Second Workshop on Exploratory Search and Interactive Data Analytics (ESIDA)

Dorota Głowacka[★] **Evangelos Milios**[○] **Axel J. Soto**[△] **Fernando V. Paulovich**[○] **Denis Parra**[▽]

[★] University of Edinburgh   [○] Dalhousie University   [△] ICIC-CONICET   [▽] PUC Chile

[★]dorota.glowacka@ed.ac.uk, [○]eem@cs.dal.ca, [○]paulovich@dal.ca,

[△]axel.soto@cs.uns.edu.ar, [▽] dparra@ing.puc.cl

ABSTRACT

This is the second edition of the Workshop on Exploratory Search and Interactive Data Analytics (ESIDA). This series of workshops emerged as a response to the growing interest in developing new methods and systems that allow users to interactively explore large volumes of data, such as documents, multimedia or specialised collections, such as biomedical datasets. There are various approaches to supporting users in this interactive environment ranging from the development of new algorithms through visualisation methods to analysing users' search patterns. The overarching goal of ESIDA is to bring together researchers working in areas that span across multiple facets of exploratory search and data analytics to discuss and outline research challenges for this novel area.

Author Keywords

exploratory search, interactive search, data analytics, personalisation

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

BACKGROUND

While retrieval techniques operating on text or semantic annotations have become an industry standard, traditional search by keyword query becomes cumbersome for other forms of data (e.g. images, video, music) and even for textual data in the case of ambiguous queries (e.g. *bank* referring to side of a river or a financial institution). Also, in complex scenarios, or when the path from data to decision is not clear, exploratory search is necessary to understand and seek information interactively. Therefore, by actively engaging the user in the information retrieval loop, the user can explore a given dataset more easily as well as gradually direct their search to a more specific area of the search space.

Despite an increasing interest in exploratory search and interactive data analytics [6], several research questions still remain open, e.g.:

- What is the scope of exploratory search [10]?
- How to design systems that can support both traditional lookup search and exploratory search [2] as well as support personalisation [9]?
- How machine learning methods can be complemented with user domain knowledge to improve information finding or sense making [4]?
- How can exploratory search and interactive analytics be used in specialised domains, such as bioinformatics [13], mobile network data [7], or social media [5]?

WORKSHOP OVERVIEW

The aim of the workshop was to investigate different aspects of the design and evaluation of interactive search and data analytics algorithms and systems that tightly couple interactive visualization with analytics of the content of the data sets. The workshop included two keynotes, four regular papers and one short paper.

The first keynote, by Giulio Jacucci (University of Helsinki, Finland), was titled "Entity Based User Interfaces for Resourceful Information Exploration". The talk proposed a new approach characterised by entity modelling of information, conversational and synergic human-computer interaction, combining search and recommendation, and mapping relationships and representations of intent, queries and results. To demonstrate these features the talk discussed recent systems developed by the speaker's team including an interactive model of intent represented by keywords used to direct exploratory search, a system utilising entity based parallel search streams, and an interactive map visualisation for multi-aspect information retrieval. The talk summarised the principles embodied in the presented system towards a framework that considers principle of how to employ entities in information exploration.

The second keynote, by Shlomo Berkovsky (CSIRO, Australia), was titled "Improving interactive information discovery". The talk presented computational ways for improving the interaction of users with the recommendation lists or Web search results, aiming to devise a method that simplifies and shortens information discovery for users. One practical application of a generalised linear search (combination of the established linear and generalised search approaches) was presented, evaluated both with offline datasets and in an online

user study. It was discussed how similar interactive information discovery approaches can be applied to scenarios beyond recommender systems and Web search.

Sense making of document collections was the subject of two papers. Hierarchically organized social media discussions were visualized as scatter plots via dimensionality reduction [11]. A variety of interaction techniques was demonstrated to help reveal semantic relationships between posts that are not associated with the given hierarchy. Interactive topic models were used for highlighting key points of discussion in online petition data [3]. Content curation of the contents of a social Q&A system, by performing interactive cluster analysis to cluster together near-duplicate entries to improve search relevance was the focus of one paper [12]. Detection and classification of sentiment, aiming to identify outliers and fake news in news corpora is addressed in [8], using a Recursive Neural Network, in particular Long Short Term Memory (LSTM) network. An interactive query interface supported an empirical study in which users classify sentiment on name entities. User input is compared with the result obtained from the LSTM network. The last two papers focus on human annotation of documents to train machine learning systems. A user interface facilitates and optimizes the interactive steps of document presentation, inspection and labelling of publications, and supports the incorporation of domain-specific features (terminology gazetteers and document metadata), in [1].

REFERENCES

1. C. Aguirre, S. Coen, M. F. D. L. Torre, W. Hsu, and M. Rys. Towards faster annotation interfaces for learning to filter in information extraction and search. In *ESIDA*, 2018.

2. K. Athukorala, A. Medlar, A. Oulasvirta, G. Jacucci, and D. Glowacka. Beyond relevance: Adapting exploration/exploitation in information retrieval. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, pages 359–369, New York, NY, USA, 2016. ACM.

3. G. Cai and F. Sun. Interactive visualization for topic model curation. In *ESIDA*, 2018.

4. D. B. Coimbra, R. M. Martins, T. T. Neves, A. C. Telea, and F. V. Paulovich. Explaining three-dimensional dimensionality reduction plots. *Information Visualization*, 15(2):154–172, 2016.

5. A. Dang, A. Moh'd, A. Gruzd, E. Milios, and R. Minghim. An offline–online visual framework for clustering memes in social media. In *From Social Data Mining and Analysis to Prediction and Community Detection*, pages 1–29. Springer, 2017.

6. D. Glowacka, E. Milios, A. J. Soto, and F. Paulovich. Exploratory search and interactive data analytics. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion*, IUI '17 Companion, pages 9–11, New York, NY, USA, 2017. ACM.

7. E. Graells-Garrido, D. Caro, and D. Parra. Toward finding latent cities with non-negative matrix factorization. *arXiv preprint arXiv:1801.09093*, 2018.

8. C. G. Harris. Searching for diverse perspectives in news articles: Using an lstm network to classify sentiment. In *ESIDA*, 2018.

9. A. Medlar, J. Pyykkö, and D. Glowacka. Towards fine-grained adaptation of exploration/exploitation in information retrieval. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces*, IUI '17, pages 623–627, New York, NY, USA, 2017. ACM.

10. E. Palagi, F. Gandon, A. Giboin, and R. Troncy. A survey of definitions and models of exploratory search. In *Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, ESIDA '17, pages 3–8, New York, NY, USA, 2017. ACM.

11. J. Peltonen, Z. Lin, K. Järvelin, and J. Nummenmaa. PIHVI: Online forum posting analysis with interactive hierarchical visualization. In *ESIDA*, 2018.

12. I. Podgorny and C. Gielow. Semi-automated prevention and curation of duplicate content in social support systems. In *ESIDA*, 2018.

13. A. J. Soto, C. Zerva, R. Batista-Navarro, and S. Ananiadou. Litpathexplorer: a confidence-based visual text analytics tool for exploring literature-enriched pathway models. *Bioinformatics*, 1:9, 2017.