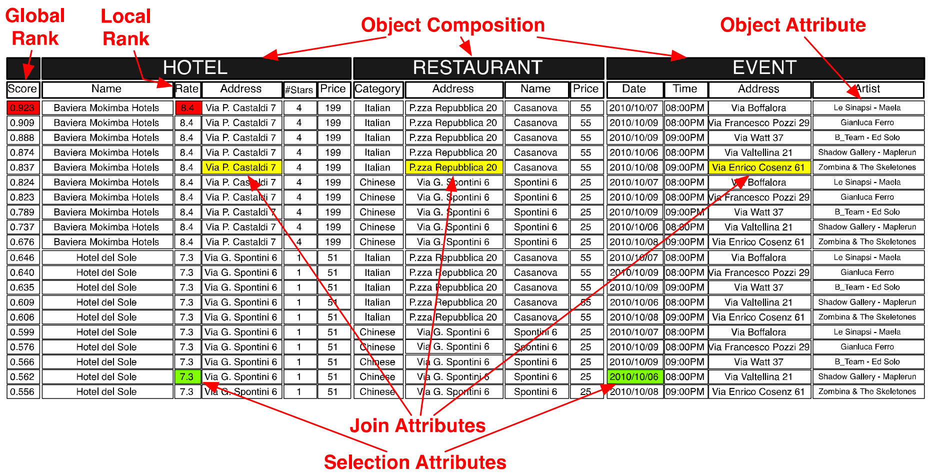
Whereas the abovementioned efforts are horizontal, spanning all possible categories of data, other works concentrate on the design, evaluation and comparison of visualization techniques for specific data: categorical (e.g., [19][11]), temporal (e.g., [7][9]), geographical (e.g., [12][20][30]), multidimensional (e.g., [31]), and graph-based data (e.g., [32]) are the most prominent sectors.

A huge amount of literature also concentrates on the presentation generation process. The pioneer approach proposed by Mackinlay [22], and several other successive works (e.g., [15]), exploit data characterization and propose rule-based approaches to map data types to visual elements [5]. In other words, elements of a *data model* are mapped to elements of a *visual model* [15]. The common aim characterizing such works is to automatically derive "adequate" visualizations [5], where adequate means *complete,* i.e., the user perceive from them all the information enclosed within the original data, and *correct,* i.e., no unrequired information is conveyed. By capitalizing on the principles introduced by the above-illustrated works, in the following sections we illustrate how the characterization of the Search computing result set can guide the definition of a visualization process.

3 Visualization Model and Process for Search Computing

3.1 Visualization Model

The result of a Search computing query is made of combinations, where each combination is a record of *N* object instances connected by join predicates and satisfying selection conditions. Each object in a combination has a schema constituted by typed attributes; a subset of the attributes defines the combination *identifier*, which is either system-generated (object identifier) or provided by the real world context (e.g., event's date and time). Objects may have *repeating groups*, where each repeating group is a non-empty set of attributes with multiple values (such as actors of a given movie, with name and gender). Object instances extracted by a query may be locally *ranked*; in such a case one or more object's attributes express the ranking, which is either system-generated (rank position among the object instances) or provided by the usage context (e.g., the rating of a hotel). Combinations have a global rank, expressed as a weighted sum of the local ranks of member objects, normalized in the [0-1] range.



	HOTEL					RESTAURANT				EVENT			
Score	Name	Rate	Address	#Stars	Price	Category	Address	Name	Price	Date	Time	Address	Artist
0.923	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Italian	P.zza Repubblica 20	Casanova	55	2010/10/07	08:00PM	Via Boffalora	Le Sinapsi - Maela
0.909	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Italian	P.zza Repubblica 20	Casanova	55	2010/10/09	08:00PM	Via Francesco Pozzi 29	Gianluca Ferro
0.888	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Italian	P.zza Repubblica 20	Casanova	55	2010/10/09	09:00PM	Via Watt 37	B_Team - Ed Solo
0.874	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Italian	P.zza Repubblica 20	Casanova	55	2010/10/06	08:00PM	Via Valtellina 21	Shadow Gallery - Maplerun
0.837	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Italian	P.zza Repubblica 20	Casanova	55	2010/10/08	09:00PM	Via Enrico Cosenz 61	Zombina & The Skeletones
0.824	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/07	08:00PM	Via Boffalora	Le Sinapsi - Maela
0.823	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/09	08:00PM	Via Francesco Pozzi 29	Gianluca Ferro
0.789	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/09	09:00PM	Via Watt 37	B_Team - Ed Solo
0.737	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/06	08:00PM	Via Valtellina 21	Shadow Gallery - Maplerun
0.676	Baviera Mokimba Hotels	8.4	Via P. Castaldi 7	4	199	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/08	09:00PM	Via Enrico Cosenz 61	Zombina & The Skeletones
0.646	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Italian	P.zza Repubblica 20	Casanova	55	2010/10/07	08:00PM	Via Boffalora	Le Sinapsi - Maela
0.640	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Italian	P.zza Repubblica 20	Casanova	55	2010/10/09	08:00PM	Via Francesco Pozzi 29	Gianluca Ferro
0.635	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Italian	P.zza Repubblica 20	Casanova	55	2010/10/09	09:00PM	Via Watt 37	B_Team - Ed Solo
0.609	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Italian	P.zza Repubblica 20	Casanova	55	2010/10/06	08:00PM	Via Valtellina 21	Shadow Gallery - Maplerun
0.606	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Italian	P.zza Repubblica 20	Casanova	55	2010/10/08	09:00PM	Via Enrico Cosenz 61	Zombina & The Skeletones
0.599	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/07	08:00PM	Via Boffalora	Le Sinapsi - Maela
0.576	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/09	08:00PM	Via Francesco Pozzi 29	Gianluca Ferro
0.566	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/09	09:00PM	Via Watt 37	B_Team - Ed Solo
0.562	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/06	08:00PM	Via Valtellina 21	Shadow Gallery - Maplerun
0.556	Hotel del Sole	7.3	Via G. Spontini 6	1	51	Chinese	Via G. Spontini 6	Spontini 6	25	2010/10/08	09:00PM	Via Enrico Cosenz 61	Zombina & The Skeletones

Fig. 5. Tabular representation of a multi-domain result set

Fig. 5 describes the main elements of the results of the query "*find the best combination of hotels, restaurants and events to spend a nice evening and night in given city* ", which composes of three objects: *events*, *restaurants*, and *hotels*. Fig. 5 can be considered the default way of visualizing a multi-domain result set: the table-based visualization lists all combinations, with their objects and attributes, in descending global rank. However, alternative visualizations are possible, which might highlight few primary visualization dimensions selected within a combination. Such result presentations can take advantage of suitable visualizations, based upon the type of some of the attributes of the result, e.g., maps for geographic locations, timelines for temporal events, or Cartesian spaces for quantitative variables. With this respect, the first step is to identify relevant data types that can guide the selection of suitable primary visualization dimensions.

3.2 Data Type Classification

Visualization of attribute values can be optimized according to their type. As usual, attribute types are classified according to their scale type, as stated in the classic definition of measurement theory [25]:

- **Interval:** quantitative attributes measured relative to an arbitrary interval (e.g., Celsius degrees, latitude and longitude, date, GPA). In this class, two important subclasses are further distinguished for visualization purposes:
  - o **Geographic points and addresses:** they admit domain specific operations like the computation of distance, the visualization on maps, the determination of routes, etc.
  - o **Time points:** they admit the representation on time scales and calendars, at different domain-specific granularity.
- **Ratio:** quantitative attributes measured as the ratio with a known magnitude unit (e.g., most physical properties).

- **Nominal:** categorical labels without notion of ordering (e.g., music genre). They can be visualized by means of textual labels (for example within tables). In case of a low number of categories, they can be represented through visual clues, for example different shapes or colors.
- **Ordinal:** data values that admit order, but not size comparison (e.g., quality levels).

Nominal data can be also associated with frequency (or relative frequency) values, i.e., the number (percentage) of elements falling into categories. This may imply the use of a combination of visual clues, such as the size of shapes or the opacity of the colors representing categories, or the "construction" of graphics (e.g., bar charts), where categories are labels over the horizontal axis, while frequencies correspond to the vertical axis dimension. Nominal data can also be further distinguished into **taxonomical**, when they admit subset relations (e.g., animal species). Tree based representations (e.g., treeMap) can be adopted in this case [16].

3.3 Visualization Process

The process of generating the visualization aims at producing a representation that maximizes the understandability of the result set, by considering the type and semantics of data and the functions of multi-domain search. The visualization process, schematically illustrated in Fig. 6, maps the result set onto a presentation space by

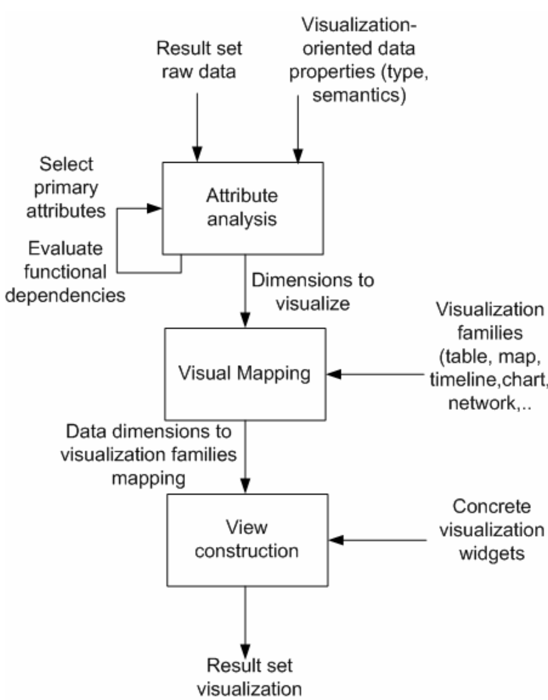


Fig. 6. Main steps in the result set visualization process

considering the types of attributes that describe the properties of combinations, the presentations functions to be implemented, and the visualization families that are best usable for rendering the data dimensions.

The most important visualization dimensions, which need to be highlighted in the visualization and that can guide the identification of the primary visualization space, are determined; to this end, information about object *identifiers*, *join* and *selection conditions*, and dependencies among the different data are exploited.

The output of the visualization process is a page layout (a view, in the terminology of [6]), which assigns data to visualization elements, implementing all the needed visualization functions. The generation process in Fig. 6 can occur at design time (when the query is fixed and non modifiable), at run-time (when the user explores the search space from scratch) and in a mixed mode (when the user starts with a known query and then expands it, by joining in more search services).

In order to identify the primary visualization dimension, one can assume that an ordering of data type exists and that it guides the selection of the primary attributes determining the visualization space. Such ordering depends on the capacity of the data types to “delimit” a visualization space where single objects and combinations included in the result set can be conveniently positioned. For example, interval attributes can be considered the best candidates for primary visualization, since they permit a precise characterization of the position of objects within a bi-dimensional visualization space, followed by ordinal, and nominal attributes, which instead can be adequately represented by means of visual clues. Therefore, if a given object O has a geo-referenced attribute, then its instances can be represented on a map by using that attribute as a primary visualization; similarly, objects with attributes representing temporal events can be placed on a timeline. Representing a combination then amounts to finding suitable representations for the majority of its objects, by highlighting them upon a given visualization space, and then relating together the object instances of a combination through orthogonal visual mechanisms. Once placed on the visualization space, an object is succinctly represented by some of its attributes (e.g., identifiers); typically the local ranking of the object can also be visually represented (e.g., through conventional shapes, or colors). Other attributes are omitted from the visualization, and can be accessed through secondary visualization methods, such as pop-up windows.

In line with the classical approaches for the automatic generation of visualizations [5], this construction can be formalized by heuristic rules, which take in input the characterization of the N dimensions to visualize and emit as output the decision of how to allocate them onto a bi-dimensional representation space, addressing the visualization functions of multi-domain search. A very general scheme is the following:

1. Pick the most relevant dimensions, on the basis of a defined order among the available data types and on the identification of the dimensions that best characterize the majority of objects involved in combinations.
2. Identify the primary visualization space to use;
3. Allocate all dimensions that can be supported by the primary visualization space;
4. Pick the remaining dimensions in order of significance and decide the secondary widgets best suited to represent them;
5. Allocate all remaining dimensions that can be supported ergonomically by secondary widgets.

At step 1, selected data types allow a domain-dependent effective visualization. The order: *geo location→time/date→other types* can be used to exploit geographical information first, then temporal data, and then all the remaining uncharacterized types. At step 2, the most appropriate visualization space is chosen. Guidelines from the literature and best practices in data visualization (e.g., [9] for temporal data, [11] for categorical data, [20] for geographical data) can be used. Also, one can exploit the huge quantity of out-of-shelf components that are more and more offered by visualization projects and providers of software components[1]. At step 4, the dimensions that are not represented by the primary widget need to be considered, and visual clues for their representation selected. Various techniques exist that descend from the classic identification of visual variables by Bertin (position, size, shape, value, color, orientation, and texture), possibly adapted to specific contexts (e.g., thematic cartography [12]).

The primary dimensions for data visualization can be further reduced by exploiting dependencies between objects. For example, if a query includes a 1:M join between two objects O1 and O2, such that a set of instances of O2 is mapped exactly to one instance of O1, then the objects of O2 can be represented by using the primary visualization chosen for O1. For instance, if O1 are hospitals and O2 are doctors, and hospitals have a geo-localized attribute (their address), then it is possible to display doctors on a map by placing them at the same attribute as their hospital.

4 Examples of Visual Representations of Query Results

To highlight the importance of dimension selection, this section presents several examples of visualizations that represent objects, their composition, and their local and global rankings; every example can be generated by suitable applications of generic visualization rules.

4.1 Visualization of Geo-referenced Objects

Fig. 7 presents a visualization example for the query "*find the best combination of hotels, restaurants and events to spend a nice evening and night in given city*". One possible result set is the one illustrated in Fig. 7. The three objects to be displayed have geographic coordinates; therefore, the primary dimension to adopt for their representation is a map. This choice allows us positioning each object instance as a point in the map, which will be conveniently selected so as to include objects in the context (e.g., a portion of a city map). A combination is then a triple of positions on the map; it can be visualized by any representation that puts the three objects together; in the example, we enclose each triple within an area, which is highlighted by means of colors. We then use darkest color for the best combination.

---

[1] Among the best known examples, the OLIVE library [24] lists data visualization environments clustered according to Shneiderman's classification scheme and the recent effort by IBM [17] offers a community space for publishing visualization widgets and data sets (at the time of writing, 70183 visualizations and 141459 data sets are enlisted).

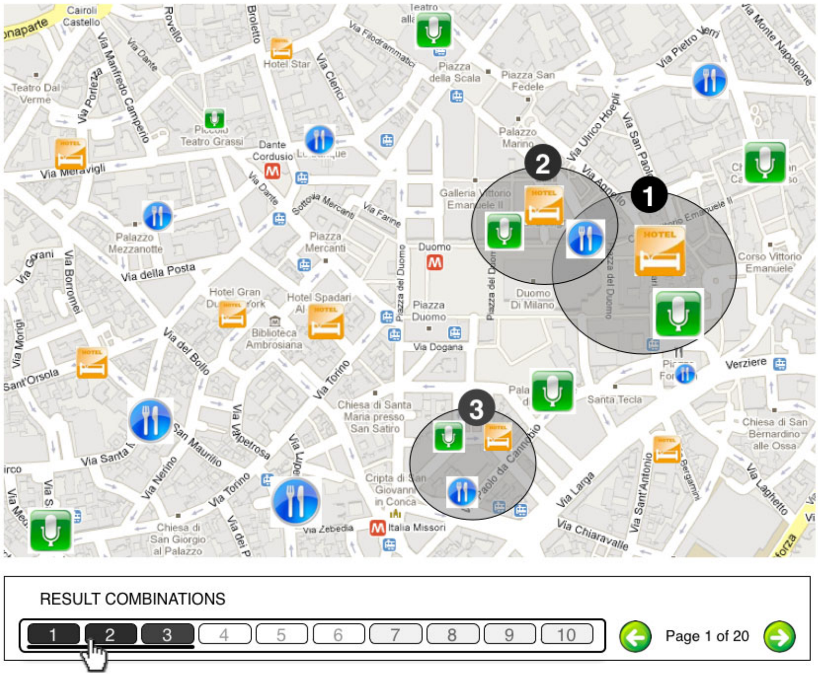


Fig. 7. Visualization of geo-referenced objects

Moreover, each object has a different icon, and the local ranking (representing the relative ranking of the object instance within the selected objects) is represented by the size of the icon. Such representation can show only a few combinations on the same screen, therefore the visual result presents only three combinations (the top-ranked) and a scrolling mechanism allows seeing the following combinations. The other object attributes (e.g., name, stars and price for the hotels) can be displayed in pop up windows, opened by pointing to given object with the mouse.

4.2 Geo-referenced Visualization with Object Dependencies

Object dependencies can be used to associate a visualization dimension to objects that do not have properties with data types that effectively determine a visualization space. Consider a query that searches for *close hospitals to a given location, such that the hospitals have specialists of a given disease (e.g., Parkinson)*; the specialists are ranked, e.g., by the relevance of their published articles on the topic. Doctors have no properties that allow their representation according to a quantitative visualization space, but they are placed at hospitals, which have a specific location. Therefore, doctors can be represented as icons, which are placed on a map at the same location as their hospitals. Their local ranking can be represented through different icon sizes.

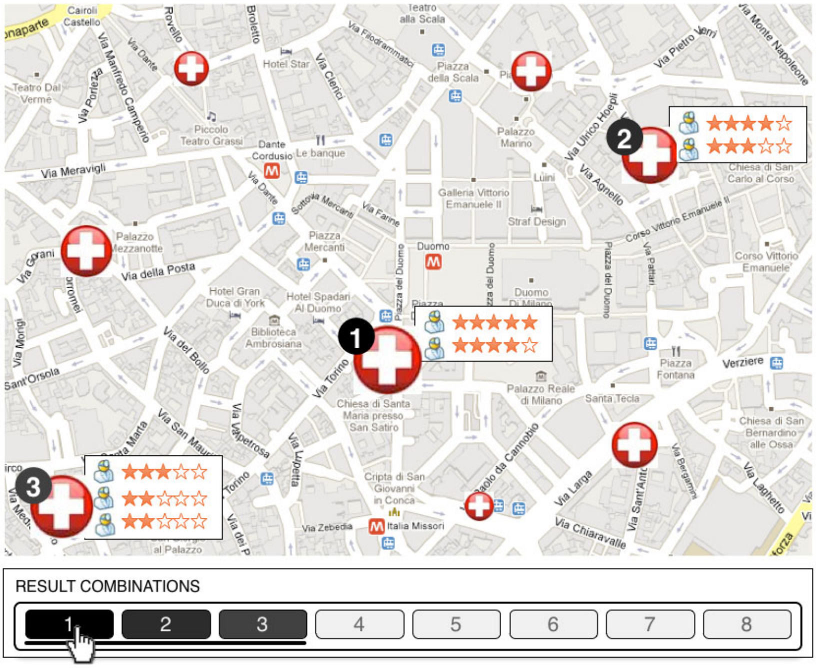


Fig. 8. Geo-visualization with object dependencies

Fig. 8 highlights the dependency of doctors from hospitals; hospitals play the role of aggregators. Being hospitals geo-referenced objects, than the map can be exploited even though addresses do not characterize doctors. The map shows that *hospital 1* has two doctors with high rating; *hospital 2*, instead, has two doctors but with a lower rating, and hospital 3 has three doctors with a lower rating. The hospital index is also a representation of the global ranking, which is in this case globally associated with a hospital, by aggregating the ratings of doctors, and by considering the user's location. Once the different combinations are displayed on the map, then pop-up windows can show more attributes about hospitals and doctors.

4.3 Timeline Visualizations

While geographic maps are very effective for relating objects that are located in space, timelines are effective for relating objects that are located in time. Fig. 9 provides a visualization example for a search about author's productivity, both in terms of publication indexes (e.g., the Hirsch Index) and of yearly production. Authors are locally ranked by their publication index; the yearly productivity of an author is measured by the number of published articles (divided in journal and conferences) and books; the local ranking is a weighted sum of such measures. The global ranking, which takes into account the publication indexes, the yearly production, and how recent is the production (recent production is most highly ranked), presents the "*most productive recent years of high-ranked authors*".

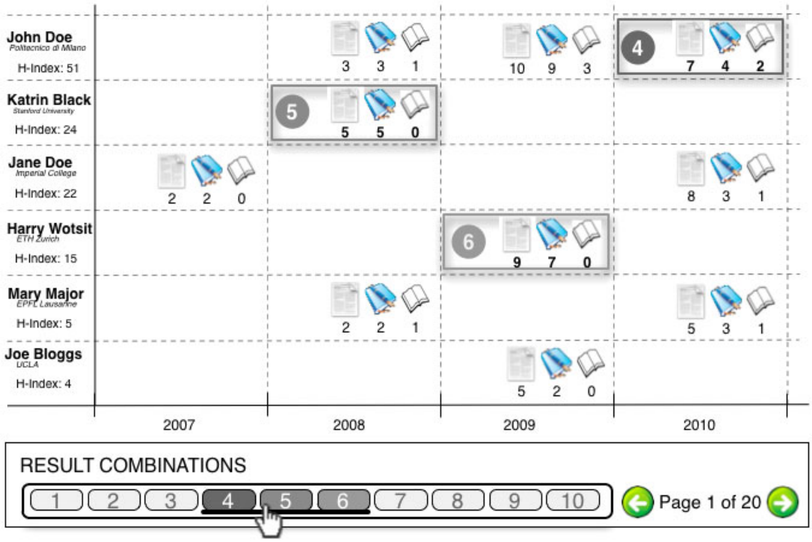


Fig. 9. Timeline representing scientists' productivity over the years

This information is presented on a timeline, having on the time axis the year of production, and presenting every author on the same line. Authors are listed in "order" of their Hirsch ranking, and the visualization highlights the most productive pair author-year, initially centered on that author and year; the author-year pair is highlighted (e.g., being encircled and colored). Author information, reported on the left of the timeline, is the name and H index; yearly production has three icons representing papers, articles and books, with the appropriate number below the icon, placed in the timeline. The three combinations with higher global index are highlighted, and a scroller allows moving to the subsequent combinations; scrolling the combinations has the effect of moving along the timeline and the authors. This representation is inspired by Envision [10].

The automatic generation of such representation uses the assumptions that yearly productivity is related to time intervals (therefore it can be represented by a timeline), and that each yearly productivity item functionally determines an author; therefore, each author can be associated with a single timeline. Then, author's local ranking is the H-index, which is represented, while yearly productivity's local ranking is a combination of the number of publications, which are also represented; and the global ranking is explicitly represented by highlighting. A similar representation could display the severity of hospitalizations of a family of chronically ill patients, with each patient represented as a line in a timeline, hospitalizations reported as intervals on the timeline, and severity of treatment reported as indicators close to the intervals. The presence, type and duration of given treatments could be also reported graphically.

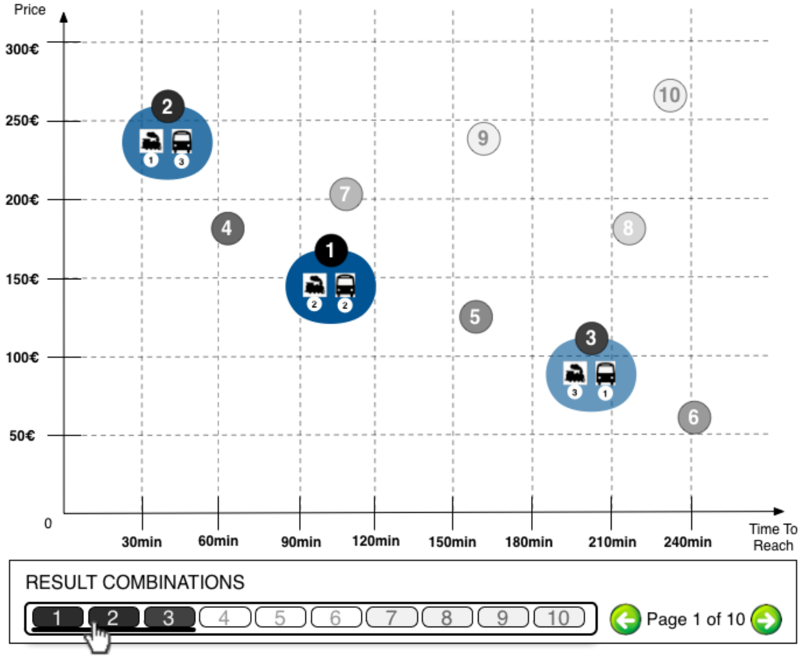


Fig. 10. Representation of combinations ranked by ratio attributes

4.4 Representation in the Lack of Suitable Interval Dimensions

While in the above examples the objects to be displayed include interval dimensions with an associated graphic representation, in the general case objects may lack such a property. Then, their visualization may resort to other object properties, especially when such properties are associated with the object's rankings.

Assume a query about reaching a given location by combining trains and local transportation (taxi or bus). Assume that each such transportation has a cost and duration, and assume that they can be ranked locally and globally by a suitable function of their local and global costs and duration. Note that price and duration both belong to "interval" attribute types according to the categorization of Section 3.2. Then, Price and Time can be used as axes of a Cartesian space.

In the representation of Fig. 10, we highlight the global time and global duration in the space, and then enclose the first, second, and third combinations according to the global ranking (and let other combinations be accessible e.g., through a scroller). The presence of two objects is represented by the fact that each location in the space is associated with two objects, and by clicking on the object one can read the details about the train and local trip, including its duration and cost.

Finally, assume the case when ranking is based on ordinal dimensions. Let's consider an example involving a vacation package where two ordinal dimensions describe the price-range of the package (e.g., in hundreds of Euro) and quality of the hotels being used (from five to three stars), described in Fig. 11.

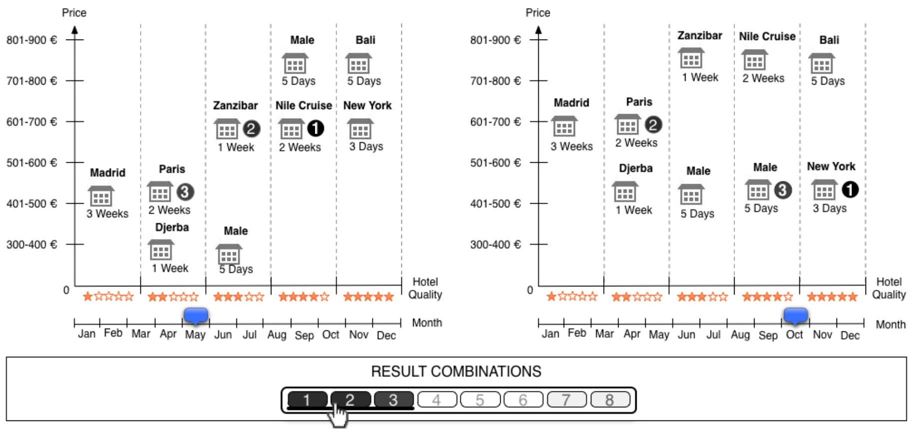


Fig. 11. Example of ordinal-quantitative visualization

Assume then that the global ranking function takes into account both these features and is based on a price-performance ratio. In such case, one of the dimensions (say, the quality of hotels) is used as reference in order to organize the space in five columns, and then the packages are presented as nodes in these columns, named by the principal attraction (e.g., "Nile Cruise"), in price-range order. The global ranking is presented by highlighting the three combinations that present the "best" price-performance according to the global ranking. Fig. 11 also shows the presence of two timelines, selected by the user in the query, and the effect of seasonal changes (e.g., of prices and hence on global ranking) on the choice of top combinations. The system may enable selecting trips (e.g., by country, availability during the year) so as to inspect the vacation packages of interest; once selected, the properties of the packages can be described further by inspecting pop-ups associated with each node.

5 Conclusions and Future Work

This paper has presented the general method that we intend to use in generating visualizations for multi-domain ranked data, and illustrated some representative examples that are currently driving us in building the method. Building a generic visualization tool, able to analyze the visualization data model and produce a suitable representation without being driven by other knowledge about the application domain, is a very challenging task. While the first examples that we have constructed ad-hoc seem promising and yielding to general rules of good applicability, the actual validation of the approach requires a formalization of the data and visualization models, aimed at identifying the most relevant properties that can guide the visualization process. Such formalization will be inspired to some past works that have already identified collections of rules to guide the visualization process for relational structured data, but it will take into account the peculiarity of the Search computing result set.

We will also try to overcome some limitations of past approaches, such as the lack of formal checking of the visualization correctness [4]. An extensive experimentation with users will also allow us to assess the effectiveness of the produced visualizations, and to investigate to which extent totally automatic processes for the visualization generation should be preferred to participatory paradigms where the user is directly involved, e.g., by means of preference expressions, to the construction of the visualization spaces.

References

[1] Bertin, J.: Semiology of Graphics: Diagrams, Networks, Maps (1983)

[2] Bostock, M., Heer, J.: Protovis: A Graphical Toolkit for Visualization. IEEE Trans. Vis. Comput. Graph. 15(6), 1121–1128 (2009)

[3] Bozzon, A., Brambilla, M., Ceri, S., Fraternali, P., Manolescu, I.: Liquid Queries and Liquid Results in Search Computing. In: Ceri, S., Brambilla, M. (eds.) Search Computing. LNCS, vol. 5950, pp. 244–267. Springer, Heidelberg (2010)

[4] Catarci, T., Dix, A.J., Kimani, S., Santucci, G.: User-Centered Data Management. Morgan & Claypool Publishers, San Francisco (2010)

[5] Catarci, T., Santucci, G., Costabile, M.F., Cruz, I.: Foundations of the DARE System for Drawing Adequate Representations. In: Proc. of DANTE 1999 (1999)

[6] Chi, E.H.: A Taxonomy of Visualization Techniques Using the Data State Reference Model. In: IEEE Symp. Information Visualization, Salt Lake City, UT, USA, pp. 69–75. IEEE CS Press, Los Alamitos (2000)

[7] Daassi, C., Nigay, L., Fauvet, M.C.: A Taxonomy of Temporal Data Visualization Techniques. Information-Interaction-Intelligence 5(2), 41–63 (2005)

[8] Elmqvist, N., Stasko, J., Tsigas, P., Meadow, D.: A Visual Canvas for Analysis of Large-Scale Multivariate Data. Information Visualization 7(1), 18–33 (2008)

[9] Fernandes Silva, S., Catarci, T.: Visualization of Linear Time-Oriented Data: A Survey. Journal of Applied System Studies 3(2) (2002)

[10] Fox, E.A., Hix, D., Nowell, L.T., Brueni, D.J., Wake, W.C., Heath, L.S., Rao, D.: Users, User Interfaces, and Objects: Envision, a Digital Library. Journal of the American Society for Information Science 44(8), 480–491 (1993)

[11] Friendly, M.: Visualizing Categorical Data. SAS Publishing (2000)

[12] Garlandini, S., Fabrikant, S.I.: Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization. In: Hornsby, K.S., Claramunt, C., Denis, M., Ligozat, G. (eds.) COSIT 2009. LNCS, vol. 5756, pp. 195–211. Springer, Heidelberg (2009)

[13] GeoNames Ontology, `http://www.geonames.org/ontology/`

[14] Giunchiglia, F., Kharkevich, U., Zaihrayeu, I.: Concept Search. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds.) ESWC 2009. LNCS, vol. 5554, pp. 429–444. Springer, Heidelberg (2009)

[15] Haber, E.M., Ioannidis, Y.E., Livny, M.: Foundations of Visual Metaphors for Schema Display. J. Intell. Inf. Syst. 3(3/4), 263–298 (1994)

[16] Hearst, M.: Search User Interfaces. Cambridge University Press, Cambridge (2009), ISBN 9780521113793

[17] IBM. Manyeyes (December 2010), `http://www-958.ibm.com/software/data/cognos/manyeyes/`

[18] Johansson, S.: Visual Exploration of Categorical and Mixed Data Sets. In: Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration, Paris, France, July 28-28, pp. 21–29 (2009)

[19] Johnson, B., Shneiderman, B.: Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In: Proceedings of the IEEE Information Visualization 1991, pp. 275–282. IEEE, Los Alamitos (1991)

[20] Maceachren, A.M.: An Evolving Cognitive-Semiotic Approach to Geographic Visualization and Knowledge Construction. Cartography and Geographic Information Systems 19, 197–200 (2001)

[21] Mackinlay, J.D.: Automatic Design of Graphical Presentations, Ph.D. Thesis, Department of Computer Science, Stanford University (1986)

[22] Mackinlay, J.D., Hanrahan, P., Stolte, C.: Show Me: Automatic Presentation for Visual Analysis. IEEE Trans. Vis. Comput. Graph 13(6), 1137–1144 (2007)

[23] North, C.: A Taxonomy of Information Visualization User-Interfaces (1998), `http://www.cs.umd.edu/~north/infoviz.html`

[24] OLIVE: On-line Library of Information Visualization Environments, University of Maryland College Park (1999), `http://otal.umd.edu/Olive/`

[25] Rajaraman, A.: Kosmix: High Performance Topic Exploration Using the Deep Web. In: Proceedings of the VLDB Endowment, August 2008, vol. 2(1), pp. 1524–1529 (2009)

[26] Rodrigues Jr., J.F., Traina, A.J.M., de Oliveira, M.C.F., Traina Jr., C.: Reviewing Data Visualization: An Analytical Taxonomical Study. In: Proceedings of the Information Visualization, IV 2006 (2006)

[27] Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: IEEE Symp. Visual Languages, pp. 336–343. IEEE CS Press, Los Alamitos (1996)

[28] Stevens, S.S.: On the Theory of Scales of Measurement. Science 103(2684), 677–680 (1946), `http://web.duke.edu/philosophy/bio/Papers/Stevens_Measurement.pdf`

[29] Stolte, C., Tang, D., Hanrahan, P.: Polaris: A System for Query, Analysis, and Visualization of Multidimensional Databases. Commun. ACM 51(11), 75–84 (2008)

[30] Takatsuka, M., Gahegan, M.: GeoVISTA Studio: A Codeless Visual Programming Environment for Geoscientific Data Analysis and Visualization. Computational Geoscience 28, 1131–1144 (2002)

[31] Ward, M.O.: A Taxonomy of Glyph Placement Strategies for Multidimensional Data Visualization. Journal of Information Visualization 1(3/4)

[32] Wattenberg, M.: Visual Exploration of Multivariate Graphs. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI 2006, pp. 811–819. ACM, New York (2006)