

Preface

With the increasing amount of available data, it becomes more and more important to have effective methods to manage data and retrieve information. Data processing, especially in the era of Social Media, is changing the users' behaviors. Users are ever more interested in information rather than in mere raw data. Considering that the large amount of accessible data sources is growing, novel systems providing effective means of searching and retrieving information are required. Therefore, the fundamental goal is making information exploitable by humans and machines.

This book collects invited chapters and extended research contributions from DART 2013 (the 7th International Workshop on Information Filtering and Retrieval), held in Turin (Italy), on December 6, 2013 and co-located with the XIII International Conference of the Italian Association for Artificial Intelligence.

The book is focused on researching and studying new challenges in intelligent information filtering and retrieval.

In Chapter "[On the Use of Matrix Bloom Filters in Data Leak Protection](#)," by Sergey Butakov, modifications to Matrix Bloom filters for data leak prevention systems are proposed. Bloom filters are fast tools for massive document comparison. The approach proposed in this chapter aims at improving density in matrix Bloom filters with the help of a special index to track documents uploaded into the system. Density is increased by combining a few documents in one line of the matrix to reduce the filter size and to address the problem of document removal. Special attention is paid to the negative impact of filter-to-filter comparison in matrix filters. A theoretical evaluation of the threshold for false positive results is described. The performed experiment outlines advantages and applicability of the proposed approach.

Chapter "[A Hybrid Approach for the Automatic Extraction of Causal Relations from Text](#)," by Antonio Sorgente, Giuseppe Vettigli, and Francesco Mele, proposes a hybrid approach for the discovery of causal relations from open domain text in English. The approach first identifies a set of plausible cause-effect pairs through

a set of logical rules based on dependencies between words, then it uses Bayesian inference to filter and reduce the number of pairs produced by ambiguous patterns. The proposed model has been evaluated in the context of the SemEval-2010 task 8 dataset challenge. The results confirm the effectiveness of the rules for relation extraction and the improvements made by the filtering process.

Chapter “[Ambient Intelligence by ATML Rules in BackHome](#),” by Eloi Casals, José Alejandro Cordero, Stefan Dauwalder, Juan Manuel Fernández, Marc Solà, Eloisa Vargiu, and Felip Miralles, focuses on an implementation of state-of-the-art ambient intelligence techniques implemented in the telemonitoring and home support system developed by BackHome. BackHome is a EU project that aims to study the transition from the hospital to the home for people with disabilities. The proposed solution adopts a rule-based approach to handle efficiently the system adaptation, personalization, alarm triggering, and control over the environment. Rules are written in a suitable formal language, defined within the project, and can be automatically generated by the system taking into account dynamic context conditions and user habits and preferences. Moreover, rules can be manually defined by therapists and caregivers to address specific needs. The chapter also presents the concept of “virtual device”, a software element that performs a mash-up of information coming from two or more sensors in order to make some inference and provide new information.

In Chapter “[Dense Semantic Graph and Its Application in Single Document Summarisation](#),” by Monika Joshi, Hui Wang, and Sally McClean, a novel graph generation approach is proposed. The proposal is an extension of an existing semantic graph generation approach. Semantic graph representation of text is an important part of natural language processing applications such as text summarisation. In particular, two ways of constructing the semantic graph of a document from dependency parsing of its sentences are studied. The first graph is derived from the subject–object–verb representation of sentence, and the second graph is derived from considering more dependency relations in the sentence by a shortest distance dependency path calculation, resulting in a dense semantic graph. Experiments show that dense semantic graphs perform better in semantic graph-based unsupervised extractive text summarization.

In Chapter “[Web Architecture of a Web Portal for Reliability Diagnosis of Bus Regularity](#),” by Benedetto Barabino, Cristian Lai, Roberto Demontis, Sara Mozzoni, Carlino Casari, Antonio Pintus, and Proto Tilocca, a methodology to evaluate bus regularity is proposed. In high-frequency transit services, bus regularity—i.e., the headway adherence between buses at bus stops—can be used as an indication of service quality in terms of reliability, by both users and transit agencies. The web architecture is the entry point of a Decision Support System (DSS), and offers an environment designed for experts in transport domain. The environment is composed of tools developed to automatically handle Automatic Vehicle Location (AVL) raw data for measuring the Level of Service (LoS) of bus regularity at each bus stop and time interval of a transit bus route. The results are represented within easy-to-read control dashboards consisting of tables, charts, and

maps, able to perform fast AVL processing and easy accessibility in order to reduce the workload of transit operators. These outcomes show the importance of well-handled and presented AVL data, in order to use them more effectively, improving past analysis done by using, if any, manual methods.

Chapter “[An Approach to Knowledge Formalization](#)”, by Filippo Eros Pani, Maria Ilaria Lunesu, Giulio Concas, and Gavina Baralla, is focused on the study of a process to formalize knowledge through an iterative approach mixing the top-down analysis for the structure of the knowledge of the specific domain and the bottom-up analysis for the information in the object of the specific domain. Just a few years after its birth, knowledge management drew the attention of the academic world, becoming a matter of intense study as the scientific community has always felt the need of making its members cooperate through the exchange and the reciprocal fruition of information. Many disciplines develop standardized formalization of knowledge, which domain experts can use to share information in the form of reusable knowledge. The proposed case study analyzes the knowledge on the domain of descriptions and reviews of Italian wines which can be found on the Internet: starting from an analysis of the information on the web, a taxonomy able to represent knowledge is defined.

Pula, Italy
Cagliari, Italy
Bari, Italy
October 2014

Cristian Lai
Alessandro Giuliani
Giovanni Semeraro

Emerging Ideas on Information Filtering and Retrieval

DART 2013: Revised and Invited Papers

Lai, C.; Giuliani, A.; Semeraro, G. (Eds.)

2018, XI, 104 p. 35 illus., 23 illus. in color., Hardcover

ISBN: 978-3-319-68390-4