1.

Record Nr.	UNINA9910299161703321
Autore	Zhang Wei Emma
Titolo	Managing Data From Knowledge Bases: Querying and Extraction / / by Wei Emma Zhang, Quan Z. Sheng
Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2018
ISBN	9783319949352
	3319949357
Edizione	[1st ed. 2018.]
Descrizione fisica	1 online resource (XIII, 139 p. 41 illus., 32 illus. in color.)
Disciplina	006.312
Soggetti	Data mining
	Information storage and retrieval systems
	Application software
	Data Mining and Knowledge Discovery
	Information Storage and Retrieval
	Computer and Information Systems Applications
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references.
Nota di contenuto	1 Introduction -- 2 Cache Based Optimization for Querying Curated Knowledge Bases -- 3 Query Performance Prediction on Knowledge Base -- 4 An Efficient Knowledge Clustering Algorithm -- 5 Knowledge Extraction from Unstructured Data on the Web -- 6 Building Knowledge Bases from Unstructured Data on the Web -- 7 Conclusion.
Sommario/riassunto	In this book, the authors first address the research issues by providing a motivating scenario, followed by the exploration of the principles and techniques of the challenging topics. Then they solve the raised research issues by developing a series of methodologies. More specifically, the authors study the query optimization and tackle the query performance prediction for knowledge retrieval. They also handle unstructured data processing, data clustering for knowledge extraction. To optimize the queries issued through interfaces against knowledge bases, the authors propose a cache-based optimization layer between consumers and the querying interface to facilitate the querying and solve the latency issue. The cache depends on a novel learning method

that considers the querying patterns from individual's historical queries without having knowledge of the backing systems of the knowledge base. To predict the query performance for appropriate query scheduling, the authors examine the queries' structural and syntactical features and apply multiple widely adopted prediction models. Their feature modelling approach eschews the knowledge requirement on both the querying languages and system. To extract knowledge from unstructured Web sources, the authors examine two kinds of Web sources containing unstructured data: the source code from Web repositories and the posts in programming question-answering communities. They use natural language processing techniques to pre-process the source codes and obtain the natural language elements. Then they apply traditional knowledge extraction techniques to extract knowledge. For the data from programming question-answering communities, the authors make the attempt towards building programming knowledge base by starting with paraphrase identification problems and develop novel features to accurately identify duplicate posts. For domain specific knowledge extraction, the authors propose to use a clustering technique toseparate knowledge into different groups. They focus on developing a new clustering algorithm that uses manifold constraints in the optimization task and achieves fast and accurate performance. For each model and approach presented in this dissertation, the authors have conducted extensive experiments to evaluate it using either public dataset or synthetic data they generated.