Suggestion Models in Geographic Exploratory Search

Suggestion Models in Geographic Exploratory Search

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

ABSTRACT

My PhD project focuses on the suggestion of information categories in exploratory search in a geographical domain. Geographical maps may challenge the user in the exploration of possibly complex information spaces, making difficult to find all the relevant data for the completion of her/his search task. I propose different models for concepts suggestion which, given a search query, allow the creation of clusters of categories useful for query expansion as a *"you might be interested in"* function. The training of these models is done by exploiting different types of information: search sessions, users' preferences and social data coming from Twitter.

CCS Concepts

•Information systems → Information retrieval; Query log analysis; Query suggestion; Query reformulation;

Author Keywords

Geographical Information Retrieval; Session-based Concept Suggestion; Query Expansion.

INTRODUCTION

Exploring large datasets can be challenging for users especially when no support for the information navigation is provided by the system. Information search often happens in an "anomalous knowledge status" [3] and the modality of expression and the conceptualization adopted by the system can differ from the user's one. To address these issues, some systems offer different schemes for browsing data categories, such as lists or hierarchies. Sometimes the user deals with the exploration of complex information spaces organized as categories. Thus, users are often exposed to a possibly difficult selection task of the concepts presented. My PhD project focuses on the definition of different suggestion models for exploratory search in a geographical context. Exploratory search is affected by the above issues because the users' information goals are ill-defined; see [7]. I aim to provide users with small clusters of concepts that have the meaning of the suggestion function *"you might be also interested in"*. This may help the exploration of categories belonging to complex information spaces and suggest information related to users' search task.
I aim at answering the following research questions:

H1: *Is it possible to exploit frequently co-occurring concepts in a search session to suggest clusters of concepts relevant to the exploration of complex geographical information spaces?*

H2: *Is it possible to extract from Twitter local common interests of a specific geographical area to suggest clusters of concepts relevant for the user's search task?*

H3: *Is it possible to personalize the suggestion of clusters of concepts taking into account the user's specific information needs extracted from her/his user model?*

In order to answer these questions, I am developing different concepts suggestion models that propose the user a set of possible relevant data categories starting from the queries that he or she has already submitted. The suggested categories are complementary to the concepts that the user has focused on. They are suggested by exploiting different sources of data using different strategies.

RELATED WORK

My PhD project starts from the development of a textual search query interpretation model at a conceptual level. It leverages ontology and linguistic information, following an associative information retrieval model. However, the main focus of my project is the creation of different concepts suggestion models. They are based on the observation of concepts co-occurrence in search sessions, on the examination of population interests in Twitter in a geographical area, and on the creation of user models that represent the specific user's interests within a search session. My work differs from [6] because I aim at finding *concept co-occurrence* in search sessions, while they observe *term co-occurrence* for query expansion. Some works on information filtering and on recommender systems acquire relations among information items observing the users' behavior. For instance, Google Knowledge Graph (https://goo.gl/c56X3D) gather relations among facts, concepts, and items starting from their co-occurrences in queries. Those works are complementary to my PhD project. The main difference is that I employ a knowledge graph to predict further concepts that the user might be interested in, i.e., I aim to suggest topics, instead of recommending individual items.

METHODOLOGY AND SUGGESTION MODELS

I defined the different models starting from the analysis of search sessions from the AOL query log[1]. Since the goal of

[1] https://archive.org/details/AOL_search_data_leak_2006

IUI 2018, March 7–11, 2018, Tokyo, Japan.
ACM ISBN 978-1-4503-4945-1/18/03.
https://doi.org/10.1145/3172944.3173148

my work is the development of an intelligent search support function for Geographical Information Systems, I chose for the experiments an example system, i.e., OnToMap [2, 1]. I pre-processed the AOL search log to work on a smaller dataset composed of search sessions relevant to the geographical information provided by OnToMap. I selected those including at least one query that refers to the concepts of the OnToMap ontology. The ultimate goal of my work is to train different query expansion models suitable for improving information search in OnToMap. The clusters of concepts used as suggestions are created in two steps: (1) First, the creation of a concepts co-occurrence graph in which each node represents a concept and the edges weight reflects the strength of concept co-occurrence in search sessions, or in Twitter (it depends on the model). (2) Secondly, the application of the COPRA [5] community detection algorithm to the co-occurrence graph. COPRA works on weighted graphs and detects overlapping communities. Each community represents a cluster of concepts used as suggestions, while a concept can belong to different clusters at the same time. This supports capturing the fact that different sessions can have different sets of common concepts. Moreover, it is possible to suggest different groups of concepts, starting from the same search queries. In my PhD project, I plan to create different types of concepts suggestion models. They differ from each other by the data source used for the graph creation and from the strategies used to select clusters for query expansion.

Session-based Suggestion of Topics

The concepts co-occurrence graph is created from the analysis of search sessions in the AOL query log and is described in [8]. After the extraction of clusters of concepts from the graph, I analyzed different strategies for the clusters' selection starting from the first query submitted in each search session. Specifically, I investigated three different strategies: (1) Selection of clusters that contain all the concepts referred in the first search query. (2) Selection of clusters that contain at least one concept referred in the first search query. (3) Selection of the best matching cluster with respects to the first query of the search session. The results show that these suggestions strategies have the same accuracy (F1 score) but they differ in precision, or recall, respectively. Thus, depending on the priority of precision with respect to recall, it might be useful to select a specific strategy for suggestions. See [8] for details.

Personalized Suggestion of Topics

Starting from the strategies described in the above section, it could be interesting to analyze if the introduction of ontology-based user models [4] can improve the strategies' accuracy and can better reflect users' information needs. User models try to capture the specific interests of a user and their combination with the general interests of the population might improve the suggestion of more suitable clusters. It is also important to investigate how to tune the temporal span for the generation of the user models.

Suggestion of Topics Exploiting Social Data

Another part of my PhD project concerns the creation of the co-occurrence graph leveraging Twitter data grouped by user in a particular geographical area. In this case, the co-occurrence graph represents the social interests of the population and this information can be used for the construction of a social suggestion model. Starting from the first search query of a search session of AOL query log, the model suggests the clusters built using a social layer of the information.

CONCLUSIONS

I presented different concepts suggestion models helpful to support query expansion during exploratory search in a geographical context. The evaluation of the first model on the AOL log provided satisfactory results about its concept prediction accuracy. I plan an evaluation of the other two models, based on the analysis of search queries (and sessions) of the AOL query log and Twitter data useful for the construction of the clusters described in Section 3.3. In particular, I will evaluate whether the clusters represent the search sessions performed by the users in the AOL log. I also plan some experimentations with real users in order to check whether they explore the suggested concepts. Finally, I plan to combine the different models in order to evaluate whether the accuracy might improve and the models can better support users in the exploration of geographical information space. This work is funded by project MIMOSA (MultIModal Ontology-driven query system for the heterogeneous data of a SmArtcity, "Progetto di Ateneo Torino_call2014_L2_157", 2015-17) and by "Ricerca Autofinanziata"of the University of Turin.

REFERENCES

1. L. Ardissono, M. Lucenteforte, N. Mauro, A. Savoca, A. Voghera, and L. La Riccia. 2017a. Semantic Interpretation of Search Queries for Personalization. In *Proc. of UMAP 2017 Adjunct*. ACM, 101–102.
2. L. Ardissono, M. Lucenteforte, N. Mauro, A. Savoca, A. Voghera, and L. La Riccia. 2017b. OnToMap - Semantic Community Maps for knowledge sharing. In *Proceedings of Hypertext 2017*. ACM, 317–318.
3. N.J. Belkin. 1980. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science* 5 (1980), 133–143.
4. S. Gauch, J. Chaffee, and A. Pretschner. 2003. Ontology-based personalized search and browsing. *Web Intelli. and Agent Sys.* 1, 3-4 (2003), 219–234.
5. S. Gregory. 2010. Finding overlapping communities in networks by label propagation. *New Journal of Physics* 12, 103018 (2010).
6. C.K. Huang, L.F. Chien, and Y.J. Oyang. 2003. Relevant term suggestion in interactive web search based on contextual in formation in query session logs. *Journal of the American Society for Information Science and Technology* 54, 7 (2003), 638–649.
7. G. Marchiorini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.
8. N. Mauro and L. Ardissono. 2018. Session-based Suggestion of Topics for Geographic Exploratory Search. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces (IUI '18)*. ACM.