Journal of Visual Languages and Computing 25 (2014) 995–1002

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

Short Paper

Uncertainty visualization of multi-providers cartographic integration ☆

Bilal Berjawi [a,*], Fabien Duchateau [b], Élisabeth Chesneau [c], Franck Favetta [d], Geoffrey Seccia [e], Claire Cunty [f], Maryvonne Miquel [a], Robert Laurini [a]

[a] LIRIS, UMR 5205, INSA de Lyon, Lyon, France
[b] LIRIS, UMR 5205, UCBL, Lyon, France
[c] EVS, UMR 5600, UJM, Saint Etienne, France
[d] LIRIS, UMR 5205, ENSNP, Lyon, France
[e] EMC, EA 3082, Université Lumière Lyon 2, Lyon, France
[f] EVS, UMR 5600, Université Lumière Lyon 2, Lyon, France

ARTICLE INFO

Article history:
Received 22 September 2014
Received in revised form
29 October 2014
Accepted 31 October 2014
Available online 11 November 2014

Keywords:
Cartography
GIS
Visual uncertainty
Spatial integration

ABSTRACT

Multiple cartographic providers propose services displaying points of interests (POI) on maps. However, the provided POIs are often incomplete and contradictory from one provider to another. Previous works proposed solutions for detecting correspondences between spatial entities that refer to the same geographic object. Although one can visualize the result of the integration of corresponding entities, users do not have any information about the quality of this integration. In this paper, we propose a solution to visualize the uncertainty inherent to a spatial integration algorithm. We present an integration process that identifies three levels of confidence for spatial and terminological integration results. Based on perceptual tests, we select visual variables to portray these three levels of confidence and we choose a visualization strategy. A prototype has been implemented to present the benefits of our proposal in a use-case scenario. This work has been realized within the framework of UNIMAP[1] project.

1. Introduction

Location-based services (LBS) are daily used in various applications, and cartographic providers play an essential role in displaying points of interest (POI) such as restaurants, hotels, and tourist places. A POI can be defined as a geographic object that has a point geometric shape. A POI has spatial attributes longitude and latitude, and terminological (non-spatial) attributes such as name and type (e.g., restaurant, hotel). Some providers may supply additional terminological attributes such as address, phone number, Website, customers' ratings, etc. A provider usually represents a POI on a map with a specific symbol or icon. Due to lack of completeness, noisy, inaccurate and contradictory data, it is interesting to propose solutions for detecting corresponding entities (i.e., which refer to the same POI) from different providers. This challenge aims at improving the quality and the relevance of information, which has a significant impact in tourist applications.

The integration of spatial information issued from different sources has been studied [9]. Earlier works so called “map conflation” were specifically devoted to vector objects such as roads [22]. In the last decade, the integration problem mainly

☆ This paper has been recommended for acceptance by Shi Kho Chang.

* Corresponding author.
E-mail addresses: bilal.berjawi@liris.cnrs.fr (B. Berjawi), fabien.duchateau@liris.cnrs.fr (F. Duchateau), elisabeth.chesneau@univ-st-etienne.fr (É. Chesneau), franck.favetta@liris.cnrs.fr (F. Favetta), geoffrey.seccia@gmail.com (G. Seccia), claire.cunty@univ-lyon2.fr (C. Cunty), Maryvonne.Miquel@liris.cnrs.fr (M. Miquel), Robert.Laurini@liris.cnrs.fr (R. Laurini).

[1] UNIMAP: http://liris.cnrs.fr/unimap (July 2014).

http://dx.doi.org/10.1016/j.jvlc.2014.10.033
1045-926X/

refers to the "entity matching" research domain, enhanced by a spatial aspect. The discovery of corresponding entities is performed either by exploiting only spatial information [25] or by computing and combining terminological similarities for selected attributes (e.g., name, type) [21]. Machine learning algorithms may be applied for tuning the parameters (e.g., weights) of a matching process [27]. When corresponding entities have been detected, an interesting use case aims at displaying a merged entity, i.e., to use a crafted algorithm to fusion the attributes' values of these corresponding entities. Such merging algorithms are not 100% confident. For instance, two corresponding entities may have a different location and the algorithm needs to determine the correct position. Similarly, the names or the phone numbers of two corresponding entities may differ, and the choice of the correct values relies on the merging algorithm. A merged entity may therefore include at different levels some uncertainties, which have to be presented to end-users [18].

In this paper, we are interested in visualizing the uncertainty resulting from the merging process of spatial entities. Our contributions can be summarized as follows: (i) identifying the "dimensions" which have to be taken into account for uncertainty, i.e., the POI type, the spatial attributes and the terminological (non-spatial) attributes; (ii) measuring the confidence level for each dimension as well as a global confidence score; (iii) proposing many visualizations of a merged entity and its uncertainty and testing them to select the best; (iv) implementing a prototype to demonstrate in a scenario the benefits for end-users.

The next section describes the related work, both in spatial integration and uncertainty visualization. Section 3 provides a detailed explanation of our solutions, tested among different users, to represent and visualize various criteria about a merged entity. In Section 4, we demonstrate the benefits of our approach in a scenario, and we conclude in Section 5.

2. Related work

This section covers the existing works in two domains: the methods for integrating spatial data and the visualization of uncertainty in a spatial context.

2.1. Spatial integration

The same reality is described with a multiplicity of geographical information. This information growth rapidly over the Internet, some may be incomplete, inaccurate or contradictory. Integration of several sources of geographical information is necessary in order to update information that changes daily [12] or to produce a more complete and accurate information [7]. In [32], authors define three categories of imperfection that occurs when integrating several spatial data sources, namely (i) inaccuracy, which concerns wrong spatial information that do not correspond to reality, (ii) imprecision, which deals with spatial information that corresponds to reality but is not sufficiently precise and (iii) vagueness, which is about ambiguity of spatial information (e.g., boundaries heterogeneity). Geospatial integration has been widely studied under the term "map conflation" where two whole maps are integrated. Integration of maps consists in identifying the corresponding entities and to fuse them [5]. In [22], authors describe existing works in map conflation regarding their formats (raster and vector) and their criteria such as spatial data, terminological data and topological relationships between entities. Some works have been proposed in map conflation using points [23,6,30], lines [24,10,31] and polygons [1,11,19].

In [2,25], the authors use only the spatial information (location) to detect the corresponding entities with a similarity measure based on probabilistic consideration. The probability that two entities are corresponding is estimated using the Euclidean distance between them. In order to improve the quality of integration, some works propose to combine similarity measures that use spatial information with similarity measures that use terminological information to identify correspondences. In [26], three algorithms were proposed using a first similarity measure to filter the entities and a second to detect the corresponding entities. For example, a string similarity measure can be applied on the name of the POI, then for each pair of entities that are not considered as corresponding, the distance between them is increased. The final step is to apply a similarity measure on spatial information with the new distances. Note that increasing the distance between two entities lowers the probability that they will be considered as corresponding entities when we apply a similarity measure on spatial information. A variety of learning-based methods including logistic regression, support vector machines and voted perceptron has been proposed to find out how to combine and tune several similarity measures in order to identify the corresponding entities [27].

The "Theory of Evidence", also called "Dempster–Shafer theory" [28], combines an evidence measure of different sources and finds a degree of belief that takes into account all the available evidence. That is, a belief mass represented by a belief function, is associated to each evidence, then Dempster's rule is used to combine them. The "Theory of Evidence" is proposed to integrate geospatial databases [21] and to match geospatial entities of several LBS providers [14].

Kang et al. propose a visual interface to detect the corresponding geospatial entities based on a neighborhood similarity [13]. It takes two sources of entities as input, and then the user chooses a similarity measure to apply on terminological information or on spatial information. Detected entities are considered as potentially corresponding. Then each pair of entities is visualized on the screen. Their shared neighborhood of entities are placed between them and non-shared neighbors on the sides. Finally, the user has to make a decision for each pair to be considered as corresponding or not.

2.2. Spatial uncertainty visualization

Thomson et al. [29] and MacEachren et al. [17] define nine categories of uncertainty paired with three components of geographic information: space, time and attribute (terminological). On this basis, Thomson et al. [29] make an empirical study to characterize the kind of visual signification that is appropriate for representing those different categories of uncertainty. The authors use a set of visual variables corresponding to the visual variables defined by [4,20,16]: Location, Size, Color Hue, Color Value, Grain, Orientation, Shape, Color Saturation, Arrangement, Clarity, Resolution and Transparency. Their symbol sets are points and for each visual variable, three levels are specified coming from high to low certainty (Fig. 1). They add iconic/pictorial symbols to compare their efficiency according to abstract/geometric symbols such as Smiley, Filled bar with Slider, and Thermometer (Fig. 2). Two tests are realized to judge the suitability of different symbol sets for representing variation in uncertainty by manipulating one single visual variable for all the categories of certainty in all the components of geographic information or for one specific category of certainty (accuracy, precision, trustworthiness) in each component of the geographic information.

3. Representing uncertainty

This section covers our contributions for representing uncertainty in spatial integration. We first introduce an overview of our approach. Next we focus on the integration process, which produces confidence scores, and on uncertainty visualization on maps.

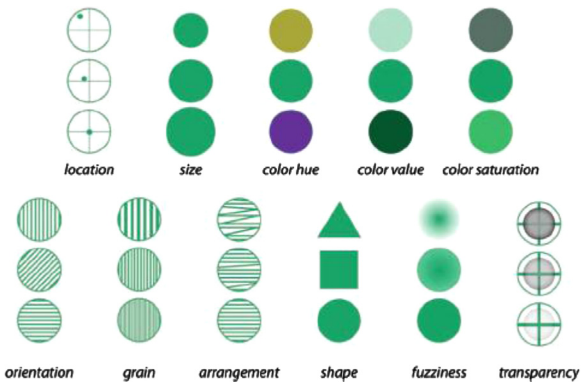


Fig. 1. Visual variables proposed by [18].

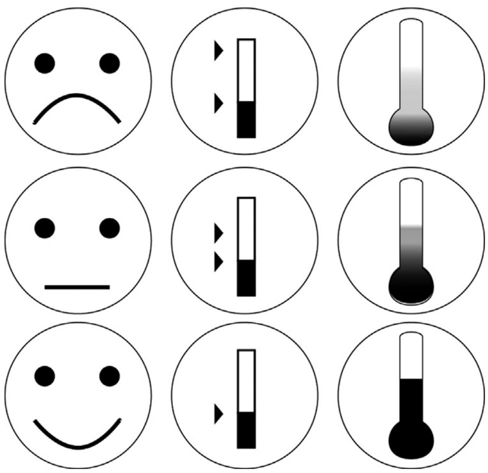


Fig. 2. Smiley, Filled bar with Slider, and Thermometer icons proposed by [18]

3.1. Approach overview

A (semi-)automatic integration process does not achieve perfect results. Depending on data quality of providers, an integration process may have to deal with various kinds of uncertainty and to take decisions. In our geospatial context, the quality is strongly variable from one provider to another, and we need to take into account the uncertainty inherent to the process. Besides, this uncertainty should be represented, especially on a map. Our integration approach consists of three consecutive processes, namely "mediation process", "integration process" and "visualization process".

- Mediation process: it is in charge of processing and rewriting a spatial query. For each LBS provider, the initial query is rewritten to comply with the schema or model of each provider. In addition, the mediation process performs a blocking process, which reduces the set of returned entities based on the location (within a radius) and the POI type specified in the query. As an example, let us imagine that a user is interested in finding the hotels in Pittsburgh. This query may be rewritten as "accommodations in Pittsburgh" for a first provider, and as "hotels in Pittsburgh, PA" for another provider. The output of this process is a set of entities returned from each provider to the mediator that are ready for the integration part. Note that the mediation is not further discussed in this paper, since the schema heterogeneity of the providers has been beforehand manually solved and that the blocking processes are performed using the providers' querying systems.
- Integration process: it aims at detecting and merging spatial entities which refer to the same POI (corresponding entities). It takes the sets obtained from the mediation process to produce a single set of entities, in which corresponding entities from different providers are merged into a single entity. Our integration process produces various confidence scores between the attributes of corresponding entities (see Section 3.2). The lower the uncertainty, the higher the confidence levels. Note that any spatial integration system, which takes the same inputs, could be used in replacement.
- Visualization process: its main objective is the transformation of the confidence scores into visual representation of confidence levels (see Section 3.3). In this process, the merged entities resulting from the integration are displayed on a map.

3.2. Integration and uncertainty computation

The challenges in entity integration traditionally deal with the selection of data and transformation functions to be used for merging. In our context, we can add the computation of relevant and useful confidence scores for spatial and terminological attributes. In this part, we describe a simple solution for detecting and matching corresponding entities and for computing confidence levels.

Many generic approaches for "entity matching" have been proposed [15]. Getting inspired by these generic approaches, we propose a simple entity matching process based on sophisticated similarity measures. The matching process is performed between all entities resulting from the mediation process. Given two entities from different sets, we compute confidence scores between their attributes. A score close to 0 means that two entities are totally dissimilar. Conversely, a score equal to 1 indicates that both entities are equivalent. The coordinates of two entities are compared according to the Euclidian distance. The shorter the distance between both entities, the closer to 1 the similarity value for coordinates is. All terminological attributes (e.g., name, phone) are compared using the Levenshtein measure. This measure is the most effective with regard to other string similarity measures [27]. Using several metrics to match the same attribute involves a new problem for combining the different similarity scores. When all the individual scores have been computed, we may also compute a global score. A weighted average is traditionally used for combining the individual similarity scores. A decision step is finally required to select the correspondences. Various methods such as a threshold or the top-K enable this automatic selection [3]. In our context, proposing the corresponding entities with the highest global score is sufficient. To select which attributes of corresponding entities should be merged, we apply statistics (mainly value frequency). This simple proposition of entity matching and merging aims at illustrating our uncertainty visualization solutions. Note that any integration algorithm, which takes the same inputs, can replace our proposition.

Concerning the output of the confidence scores, they are deduced from the similarity scores computed during the entity matching. The score computed between the coordinates constitutes the spatial confidence score. All terminological scores (between names, phones, etc.) are averaged to become the terminological confidence score. The global confidence score aims at evaluating the global confidence about a merged entity. For instance, the integration process produces a high score of spatial confidence when two providers locate the same POI at the same place and a low score of terminological confidence when two providers provide the user with totally different names, addresses, telephones, websites, etc. At the end of the integration process, corresponding entities have been merged and three confidence scores have been computed for each merged entity. The next step consists in visualizing these scores on a map.

3.3. Uncertainty visualization proposal and assessment

Visualization of integrated information may be insufficient in various cases. For instance, a user needs to check original information when observing strange outcome from the integration process. Therefore, the user requires to estimate himself the confidence of integration process visualizing (i) the spatial and terminological uncertainty for each integrated POI and (ii) the whole providers' source information. This requirement generates a large amount of information that might become an issue to visualize. To meet this requirement, in our approach, we first convert the spatial, terminological and global scores output from the integration process into three confidence levels (similar to the three uncertainty levels in [29,18]): uncertain (low confidence level), moderately certain (medium confidence level), certain (high confidence level). The first range [0, 0.5] is associated to the uncertain level. The middle range (]0.5, 0.75]) includes the moderately certain values. And the certain level stands for highest values in the range]0.75, 1]. These ranges have been fixed according to experiments performed with similarity measures [8]. In the future, we intend to learn the best ranges for each level.

We are interested in monitoring uncertainty of two dimensions: the confidence level of spatial attributes (the spatial confidence score from integration process) and the confidence level of terminological attributes (the terminological confidence score from integration process). Moreover, to create a map easier to read and understand for a tourist, we propose to group these two dimensions of confidence to display one global confidence level. Then, a POI has a global (spatial and terminological) high confidence level when the data of the providers are consistent and complete between them. On the contrary, a POI has a global low confidence level when the providers are not consistent and/or not complete between them.

An analysis of the results obtained by [18] leads us to select the most relevant data useful in our context. Location, Size associated to Fuzziness variables are relevant to portray spatial uncertainty. Smiley, Filled bar associated with Slider and Thermometer are interesting to portray terminological uncertainty. Finally, Fuzziness, Location and Color Value are well suited to portray global uncertainty.

We define various cartographic proposals to portray confidence levels of POI that are oriented in two directions: first the choice of the visual variables, second the choice of the dimension(s) of the geographic information to display on the map.

3.3.1. Visual variables to portray confidence levels

On the basis of conclusions made in Section 2.2, we propose two alternative visual variables to portray the confidence level of each dimension of geographic information (spatial, terminological, global). Fig. 3 illustrates them.

Concerning the spatial attributes, we decide to compare Location with Size associated to Fuzziness. We choose Location because it is intuitively connoted to space. We aggregate Size and Fuzziness. The taller the sign is, the fuzzier the sign is. We do this combination because independently, an order would be created between the signs with large or distinct signs seen before the others. This combination reduces this perception of order.

Concerning the terminological attributes, the proposals of [29,18] have been investigated. For our application,

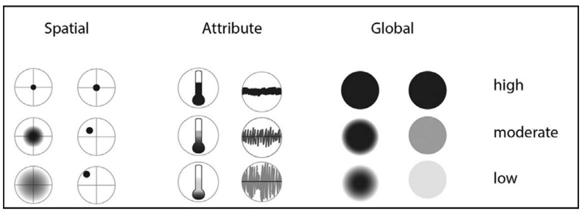


Fig. 3. Visual variables chosen for our study.

Smiley is too connoted to a score relative to the quality of a POI obtained from the opinions of different users. Then if the smiley is happy, it will be interpreted as a good POI for the public (e.g., a "good" restaurant) and this is not what we want to represent. Concerning Filled bar associated with Slider, we think it is difficult to correctly perceive the differences between its three degrees because only one small element of the slider is modified. For the previous reasons, the Thermometer icon is selected and is compared to a new visual variable: Frequency, based on graphic representations created to show uncertain chaotic behaviors of signals in Electronics Science.

Finally, for the global confidence level (spatial and terminological attributes together), we choose to combine Fuzziness with Color Value, the Location visual variable is eliminated because it is too connoted to the spatial dimension.

Perceptual tests have been conducted to determine which semiotic solution is best perceived and understood for each level of confidence based on the statistical Student *t*-test [34] with a threshold p of acceptability of risk up to 5% (i.e., if p is lower than 5% then the results are significant). For each visual variable (Location, Size associated to Fuzziness, Thermometer icon, Frequency, Fuzziness, Color Value), a couple of icons (weak vs. medium, strong vs. weak, strong vs. medium) are presented to 36 participants (14 men and 22 women, aged 18–30 years) based on a map. Participants must indicate as quickly as possible which icon represents the highest degree of confidence. These tests are repeated for each level of confidence with a counterbalance in the order of proposals to avoid learning effect. These tests enable us to evaluate the comprehension level of variables' icons between each others.

Results are measured in terms of response time and correctness. The Student *t*-test reveals significant effects in both results ($p < 3{,}14\%$) for Size associated to Fuzziness (spatial level), Thermometer (terminological level) and Color Value (global level), according to both response time and correctness. These three visual variables are then selected for the next step of assessment.

3.3.2. *Dimensions of geographic information to be displayed on map*

Portraying whole uncertainty information may overload the interface. Our approach proposes instead to portray the confidence levels with a cartographic interactive application that gives the advantage to provide the user with only main confidence information and get more confidence details on demand opening a tool-tip to display complementary information. The user can also interact with the map (zoom in/out, move around, etc.). In such an application, various visualization strategies can be proposed depending on various confidence information we can highlight on the map.

A new test has been realized with 25 participants (14 men and 11 women, aged 22–59 years) to determine which confidence level is the most important to the user. This test is conducted in two stages. Results have been analyzed based on Chi-squared test χ^2 [35] in order to ensure the significance of responses and to reject any random behaviors, always with a threshold p of acceptability of risk up to 5%.

First, we explain to the participants the meaning of the three distinct geographic dimensions. Then we ask them to rank the dimensions without considering any map or legend. Significantly ($\chi^2 = 20.8$, $p = 0.034\%$), participants place the spatial confidence level as the most important (for 13 participants) followed by the global confidence level (for 7 participants).

Secondly, participants see five different proposals. In the first two proposals (proposal 1 and proposal 2), spatial and terminological confidence levels are represented respectively. In these proposals, for each POI, we portray only the confidence level of this more significant dimension whereas the other one is shown in its tool-tip (Fig. 4 and 5). In the third proposal, both spatial and terminological levels are represented (proposal 3). In Fig. 6, spatial and terminological confidence levels are both portrayed on each POI using two signs. In the fourth proposal, a global confidence level is displayed for each POI, corresponding to the confidence combination of spatial and terminological attributes (proposal 4). In this case, the tool-tip of each

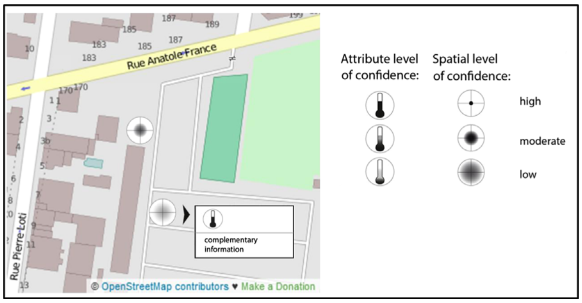


Fig. 4. Proposal 1: spatial confidence level displayed on the map.

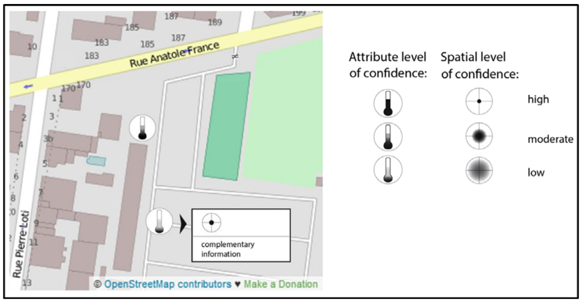


Fig. 5. Proposal 2: terminological confidence level displayed on the map.

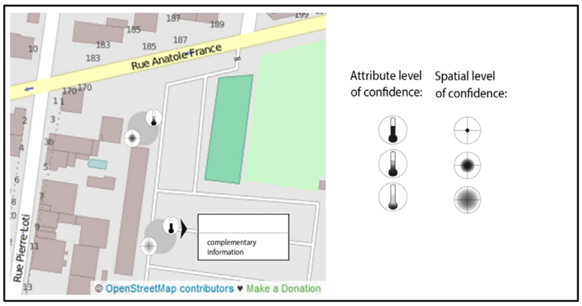


Fig. 6. Proposal 3: spatial and terminological confidence levels are both portrayed.

POI shows the spatial and terminological confidence levels (Fig. 7). Finally, global, spatial and terminological confidence levels are all portrayed together for each POI as shown in Fig. 8 (proposal 5).

After presenting the five proposals to the participants, they must give an appreciation of the application on a scale of 0 (not satisfied) up to 7 (totally satisfied). The retrieved data were subjected to analysis of variance (ANOVA), based on Fisher's test [33], with repeated measures which reveal significant effect on the type of the proposal ($F=3.19$, $p=1.6\%$). Proposal 1 (spatial confidence) and proposal 4 (global confidence) are the most popular. Fig. 9 represents the average score on the appreciation scale according to proposals. It shows that the preferences do not increase neither by adding the terminological confidence (proposal 3) nor by presenting together the spatial and global confidences (proposal 5).

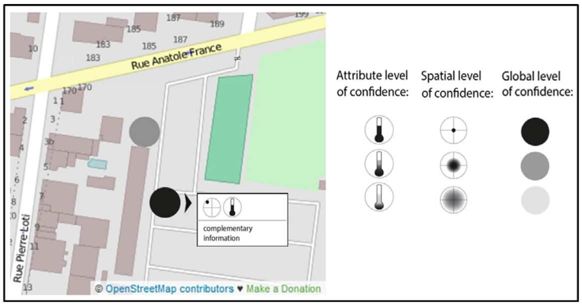


Fig. 7. Proposal 4: global confidence level is portrayed. Spatial and terminological confidence levels are displayed in the tool-tip.

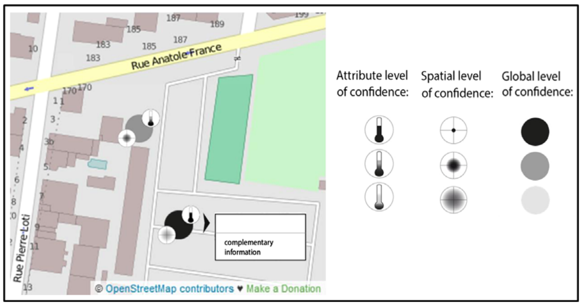


Fig. 8. Proposal 5: global, spatial and terminological confidence levels are all portrayed together.

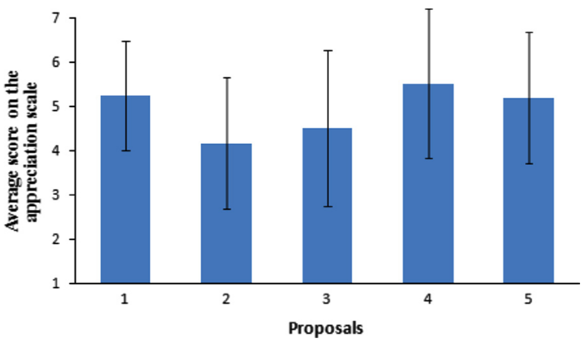


Fig. 9. Average score on the appreciation scale according to proposals.

Table 1
Number of times the POIs were selected based on strategies.

Strategy	Number of selected POIs
Precision of global level	74
Precision of spatial level	17
Precision of terminological level	11
Map	14
Other	9

Then participants must choose, among the five proposals, the one that is the most relevant. A Chi-squared test χ^2 has been conducted on the distribution of preferred proposals for participants. A trend effect ($\chi^2=117.64$, $p=5.6\% > 5\%$) has confirmed a preference for proposal 1 (spatial confidence) and proposal 4 (global confidence).

Finally, on each proposal, a set of POIs are presented to participants and we ask them to freely choose a POI where they would like to go by explaining their strategy of choosing. Emerged strategies may be based either on the confidence levels of geographical dimensions or on other elements such as the background map, the type of POI, etc. Table 1 shows various strategies and for each the number of times the POIs were selected based on this strategy. The analysis of POIs choices shows significantly ($\chi^2=117.64$, $p<0.1\%$) that participants select, as a priority, the POIs where the level of confidence is the highest for the global dimension of whatever proposal. Choosing a POI is less frequent based on the level of spatial confidence or according to other elements on the map (e.g., proximity to transportations, green spaces) and even less frequent according to the terminological confidence level.

The next section illustrates some of our proposals by describing a use-case navigation scenario of a prototype we have implemented. The proposal retained for the selection of visual variables is the result of the cognitive test namely Color Value for global confidence level and in the tool-tip of each POI: Size and Fuzziness to portray spatial confidence level and Thermometer to portray terminological confidence level.

4. Prototype

Our proposal has been integrated in a LBS prototype. The POIs of this service are the result of the integration of the POIs from several LBS providers. This prototype implements the choice of solution presented above for visualizing uncertainty of integrated spatial data.

The prototype runs on an ad-hoc POI database that has been created collecting POIs of several types from three real providers using their Application Programming Interface (API). The integration process is pre-performed on the whole POI database and the prototype interface navigates through the result. The prototype interface is composed of three components as shown in Fig. 10: (1) POI types selector: a list that the user check/uncheck to display or hide, (2) legend: the visualization solution used to portray global, spatial and terminological confidence levels and (3) map inheriting OpenStreetMap background and features (zoom in/out, satellite/map view, etc.). The user can choose two modes for the map, the former denoted as “Integrated mode” displays integrated POIs with their global

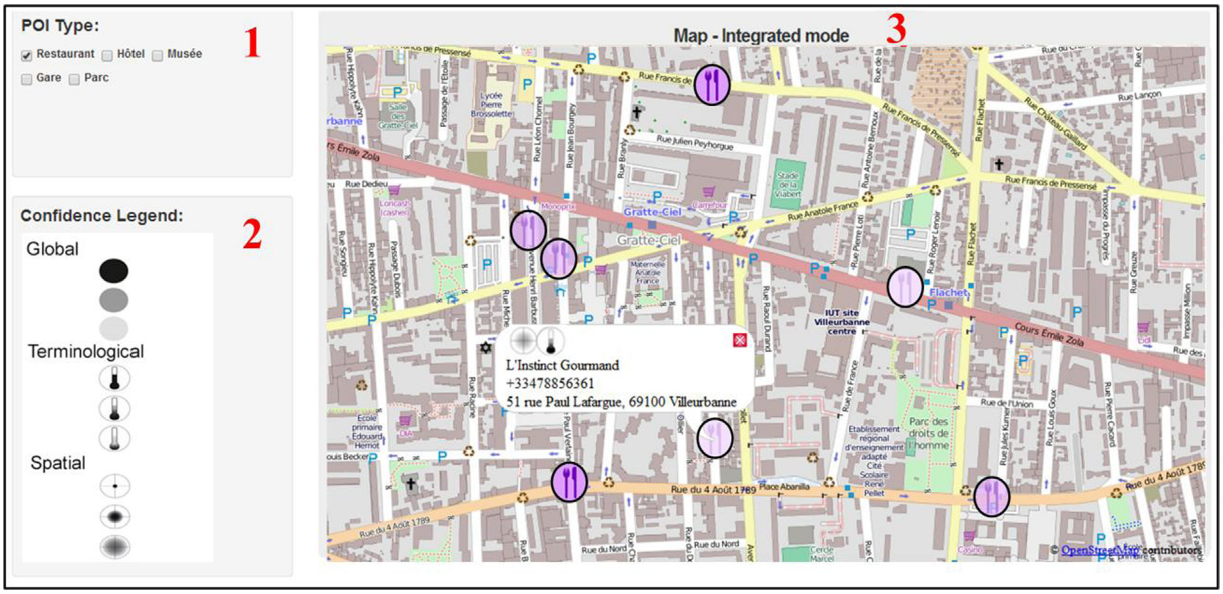
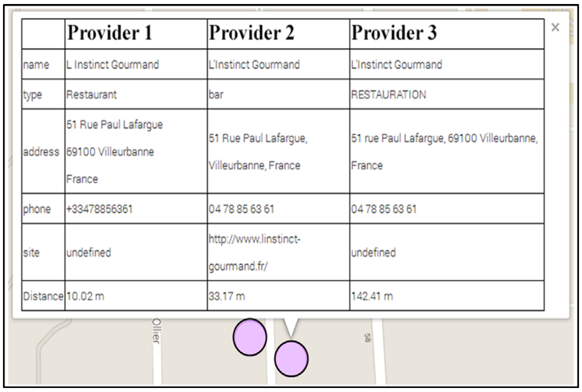


Fig. 10. The prototype interface is composed of three components: (1) POI types, (2) legend and (3) map (here in Integrated mode).



	Provider 1	Provider 2	Provider 3
name	L Instinct Gourmand	L'Instinct Gourmand	L'Instinct Gourmand
type	Restaurant	bar	RESTAURATION
address	51 Rue Paul Lafargue 69100 Villeurbanne France	51 Rue Paul Lafargue, Villeurbanne, France	51 rue Paul Lafargue, 69100 Villeurbanne, France
phone	+33478856361	04 78 85 63 61	04 78 85 63 61
site	undefined	http://www.linstinct-gourmand.fr/	undefined
Distance	10.02 m	33.17 m	142.41 m

Fig. 11. Comparison of terminological information offered by several providers for the same POI.

confidence levels. The latter, denoted as "Source mode", portrays the POIs of the source providers of an integrated POI with full information.

When the user starts navigating, the prototype detects and centers the map at user location, and the "Integrated mode" is set by default. The user selects the POI types from the selector. All the POIs of the selected types that are near the user location are collected from the integrated dataset and displayed on the map. The map window of Fig. 10 illustrates the POIs of type Restaurant, Color Value variable indicating global confidence. Two deep colored restaurants have a high global confidence level (top and bottom), two light colored restaurants have a low global confidence level (in the center) and the three remaining have a medium global confidence level. The user can click on a POI to display the tool-tip that contains the full POI terminological information, spatial confidence and terminological confidence of the integration as shown in Fig. 10. At the right top corner of the tool-tip, the Thermometer icon indicates that the terminological confidence is medium while the Location, Size, Fuzziness icon indicates that the spatial confidence is low for the selected POI.

As well, the user can check the source providers of an integrated POI by switching to the "Source mode" where all the integrated POIs are hidden except the checked one. In this mode, the user can consult the full POI information delivered by all the source providers. This mode shows the location of the integration result and of all source providers that the user can compare. The user can also check out terminological information of every source and compare them all. A table that contains POI terminological information delivered by each provider can be displayed (Fig. 11). Also, the distance between each source POI and the integrated one is indicated at the bottom of the table for each provider.

5. Conclusion

In this paper, we have proposed and studied different representations of uncertainty in a spatial integration context. Our approach is generic and the simple integration process that we have presented can be replaced by another generic approach possibly based on different assumptions. The integration process merges corresponding entities and produces confidence scores at spatial, terminological and global levels. These confidence scores are converted into confidence visualization solutions that have been evaluated among many users. Solutions have been implemented into a first application prototype to demonstrate the feasibility and the benefits in a scenario.

One of our future objectives is to customize the visual representation and the navigation process according to users' profiles. To reach this goal, we plan to explore and evaluate how tourists navigate with interactive maps. These evaluations should allow us to select the solution that is both the best perceived and the most useful for the tourists according to their expectations. They could also demonstrate how such uncertainty representation is partly user-dependent. In that case, learning automatically the best representation for a new user could be an interesting challenge. For instance, a dynamic prototype which allows users to customize the mode of representation would allow us to evaluate the preferred solutions and to identify their correlations with various criteria such as user profile and device type (e.g., computer or smart phone). Otherwise, the prototype should also allow users to send their feedback about the quality of integrated POIs in order to improve the quality of data and the integration process.

Acknowledgment

This work was supported by the LABEX IMU (ANR-10-LABX-0088) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). Special thanks go to OpenStreetMap for providing the background maps and to Professor VERSACE Rémy for his advice for perceptual tests.

References

[1] E.M. Arkin, L.P. Chew, D.P. Huttenlocher, K. Kedem, J.S.B. Mitchell, An efficiently computable metric for comparing polygonal shapes, IEEE Trans. Pattern Anal. Mach. Intell. 13 (3) (1991) 209–221.
[2] C. Beeri, Y. Doytsher, Y. Kanza, E. Safra, Y. Sagiv, Finding corresponding objects when integrating several geospatial datasets. ACM International Workshop on Geographic Information Systems, 2005, pp. 87–96.
[3] Z. Bellahsene, A. Bonifati, E. Rahm, Schema Matching and Mapping, Springer-Verlag, Heidelberg, 2011.
[4] J. Bertin. Semiology of Graphics: Diagrams, Networks, Maps, 1983.
[5] M.L. Casado, Some basic mathematical constraints for the geometric conflation problem, in: International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 2006, pp. 264–274.
[6] C.-C. Chen, S. Thakkar, C.A. Knoblock, C. Shahabi, Automatically annotating and integrating spatial datasets, in: Symposium on Spatial and Temporal Databases, 2003, pp. 469–488.
[7] M.A. Cobb, F.E. Petry, K.B. Shaw, Fuzzy spatial relationship refinements based on minimum bounding rectangle variations, Fuzzy Sets Syst. 113 (1) (2000) 111–120.
[8] W. Cohen, P. Ravikumar, S. Fienberg, A comparison of string metrics for matching names and records, Workshop on Information Integration on the Web, 2003, pp. 73–78.
[9] T. Devogele, C. Parent, S. Spaccapietra, On spatial database integration, Int. J. Geogr. Inf. Sci. 12 (4) (1998) 335–352.
[10] Y. Doytsher, A rubber sheeting algorithm for non-rectangular maps, Comput. Geosci. 26 (9) (2000) 1001–1010.
[11] M. Gombosoi, B. Zalik, S. Krivograd, Comparing two sets of polygons, Int. J. Geogr. Inf. Sci. 17 (5) (2003) 431–443.
[12] I.N. Gregory, Time-variant gis databases of changing historical administrative boundaries: a European comparison, Trans. GIS 6 (2) (2002) 161–178.
[13] H. Kang, V. Sehgal, L. Getoor, Geoddupe: a novel interface for interactive entity resolution in geospatial data, in: Proceedings of International Conference on Information Visualisation, 2007, pp. 489–496.
[14] R. Karam, F. Favetta, R. Kilany, R. Laurini, Integration of similar location based services proposed by several providers, Networked Digital Technologies, 2010, 136–144
[15] H. Kopcke, E. Rahm, Frameworks for entity matching: a comparison, Data Knowl. Eng. 69 (2) (2010) 197–210.
[16] A. MacEachern, How Maps Work, The Guilford Press, New York, London, 1995.
[17] A.M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, E. Hetzler, Visualizing geospatial information uncertainty: what we know and what we need to know, Cartogr. Geogr. Inf. Sci. 32 (3) (2005) 139–160.
[18] A.M. MacEachren, R.E. Roth, J. O'Brien, B. Li, D. Swingley, M. Gahegan, Visual semiotics and uncertainty visualization: an empirical study, IEEE Trans. Vis. Comput. Gr. 18 (12) (2012) 2496–2505.
[19] A. Masuyama, Methods for detecting apparent differences between spatial tessellations at different time points, Int. J. Geogr. Inf. Sci. 20 (6) (2006) 633–648.
[20] J.L. Morrison, A theoretical framework for cartographic generalization with the emphasis on the process of symbolization, Int. Yearb. Cartogr. 14 (1974) 115–127.
[21] A. Olteanu, A multi-criteria fusion approach for geographical data matching, International Sympoion in Spatial Data Quality, 2007.
[22] J.J. Ruiz, F.J. Ariza, M.A. Urena, E.B. Blazquez, Digital map conflation: a review of the process and a proposal for classification, Int. J. Geogr. Inf. Sci. 25 (9) (2011) 1439–1466.
[23] A. Saalfeld, A fast rubber-sheeting transformation using simplicial coordinates, Am. Cartogr. 12 (2) (1985) 169–173.
[24] A. Saalfeld, Conflation automated map compilation, Int. J. Geogr. Inf. Syst. 2 (3) (1988) 217–228.
[25] E. Safra, Y. Kanza, Y. Sagiv, C. Beeri, Y. Doytsher, Location-based algorithms for finding sets of corresponding objects over several geo-spatial data sets, Int. J. Geogr. Inf. Sci. 24 (1) (2010) 69–106.
[26] E. Safra, Y. Kanza, Y. Sagiv, Y. Doytsher, Integrating data from maps on the world-wide web, Web and Wireless Geographical Information Systems, 2006, 180–191
[27] V. Sehgal, L. Getoor, P. Viechnicki, Entity resolution in geospatial data integration, in: ACM International Symposium on Geographic Information Systems, 2006, pp. 83–90.
[28] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[29] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, M. Pavel, A typology for visualizing uncertainty, Electronic Imaging, , 2005, 146–157.
[30] S. Volz, An iterative approach for matching multiple representations of street data, in: Proceedings of the JOINT ISPRS Workshop on Multiple Representations and Interoperability of Spatial Data, 2006, pp. 101–110.
[31] M. Zhang, W. Shi, L. Meng, A generic matching algorithm for line networks of different resolutions, in: Workshop of ICA Commission on Generalization and Multiple Representation Computering Faculty of A Coruna University-Campus de Elvina, Spain, 2005.
[32] M.F. Worboys, E. Clementini, Integration of imperfect spatial information, J. Vis. Lang. Comput. 12 (1) (2001) 61–80.
[33] R.A. Fisher, Statistical Methods for Research Workers, 4th ed. Oliver & Boyd, Edinburgh, 1932.
[34] Student, The probable error of a mean, Biometrika (1908) 1–25.
[35] Robin L. Plackett, Karl Pearson and the chi-squared test, Int. Stat. Rev. (1983) 59–72.