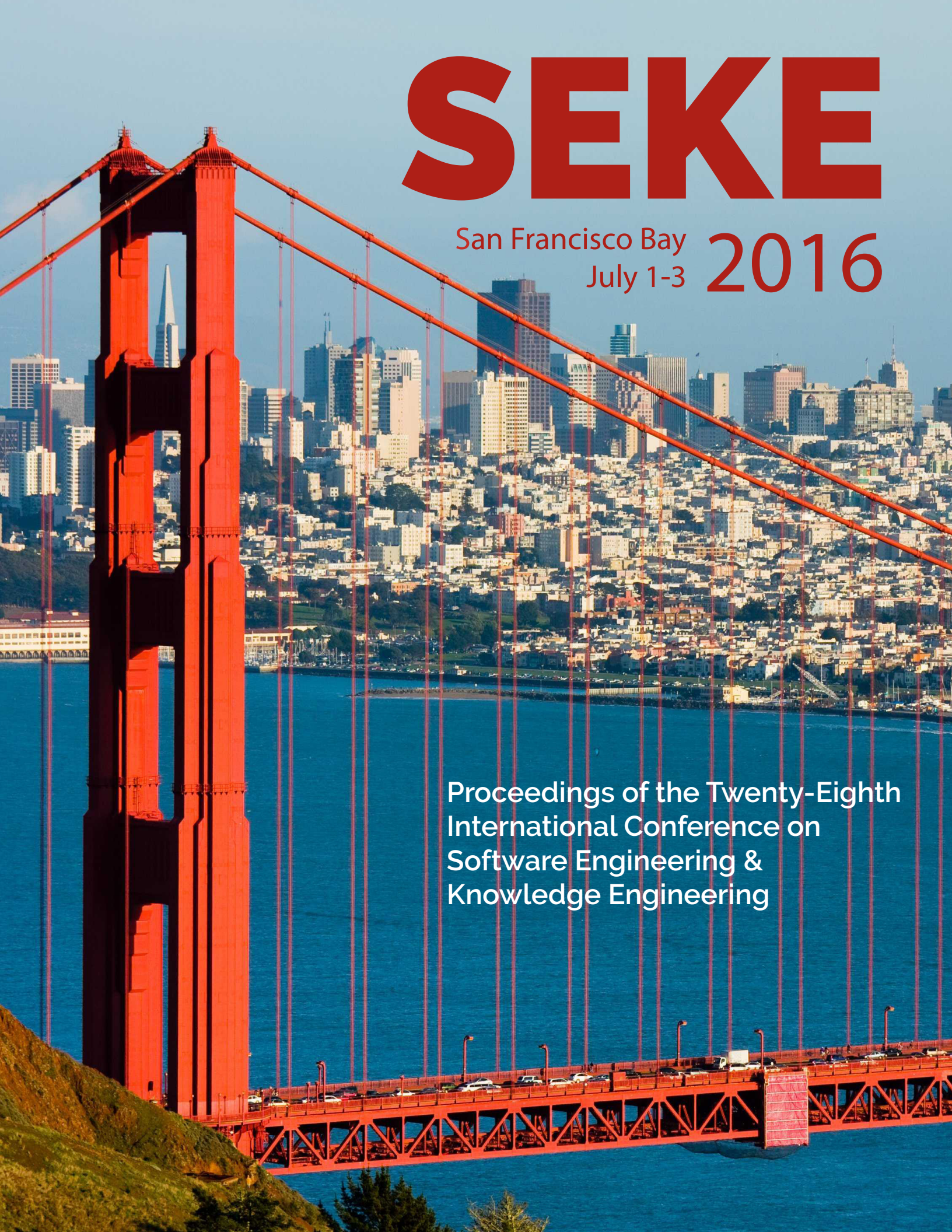

SEKE

San Francisco Bay
July 1-3 2016

Proceedings of the Twenty-Eighth International Conference on Software Engineering & Knowledge Engineering

Proceedings of the Twenty-Eighth International Conference on Software Engineering & Knowledge Engineering

SEKE

San Francisco Bay
July 1-3

2016

Printed by
KSI Research Inc. &
Knowledge Systems Institute
Graduate School
156 Park Square
Pittsburgh, PA 15238 USA
Tel: +1-412-606-5022
Fax: +1-847-679-3166
Email: seke@ksiresearch.org
Web: http://ksiresearchorg.ipage.com/seke/seke16.html
Printed in USA, 2016
ISBN 1-891706-39-X (paper)
ISSN 2325-9000 (print)

Software Engineering Methodology

Poster

Note:
(S) indicates a short paper.
(P) indicates a poster.

Towards a Systematic Approach to Graph Data Modeling: Scenario-based Design and Experiences

Mengjia ZHAO*, Yan LIU†, Peng ZHOU‡
School of Software Engineering, Tongji University
Shanghai, China
Email: *1434319@tongji.edu.cn, †yanliu.sse@tongji.edu.cn, ‡1435855@tongji.edu.cn

***Abstract*—Graph database is recently being adopted by data analytic systems as an appealing alternative to relational database for the management of large-scale inherent graph-like data. A great challenge of leveraging graph database technologies is to model a problem domain into graph. However, in the absence of considering application requirements or goals, current graph data modeling approaches seem to be invalid. This paper presents an exploration of a systematic approach for graph data modeling—SuMo. Starting from real world scenarios, requirements are transformed into a domain model, which acts as an intermediate model in SuMo, captures modeling features of that domain. SuMo defines a set of rules for the subsequent transformation of this domain model to produce a graph model. We applied SuMo to the modeling of a data-intensive analytic system using real datasets as an illustrating example to clarify our main idea. SuMo is empirically evaluated in terms of query performance, the experimental results indicate promising feasibility and efficiency. The major contribution of our work is a preliminary graph data modeling approach based on scenarios.**

Keywords-graph data modeling; data analytic system; scenario-based modeling; model-driven design

I. Introduction

Graph database has recently gained popularity rapidly because of a need to effectively manage large-scale inherent graph-like data [1]. Social network, biology, semantic web and health-care are typical application domains that contain such kind of data [2][3][4]. It has been observed that graph database is usually preferable to relational database in managing data of these domains since the latter hardly captures the inherent graph structure [5]. Motivated by scalability and performance needs of current applications, graph database is increasingly adopted by data analytic systems.

The analysis of relationships becomes important when dealing with highly connected data. A graph consists of a finite set of nodes, and a finite set of edges defining relationships between these nodes. In graph database, data is stored as graph and is accessed by queries as graph traversal operations. Compared with the stores of relational database, relationships are treated as first-class citizens in graph database [6], expressed in a more straightforward way. Moreover, queries involving complex and inefficient join operations are transformed into graph traversals. The execution time of graph traversal is proportional only to the size of traversed part [6].

DOI reference number: 10.18293/SEKE2016-119

Graph data modeling is the process in which an arbitrary application domain described as a connected graph of nodes and edges, it is not a context-free process, but a purposive abstraction related with application requirements. As it happens with relational database, the design of relational model can start from ER model, such kind of "springboard" is also conducive to the design of graph model.

Guidelines and principles on graph data modeling can be found from various sources, including books, technical reports, graph database online community and practitioners' blogs. There has also been an involvement of academic papers in the study of graph data modeling, many works focus on converting existing data from relational to graph model automatically; a number of researchers aim at generic graph data modeling approaches. However, most of these works, simply demonstrated some rudimentary strategies of graph data modeling, rather than connecting to practical application requirements. As a result, practitioners can only find some scattered modeling guidelines.

Currently, graph data modeling is still based on best practices and probably unproved guidelines, which are usually relevant to specific systems [5]. This paper aims to propose a systematic graph data modeling approach at a "proof of concept" stage. The primary challenge that come is, how to analyze requirements of application domain to support graph data modeling? Translating these requirements to produce a model brings more challenges.

In this paper, we are exploring a systematic graph data modeling approach, using scenarios to start out the modeling process. We suggest an adapted domain modeling approach with strategies to abstract key concepts from scenarios. A set of rules are defined for the transformation of domain model to graph model. Our aim is to match application requirements and facilitate the design of graph-based analytic systems. We also provide experiments, in terms of query performance, showing the advantages of our approach with respect to a naïve approach. The main contribution of this paper is a sketch of a systematic graph data modeling approach.

The rest of this paper is organized as follows, Section II presents related work. The proposed approach is presented in Section III. A case study is demonstrated in Section IV. We design and conduct a group of experiments to evaluate our approach, as well as discuss the results in Section V. The conclusion and future works are in Section VI.