Example-based Exploration: Exploring Knowledge through Examples

Matteo Lissandrini  
Aalborg University  
matteo@cs.aau.dk

Davide Mottin  
Aarhus University  
davide@cs.au.dk

Themis Palpanas  
University of Paris  
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 to progressively gather the necessary knowledge by starting from a tentative query 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 or SPARQL) and convoluted mechanism, 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 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 [49]. The need for new and effective exploratory search methods is particularly relevant given the current abundance and richness of today’s large datasets (e.g., Linked Open Datasets). In common exploratory settings, the user progressively acquires the knowledge by issuing a sequence of generic queries to gather intelligence about the data. The existing body of work in data analysis assumes the user is willing to pose several well defined or structured queries to the underlying database to progressively gather the required information. This assumption stems from the intuition that the user is accustomed to data analysis techniques. Yet, this assumption is not always true.

Recently, examples became a popular proxy for data exploration. Examples avoid the need for complex query languages (e.g., SQL or SPARQL). One of the earliest attempts to bring examples as a query method is query-by-example (QBE) [54]. The main idea was to help the user in the query formulation, allowing them to specify the results in terms of templates for tuples. This idea, though, was particularly bounded to the relational model. 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 QBE, and are successfully applied not only to relational data [11, 48], but also to textual [7, 51], and graph [9, 16, 28] data as well, with several applications to semantic web data.

The flexibility examples provide 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 several exploratory queries, the use of examples allows the searcher to provide more information with less effort, making example-based methods a more palatable choice for novice users, as well as for practitioners. This new functionality can empower existing search systems with a complementary tool: whenever a query is too complex to be expressed with a detailed set of conditions, examples represent a natural alternative. In this respect (see Figure 1) example-based exploration is a middle ground between the user interface and the data-management layer, enabling new functionalities for the former and allowing more natural exploitation of the latter. Moreover, the use of examples has been demonstrated to be very effective in visual query interfaces [21]. We describe how example-based methods can be employed as an expressive and powerful method for exploratory search systems that naturally enhance user capabilities in accessing information from web documents, knowledge graphs, and other datasets.

2 OBJECTIVES

We survey the main approaches for exploratory queries, highlighting the main differences among data models, and presenting in-depth insights into 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. In particular, we will highlight the existing example-
based methods that have been already studied to improve knowl-
edge graph search, SPARQL query formulation, and data explo-
ration of RDF data. Moreover, we aim to present techniques that
have been studied in other research areas and that could be suc-
cessfully applied in the Semantic Web domain.

The first and second part of the tutorial introduce the broad
topic of data exploration, highlighting the hardness of query lan-
guages for simple users and advocating the need for different query
methods. We will introduce the example-based methods as flexi-
bile delegates for more complex search tasks that would otherwise
need to be expressed through very complex traditional queries. In
this part, we will discuss various cases where queries cannot be
expressed in declarative languages without requiring complex con-
structs. We will also present an expressive formulation of example-
based approaches as seeking a similarity among objects. This part
will also introduce the concept of query reverse engineering and
the original solution for reverse engineering of SQL queries from
examples and more advanced techniques. These concepts will pro-
vide the necessary background to understand similar solutions for
SPARQL queries that will be presented later.

The third part of the tutorial discusses the current main tech-
niques for textual and graph data to provide a complete picture of
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 techniques for text exploration include search approaches
based on documents used as representatives for the set of results [51],
and serendipitous search based on the current visited pages [7].
These approaches focus on documents as examples for retrieving
related information and are well versed to be expanded with addi-
tional information (e.g., from graphs and ontologies in the Semantic
Web). Recently, examples have been successfully employed in entity
extraction [14, 41], in which the user provides either mentions of en-
tities in a text [14], or tuples and similarities among attributes [41],
and the system automatically returns extraction rules that can be
applied to the given dataset.

For graph data there are two prominent approaches: the first use
subgraphs, or partially specified structures as input examples [9, 16, 
20, 28], while the second focus on the vertices of the graph (usu-
ally entities [26]), which are used for making the selections [17, 34].
Among the existing approaches Exemplar Queries [27, 28] and
Graph Query by Example (GQBE) [16] use subgraph isomorphism
or structural similarities to identify structures related to the one
the user-provided. A different approach is the reverse engineering
of SPARQL queries [9] in which the input is a set of positive and
negative entity mentions in an RDF dataset. Examples can also be
employed for targeted analysis of networks, to discover communi-
ties [17], dense regions [12], or subspaces along with outliers [34].

Particular focus will be given on how example-based search can
exploit knowledge graphs to provide semantic search capabilities
for documents as well as on example-based methods for entity and
structure search within knowledge graphs.

The fourth part of the tutorial focuses on the latest develop-
ments 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 [24, 43].

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.
The need for understanding the semantics behind the user query
is also of paramount importance. Some challenges have already
been discussed in recent vision papers [4, 47, 50]. 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.

Tutorial 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 data management
   - Connection to data exploration
   - Problem formulation as similarity discovery

II. The origin: Example-based approaches for structured data
   - Query-by-example: [54]
   - Example methods in relational databases:
     - Reverse engineering of SQL queries [18, 32, 35, 40, 44, 
       45, 48, 52];
     - Schema mapping [1, 6, 13];
     - Exploratory Analytics [8, 38, 39].

III. Example-based approaches for semi-structured and un-
     structured data
   - Example methods in textual data:
     - Exploring Web documents as examples [7, 53];
     - Example based Entity and Relation extraction [14, 41];
     - Web table search and augmentation [51];
     - Goal oriented content discovery [33];
   - Example methods in graphs:
     - Cluster and Community exploration by Example Nodes [12, 
       17, 34, 37];
     - Semantic Entity Search [26, 42];
     - Reverse Engineering Path Queries [5] and SPARQL [2, 9]
       from Examples;
   - Example-based Knowledge Graph search [16, 20, 23, 28].

IV. Learning methods based on examples
   - Passive similarity learning: MindReader [15]
   - Active learning:
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 build a Personal Knowledge Assistant?

3 RELEVANCE

The topic of exploration has been of interest in many communities related to information retrieval, data management, and semantic web for many years now [9, 26, 49]. Exploratory search involves the study of information retrieval paradigms that move the process beyond predictable fact retrieval [25]. The large availability of knowledge graphs and open data provide both abundant resources and unique challenges to users aiming to find relevant information on the web. This tutorial represents a research bridge across data-management, information-retrieval, and semantic web techniques. In particular, this tutorial will show how to combine results from research areas that are already prominent in the semantic web community (e.g., search and retrieval, knowledge graphs, and machine learning) to novel techniques based on example-driven query paradigms from the data management world to the benefit of enabling user-friendly knowledge exploration. Past tutorials that cover relevant topics are for instance, “Utilizing Knowledge Graphs in Text-centric Information Retrieval” [10] by Dietz et al., presented at SIGIR 2018 (and earlier at WSDM 2017); “Graph Exploration: Let me Show what is Relevant in your Graph” [31] by Mottin and Müller at KDD 2018; and “Information Discovery in E-commerce” [36] by Ren et al. at SIGIR 2018. Yet, none of them focuses on the topic of exploratory search in general, nor they cover example-driven query paradigms in detail.

In contrast, this tutorial builds upon the earlier “New Trends on Exploratory Methods for Data Analytics” presented at VLDB 2017 [29] that has been expanded with the material from the book “Data Exploration using Example-based Methods” [19] for SIGMOD 2019 [30], as well as on the material of the tutorial on Exploratory Search presented at SIGIR 2019 [22] and will introduce the audience to these novel methods to empower data exploration for and with semantic web resources and knowledge graphs.

REFERENCES


