Example-based Search:  
a New Frontier for Exploratory Search

Extended Abstract

Matteo Lissandrini  
Aalborg University  
matteo@cs.aau.dk

Davide Mottin  
Aarhus University  
davide@cs.au.dk

Themis Palpanas  
Paris Descartes University  
themis@mi.parisdescartes.fr

Yannis Velegrakis  
Utrecht University  
i.velegrakis@uu.nl

ABSTRACT
Exploration is one of the primordial ways to accrue knowledge about the world and its nature. As we accumulate, mostly automatically, data at unprecedented volumes and speed, our datasets have become complex and hard to understand. In this context exploratory search provides a handy tool for progressively gather the necessary knowledge by starting from a tentative query that hopefully leads to answers at least partially relevant and that can provide cues about the next queries to issue. An exploratory query should be simple enough to avoid complicate declarative languages (such as SQL) and mechanisms, and at the same time retain the flexibility and expressiveness required to express complex information needs. Recently, we have witnessed a rediscovery of the so called example-based methods, in which the user, or the analyst circumvent query languages by using examples as input. This shift in semantics has led to a number of methods receiving as query a set of example members of the answer set. The search system then infers the entire answer set based on the given examples and any additional information provided by the underlying database. In this tutorial, we present an excursus over the main example-based methods methods for exploratory analysis. We show how different data types require different techniques, and present algorithms that are specifically designed for relational, textual, and graph data. We conclude by providing a unifying view of this query-paradigm and identify new exciting research directions.

1 MOTIVATION
Exploratory search includes methods to efficiently extract knowledge from data repositories, even if we do not know what exactly we are looking for, nor how to precisely describe our needs [46]. The need for new and effective exploratory search methods is particularly relevant given the current abundance and richness of today’s large datasets. In common exploratory settings, the user progressively acquires the knowledge by issuing a sequence of generic queries to gather intelligence about the data. However, the existing body of work in data analysis, data visualization, and predictive models, assumes the user is willing to pose several well defined or structured queries to the underlying database in order to progressively gather the required information. This assumption stems from the intuition that the user, being accustomed to data analysis, can more intuitively dig into the data. Yet, very often, this assumption is not true.

Recently, examples became a popular proxy for data exploration. Examples avoid the need for complex query languages. One of the earliest attempts to bring examples as a query method is query-by-example [51]. The main idea was to help the user in the query formulation, allowing them to specify the results in terms of templates for tuples, i.e., examples. Nowadays, examples are not anymore a mere template for relational queries, but rather the representative of the intended results the user would like to have. These example-based approaches are fundamentally different from the initial query-by-example idea, and are successfully applied not only to relational data [10, 36, 42], but also to textual [6, 48, 50], and graph [11, 16, 25] data as well.

We note that the flexibility of examples does not compromise the richness of the results, yet, it can overcome the ambiguity of generic keyword searches, which are frequently found in information retrieval. On the other hand, while data exploration techniques assume the user is willing to pose
We survey the main approaches for exploratory queries, highlighting the main differences among data models, and presenting in-depth insights of the current status of research in this area. The final goal is to provide a comprehensive overview of novel data-management techniques that can empower advanced exploratory search systems.

The first part of the tutorial introduces the broad topic of data exploration, highlighting the hardness of query languages for simple users and advocating the need of different query methods. We will introduce the example-based methods as flexible delegates for more complex queries that would otherwise need to be expressed through a very complex traditional query. In this part, we will discuss various cases, where queries cannot be expressed in declarative languages without requiring complex constructs. We will also present an expressive formulation of example-based approaches as seeking a similarity among objects.

The second part of the tutorial discusses the current main techniques for textual, and graph data, with an excursus on relational data as well in order to provide a complete picture on the power of the approach. In this part, we will present the algorithms, show how they work, and demonstrate their ability to (conceptually) solve complex search tasks (e.g., goal oriented search, focused community retrieval, graph search) from simple examples. We will also highlight the differences among data models, focusing on the scalability perspective, presenting the motivations and drawing parallels among methods for different data types.

The third part of the tutorial focuses on the latest developments of machine learning to progressively discover user intention. We will introduce the general area of online learning, some early methods based on relevance feedback [15], and show some recent applications of multi-armed bandits theories, that include active search.

Challenges and open research questions. The last part of the tutorial is dedicated to the challenges and open research questions. Exploratory search based on examples is rapidly attracting attention and getting traction, though, the support for such techniques in modern search and data management systems is lagging behind. Some challenges have already been discussed in recent vision papers [44, 47].

Finally, we will conclude the tutorial with remarks about the current state of affairs, and engage the audience in a discussion about their experiences with needs, tools, and challenges in this area.

3 RELEVANCE

The topic of exploration has been of interest in the IR community for many years now [46]. Exploratory search involves the study of information retrieval paradigms that move the process beyond predictable fact retrieval. [23]. This tutorial represents a bridge research in data-management and information-retrieval. In particular, in this tutorial, we will show how to combine results from research areas that already prominent in the IR community (e.g., knowledge graphs and machine learning) to novel techniques based on example driven query paradigms from the data management word to the benefit of enabling user-friendly exploratory search system. Past tutorials that cover relevant topics are for instance, “Utilizing Knowledge Graphs in Text-centric Information Retrieval” [9] by Dietz et al., presented at SIGIR 2018 (and earlier at WSDM 2017); “Graph Exploration: Let me Show what is Relevant in your Graph” [27] by Mottin and Müller at KDD 2018; and “Information Discovery in E-commerce” [32] by Ren et al. at SIGIR 2018. Yet, none of them focus on the topic of exploratory search in general, nor they cover example-driven query paradigms.

In contrast, this tutorial builds upon our tutorial “New Trends on Exploratory Methods for Data Analytics” presented at VLDB 2017 [26] and has been expanded with the material from our on “Data Exploration using Example-based Methods” [19], and will introduce the audience to these novel methods to empower data exploration.

4 TUTORIAL OUTLINE

In this tutorial1, we will provide a detailed overview of the new area of example-based methods for exploratory search, surveying the relevant state-of-the-art techniques. We will detail the overall problem formulation and taxonomy of methods, related to the questions they answer. Moreover, we will present future directions discussing various machine

1Slides of the tutorial will be available at https://data-exploration.ml
learning techniques used to infer user preferences in an online fashion. Finally we will give time to the audience to sustain a debate on future developments.

Next, we report the summary of the outline.

I. Introduction, motivation, and formulation
- Why example-based approaches are important
  - Usefulness of exploratory analysis
  - Main characteristics of exploratory analysis
  - Example-based methods for exploratory analysis
  - Use cases of failing keyword and declarative queries
  - Applications in current database systems and data analysis
- Connection to data exploration
- Problem formulation as similarity discovery

II. The origin: Example-based approaches for structured data
- Query-by-example: [51]
- Example methods in relational databases:
  - Reverse engineering of SQL queries [18, 28, 31, 36, 41, 42, 45, 49];
  - Schema mapping [1, 5, 12];
  - Data cleaning: entity matching [38], data repairing [14];
  - Exploratory Analytics [7, 34, 35].

III. Example-based approaches for semi-structured and unstructured data
- Example methods in textual data:
  - Exploring Web documents as examples [6, 50];
  - Example based Entity and Relation extraction [13, 37];
  - Web table search and augmentation [48];
  - Goal oriented content discovery [29];
- Example methods in graphs:
  - Cluster and Community exploration by Example Nodes [11, 17, 30, 33];
  - Entity Search [24, 39];
  - Reverse Engineering Path Queries [4] and SPARQL queries [2, 8] from Examples;
  - Example-based Knowledge Graph search [16, 20, 25].

IV. Learning methods based on examples
- Passive similarity learning: MindReader [15]
- Active learning:
  - Multi-armed bandits and the Upper Confidence Bound algorithm [3]
  - Gaussian processes and GP-Select [43]
  - Relevance feedback learning [10] and for graphs [22, 40]

V. Challenges and Discussion
- Can we interactively assist the user toward the retrieval of the correct answer?
- Can we provide explanations for the query results?
- How can machine learning help in exploratory analysis?
- Can we easily integrate these techniques into existing IR systems?

REFERENCES

[16] Nandish Jayaram, Arijit Khan, Chengkai Li, Xifeng Yan, and Ramez Elmasri. 2015. Querying knowledge graphs by example entity tuples. TKDE 27, 10 (2015), 2797–2811.
[17] Isabel M Kloumann and Jon M Kleinberg. 2014. Community membership identification from small seed sets. In KDD.