

Personalization in Geographical Information Search

Noemi Mauro
University of Turin
Computer Science Dept.
Torino, Italy 10149
noemi.mauro@unito.it

My PhD project focuses on the personalization of Participatory GIS (PGIS). In the project I analyze two methodologies to offer personalized search results in community maps and a natural interaction with the system. The first consists of automatically gathering the users interests at a concept level in order to generate clusters of concepts useful for the presentation of thematic maps. This is done by creating ontology-based user models mapped to the domain ontology adopted by the PGIS. The second concerns the creation of content-based user models useful for filtering the items belonging to each concept in a multi-faceted way: the goal is that of reducing and adapting the information space presented in the map. In the project I also analyze how these techniques may be jointly used during the query expansion process to retrieve more accurate and relevant search results.

Semantic search, personalization, user model, participatory GIS, linked data, ontologies

1. INTRODUCTION

With the digitization of the information and the development of crowdsourcing a large amount of data concerning the territory is available. Geographical maps have proven to represent the territory in an intuitive and expressive way, in order to help users orientate themselves during the exploration of the information space leveraging the geographical position. Specifically, the management and analysis of spatial data is supported by PGISs (Participatory Geographic Information Systems), because they offer a visual representation of geographical information based on dynamic map generation and map sharing within and between stakeholder groups. However, users can be challenged with an information overload issue in dynamic maps, that has to be solved to make information really usable to people (Ardissono et al. (2016)). For this purpose, an efficient and accurate method is required for building an information system that correctly answers the user's information needs, possibly tailoring results for improving their fruition. The main goal is to enable the user to specify her/his information needs in a natural way and tailor search results accordingly. In the related research, query expansion techniques, a natural language model for interpreting the queries, and personalization techniques for showing on the map tailored suggestions have been used. For this

purpose, the semantic representation of geographic information is exploited, following the Geospatial Semantic Web's approach (Janowicz et. al (2012)). The idea is to guide the user in the exploration of the information space, but also providing to her/him extra-helpful information in order to support serendipity and not only helping her/him find the data (s)he explicitly searches.

In order to improve the fruition of information, in this PhD project I will analyze an interaction method based on natural language processing that interprets the search queries submitted by the user and offers different visualizations of the results, based on the her/his interests and behavior. The idea is that, depending on the user's behavior, different aspects of information are relevant, and an information system should take this into account for the presentation of data.

In order to support the exploration of territorial data under different points of view, I propose a model that combines the interpretation of the user's queries with a map management model to generate dynamic maps that reflect individual information needs. This model is composed of two elements at two different levels of abstraction:

1. The former concerns the creation of ontology-based user models useful for suggesting

clusters of concepts that are usually searched together by people or by a specific user.

2. The latter concerns the creation of content-based user models useful for gathering the users' interests about the information items belonging to each concept (e.g., price range of restaurant, ratings, etc.). This type of model enables the system to filter information items depending on the preferences expressed by users during the interaction with the system.

The remainder of this paper is organized as follows: Section 2 positions our work in the related one. Section 3 presents the main features of the OnToMap application. Section 4 describes the adoption of ontology-based user models useful for the creation of thematic maps, Section 5 presents the adoption of content-based user models and Section 6 describes the evaluation plan of the work. Section 7 concludes the paper and outlines some future work.

2. BACKGROUND

The Web helps the diffusion of information by making available the results of the massive digitalization effort, carried out by cultural organizations and/or crowdsourced as Voluntary Geographic Information (Goodman and Litman (2007)), by means of online information sources and archives. However, the lack of a framework for semantic data integration, contextualized in a geographical perspective, limits the fruition of this type of information. Specifically:

- Museums, exhibitions and historical sites publish data about physical and digital collections in their own web sites; e.g., see the Tate Online (<http://www.tate.org.uk/>) and (Museums and the web (2017)). They present detailed information about art pieces and historical objects. However, they are disconnected from each other and thus provide a partial view on the resources available in the territory.
- Large projects have been developed to support the digitalization and diffusion of artistic and historical content across Europe. However, the resulting services, such as Europeana Collections (Europeana (2017)), are managed as catalogs of artworks abstracting from the geographical dimension of information items. These services enable the user to retrieve data according to a fixed number of facets (e.g., by source, color, topic). However, they fail to integrate such search criteria with the geographical aspect.
- Modern search engines, location-based services (e.g., TripAdvisor (TripAdvisor (2017)))

and Web-GIS (e.g., OpenStreetMap (OpenStreetMap (2017))) offer graphical interfaces that support information search and a map-based visualization of results. However, they cannot tune the presentation of the territory to the aspects underlying search queries because they lack a semantic approach to their interpretation and to the identification of the relevant types of information.

It should be noticed that other works, such as Google's Knowledge Graph (Google Knowledge Graph (2017)), acquire links among individual information items, that people often search together, in order to recommend their inspection, when the user selects one of them. In a complementary way, the clusters we propose are formed by *concepts*, in order to suggest relevant types of information for expanding the user's search. They support the exploration of territorial data at an abstract level and thanks to the content-based user models the information space is still filtered, instead of presenting a possibly large sets of individual information items to analyze.

3. ONTOMAP PROJECT

The PhD project is framed in the context of "OnToMap - Mappa di Comunità 3.0" (<https://ontomap.ontomap.eu>), that aims at supporting citizens' participation in public policy design, as well as knowledge sharing about territorial information. The main result of that project is the development of a web application which can be used to consult spatial data, create custom community maps, report to local administrators critical issues or new proposals in order to give a local representation of the territory. The main aspects of OnToMap are (i) the use of an ontological layer supporting the integration of heterogeneous data (managed as Linked Data) and describing semantic relations to enhance the exploration of the information space, (ii) the support to the management of personalized and persistent community maps for project development, and (iii) the crowdsourcing of new items related to the concepts of the ontology to enrich the knowledge base of the application (Ardissono et al. (2017)). Within project MIMOSA ("progetto di Ateneo Torino_call2014_L2_157", 2015-17), OnToMap was extended with information search functions for improving the users' experience in browsing and searching data, developing a semantic-based search engine (Mauro et al. (2017)). Several applications use a category-based search or a keyword-based search approach. Differently, in MIMOSA a semantic search engine was developed that semantically interprets the search query and, by leveraging the

ontological layer, offers suggestions relevant to the identified concepts.

4. ONTOLOGY-BASED USER MODELS

The first aspect of my PhD project concerns the creation of ontology-based user profiles to customize the system and the information search task taking individual interests and knowledge needs into account (Jiang and Tan (2009)). The user profiles can be combined with the domain knowledge with the purpose of enriching the search queries and increasing the likelihood that none of the potentially interesting data is missing.

Each user model reflects the structure of the ontology. The system gives a weight at each concept contained in the user model, considering the user's interest gathered by analyzing her/his behaviour. Furthermore, each relation between concepts has a weight, in order to collect the relevant semantic relations for every user. For instance, if we consider a mother interested in the *school* and *park* concepts, it is likely that the weight of the relation between these two concepts in her model is high. The ontology-based user models allows the generation of groups of suggestions that can be called thematic maps. The thematic maps play the role of indexes on information and they support query expansion towards relevant topics that the user might not be aware of. For this purpose, I model a thematic map at a meta-level, as a cluster of concepts that are usually searched together by people. The visualization of data items on a map comes as a result of selecting one or more concept clusters, or portions of them, to retrieve the corresponding data. The organization of a map as a cluster of concepts enables the user to select the types of information (s)he is interested in and, as a by product, it allows the system to observe what (s)he focuses on. This is the basis to propose personal views on information.

I aim at creating dynamic thematic maps, by finding new clusters of concepts from users behavior, but also by updating the existing ones. I consider two types of dynamic thematic maps:

- The former is the most general one, and regards concepts that typically co-occur in the search sessions performed by the user population. By analyzing the system's log, it is possible to identify sets of concepts that are jointly searched by users. Usually, if a concept cluster is recurrent, it is relevant, because it reflects the most frequent combined search done by people. In this case, the cluster defines a thematic map whose concepts are weighted by frequency of usage

during the interaction with the system in a search task. I aim at automatically generating these clusters, by dynamically updating the occurrences of actions on concepts logged by the system during the interaction with users; e.g., selection of concepts suggested by the system, subsequent searches, and so forth.

- The latter reflects the individual user's search history and is aimed at capturing the maps that (s)he frequently uses by updating the ontology-based user models. The concepts belonging to each user model must be weighted by analyzing the behavior of the specific user. Starting from this, it is possible to deduce which are the most relevant concepts for each user, that can be used to expand the general dynamic maps presented above. Indeed, the user models can also be employed to rank the concepts of the dynamic thematic maps, and promote the most interesting ones. This supports a user-adaptive presentation of maps, instead of adopting a single visualization style, the same for everybody.

I hypothesize that the management of dynamic thematic maps, adapting to continuous observations, can be particularly useful to provide users with personalized groups of suggestions, represented as sets of concepts that can be easily explored, without browsing the complete information space.

5. CONTENT-BASED USER MODELS

The previously described method can be useful for providing clusters of concepts as suggestions. However, in order to guide the user in the exploration of the information space and filter the large amount of data belonging to each concepts, content-based user models are needed. In this case an information filtering approach that recommends information based on user profiles could be applied: these profiles are built by analyzing the content of the items that the user accessed in the past (Wanrong (2016)), possibly correlated with the activities' context he/she is involved in. The idea is that observing the users' behaviour, in particular the actions done on the information items (i.e., inspection and crowdsourcing of data), the system is able to learn the value of a set of features that represent their interests. This user model can be combined with the previous one and with the geographical component of the search queries to retrieve more relevant and accurate results. In this way the system can create adaptive geographic maps that reflect user information needs.

6. EVALUATION PLAN

The model described above will be evaluated as follows:

- Regarding the ontology-based user models, I plan an evaluation based on the analysis of search queries (and sessions) in a log collected by a search engine. It is difficult to have access to that kind of logs, but a possible one is the AOL one. In particular, starting from them, I will dynamically generate personalized clusters of concepts and I will evaluate if they represent the search sessions performed by the users. Indeed, also some experimentations with real users will be performed.
- For what concerns the content-based user models I plan to analyze logs of check-ins of people in services in order to match them to the preferences inferred in the content-based user models and evaluate if the suggestions on geographical items are reliable or not.

7. CONCLUSIONS AND DISCUSSION

The described PhD project aims at improving the interaction between the users and intelligent systems and at making it more natural, leveraging learning mechanisms and user modeling techniques to automatically generate an internal representation of the user profiles. The current natural language model adopted in OnToMap will be refined for retrieving better search results in terms of precision and recall, as well as user satisfaction. The developed methods will be tested by applying traditional HCI evaluation techniques. A relevant social impact of this project is related to knowledge diffusion and education. Thanks to a flexible information search model, people will be guided in the exploration of a large amount of data about the territory that, otherwise, would be hardly usable. Another impact is the support to public engagement in policy making, which presupposes the existence of an efficient knowledge sharing and information search model raising general awareness about the territory, as well as inspection of particular projections of information. Concerning the search engine's accuracy, it will be important to offer useful and personalized results and reduce the noise of the search results. In order to improve this capability, the model will not only analyze punctual information: it will also consider the context in which the items are placed and the users' interests learned by analyzing user behaviour.

8. ACKNOWLEDGEMENTS

I would like to thank my supervisor Prof. Liliana Ardissono from Computer Science Department of

University of Turin and SIGIR (ACM Special Interest Group on Information Retrieval).

REFERENCES

- Archives and Museum Informatics. 2017. Museums and the web. <http://www.museumsandtheweb.com/>
- L. Ardissono, M. Lucenteforte, N. Mauro, A. Savoca, A. Voghera, and L. La Riccia. 2016. Exploration of Cultural Heritage Information via Textual Search Queries. *In Proc. of MobileHCI '16 Adjunct. ACM*, 992–1001.
- L. Ardissono, M. Lucenteforte, N. Mauro, A. Savoca, A. Voghera and L. La Riccia. 2017. OnToMap: Semantic Community Maps for Knowledge Sharing. *Proceedings of the 28th ACM Conference on Hypertext and Social Media, ACM*, 317–318.
- Europeana. Europeana Collections. www.europeana.eu/portal/it
- B. Goodman and D. Litman. 2007. Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research 2 (2007)*, 24–32.
- Google. 2017. Knowledge Graph. <https://www.google.com/intl/it..it/insidesearch/features/search/knowledge.html>.
- K. Janowicz, S. Scheider, T. Pehle, and G. Ha. 2012. Geospatial Semantics and Linked Spatiotemporal Data – Past, Present, and Future. *Semantic Web - On linked spatiotemporal data and geo-ontologies 3, 4 (2012)*, 321–332.
- X. Jiang and A.-H. Tan. 2009. Learning and inferencing in user ontology for personalized Semantic Web search. *Information Sciences 179 (2009)*, 2794–2808.
- N. Mauro and L. Ardissono. 2017. Concept-aware Geographic Information Retrieval. *Proceedings of 2017 IEEE/WIC/ACM Int. Conf. on Web Intelligence (WI), ACM*, 2017.
- OpenStreetMap Contributors. 2017. OpenStreetMap. <https://www.openstreetmap.org>
- TripAdvisor. 2017. TripAdvisor. <https://www.tripadvisor.it/>
- G. Wanrong, D. Shoubin and C. Mingquan. 2016. Personalized news recommendation based on articles chain building. *Neural Computing and Applications 27, 5 (2016)*, 1263–1272.