

Preface

This volume features the extended versions of a selection of the best papers that were originally presented at the French-speaking conference EGC'2016 held in Reims (France) during 18–22 January, 2016. The papers have been accepted after a peer-review process among papers already accepted in long format at the conference. For the conference, the long papers are themselves the result of a double-blind peer-review process among the 94 papers initially submitted to the conference (conference acceptance rate for long papers of 24%). This conference was the 16th edition of this event, which takes place each year and which is now successful and well known in the French-speaking community. This community was structured in 2003 by the foundation of the International French-speaking EGC society (EGC in French stands for “Extraction et Gestion des Connaissances” and means “Knowledge Discovery and Management” or KDM). This society organizes every year its main conference (about 200 attendees) but also workshops and other events with the aim of promoting exchanges between researchers and companies concerned with KDM and its applications in business, administration, industry, or public organizations. For more details about the EGC society, please consult <http://www.egc.asso.fr>.

Structure of the Book

This book is a collection of six representative and novel works done in Data Mining and Knowledge Discovery. It is intended to be read by all researchers interested in these fields, including Ph.D. or M.Sc. students, and researchers from public or private laboratories. It concerns both theoretical and practical aspects of KDM.

Alec et al. propose in Chapter “[A Combined Approach for Ontology Enrichment from Textual and Open Data](#)” an approach for ontology enrichment for automatically labeling documents describing entities, with very specific concepts reflecting specific users’ needs. This new approach to ontology population and enrichment exploits the foundations of the Semantic Web by combining contributions of text

analysis, linked open data extraction, machine learning, and reasoning tools. An evaluation, on two application domains, provides quality results and demonstrates the interest of the approach.

Chapter “[C-SPARQL Extension for Sampling RDF Graphs Streams](#)”, by Fall Dia et al., presents an extension of the semantic data stream processing system C-SPARQL for sampling RDF graphs streams. The authors add new operators to C-SPARQL query syntax and propose a solution based on existing sampling algorithms (Uniform, Reservoir, and Chain) for reducing load while keeping data semantics. The evaluation shows an improved response time of query execution and a perfect preserving of data semantics.

Chapter “[Efficiency Analysis of ASP Encodings for Sequential Pattern Mining Tasks](#)”, written by Guyet et al., presents how the declarative paradigm of Answer Set Programming may be used to define the foundations for developing a flexible sequential pattern mining tool. It proposes encodings of the classical sequential pattern mining tasks within two representations of embeddings and for various kinds of patterns. The computational performance of these encodings is compared each other to get a good insight into the efficiency of ASP encodings and to a constraint programming approach.

In multi-writer, multi-reader systems, data consistency is ensured by the number of replica nodes contacted during read and write operations. Chapter “[Consistency-Latency Trade-Off of the LibRe Protocol: A Detailed Study](#)”, by Kumar et al., develops a consistency protocol called LibRe, which helps to read the latest version of a data item by contacting a minimum number of replica nodes. This protocol provides a new trade-off between consistency and latency for distributed data storage systems. The effectiveness of the approach in practice is shown by providing a proof-of-concept implementation of the protocol inside the Cassandra distributed data store.

Taleb et al. in Chapter “[A Fault Detection Approach for Robotic Systems Combining the Data Obtained from Sensor Measurements and Linear Observer-Based Estimations](#)” propose a new approach for fault detection in robotic systems which is performed by using classification algorithms. In the first stage, data are collected from the different sensors installed on a real system. A linear state observer is used to estimate non-measurable state variables in order to enrich the database. Then, a classifier is used to diagnose the faults occurring on the considered system. An experiment on a robotized actuated seat is considered to illustrate the proposed fault detection approach and demonstrate the benefits of adding estimations of unmeasured state variables to the classification process.

Chapter “[A Collaborative Framework for Joint Segmentation and Classification of Remote Sensing Images](#)” by Troya-Galvis et al. presents a collaborative framework for joint segmentation and classification which is quality-centric and attempts to simultaneously improve both segmentation and classification for a given thematic class by the interaction of these two paradigms. The framework is applied to vegetation extraction in a very high spatial resolution image of Strasbourg, France. The experiments show that the proposed method reaches good classification results while remarkably improving the segmentation results.

Advances in Knowledge Discovery and Management
Volume 7

Pinaud, B.; Guillet, F.; Cremilleux, B.; de Runz, C. (Eds.)
2018, XIII, 147 p. 46 illus., 35 illus. in color., Hardcover
ISBN: 978-3-319-65405-8