

1. Record Nr.	UNINA9910881098903321
Autore	Strauss Christine
Titolo	Database and Expert Systems Applications : 35th International Conference, DEXA 2024, Naples, Italy, August 26–28, 2024, Proceedings, Part I // edited by Christine Strauss, Toshiyuki Amagasa, Giuseppe Manco, Gabriele Kotsis, A Min Tjoa, Ismail Khalil
Pubbl/distr/stampa	Cham : , : Springer Nature Switzerland : , : Imprint : Springer, , 2024
ISBN	9783031683091 9783031683084
Edizione	[1st ed. 2024.]
Descrizione fisica	1 online resource (289 pages)
Collana	Lecture Notes in Computer Science, , 1611-3349 ; ; 14910
Altri autori (Persone)	AmagasaToshiyuki MancoGiuseppe KotsisGabriele TjoaA. Min KhalilIsmail
Disciplina	005.74
Soggetti	Database management Artificial intelligence Information technology - Management Software engineering Information storage and retrieval systems Data mining Database Management Artificial Intelligence Computer Application in Administrative Data Processing Software Engineering Information Storage and Retrieval Data Mining and Knowledge Discovery
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Organization -- Abstracts of Keynote Talks -- Multimodal Deep Learning in Medical Imaging -- Digital Humanism as an Enabler for a Holistic Socio-Technical Approach to the Latest

Developments in Computer Science and Artificial Intelligence -- Deep Entity Processing in the Era of Large Language Models: Challenges and Opportunities -- Contents - Part I -- Contents - Part II -- Financial and Economic Data Analysis -- CSPRD: A Financial Policy Retrieval Dataset for Chinese Stock Market -- 1 Introduction -- 2 Related Work -- 2.1 Retrieval Augmented Generation -- 2.2 Dense Retrievers -- 2.3 Specialised Financial Datasets -- 3 The Policy Retrieval Dataset for Stock Market in China -- 3.1 Data Collection -- 3.2 Data Processing -- 3.3 Unsupervised MoE Selection -- 3.4 Expert Annotation -- 3.5 Dataset Release -- 4 CSPR-MQA Pre-training -- 4.1 Data Preprocessing -- 4.2 Encoding -- 4.3 Decoding -- 5 Experiments -- 5.1 Models -- 5.2 Evaluation Metrics -- 5.3 Results and Analysis -- 6 Limitations -- 6.1 Compromise Between Labor Costs and Decision Comprehensiveness -- 6.2 Limited Experiments on Pre-Trained Language Models -- 7 Conclusion -- References -- Leveraging Heterogeneous Text Data for Reinforcement Learning-Based Stock Trading Strategies -- 1 Introduction -- 2 Related Work -- 3 Preliminaries -- 3.1 Terms -- 3.2 Problem Definition -- 3.3 Base Model -- 4 Proposed Method -- 4.1 Encording -- 4.2 Feature Generation -- 4.3 Action Determination -- 5 Experiment -- 5.1 Datasets -- 5.2 Experimental Setup -- 5.3 Experimental Results -- 5.4 Comparison with Index-Based Method -- 5.5 Effects of Periods and Initial Assets -- 6 Conclusion -- References -- TCMIDP: A Comprehensive Database of Traditional Chinese Medicine for Network Pharmacology Research -- 1 Introduction -- 2 Database Contents and Access -- 3 Data Mining -- 4 User Evaluation. -- 5 Conclusion -- References -- Graph Theory and Network Analysis -- Fast Subgraph Search with Graph Code Indices -- 1 Introduction -- 2 Preliminaries -- 3 Related Work -- 4 Basic Concept of the Proposed Method -- 5 Graph Representation and Indexing of Databases -- 6 Subgraph Search with the Code Tree -- 7 Experimental Evaluation -- 7.1 Experimental Settings -- 7.2 Experimental Results -- 8 Conclusion -- References -- Completing Predicates Based on Alignment Rules from Knowledge Graphs -- 1 Introduction -- 2 Motivating Example -- 3 The SYRUP Approach -- 4 Experimental Study -- 5 Related Work -- 6 Conclusions and Future Work -- References -- Enriching Hierarchical Navigable Small World Searches with Result Diversification -- 1 Introduction -- 2 Preliminaries and Related Work -- 3 Material and Methods -- 4 Empirical Evaluation -- 5 Conclusions -- References -- An Efficient Indexing Method for Dynamic Graph kNN -- 1 Introduction -- 1.1 Existing Approaches and Challenges -- 1.2 Our Approaches and Contributions -- 2 Preliminary -- 2.1 Problem Definition -- 2.2 Previous Method: CT Index -- 3 Proposed Method: Dynamic CT -- 3.1 Ideas -- 3.2 Adding Nodes and Edges -- 3.3 Removing Nodes and Edges -- 3.4 Complexity Analysis -- 4 Experimental Evaluation -- 4.1 Efficiency for Updating -- 4.2 Efficiency for Adding/Removing Edges -- 5 Conclusion -- References -- Database Management and Query Optimization -- Improving the Accuracy of Text-to-SQL Tools Based on Large Language Models for Real-World Relational Databases -- 1 Introduction -- 2 Related Work -- 3 A Real-World Benchmark for the Text-to-SQL Task -- 3.1 The Real-World Relational Database -- 3.2 The Sets of Views -- 3.3 The Test Questions and Their Ground Truth SQL Translations -- 4 The Proposed RAG-Based Technique -- 4.1 Generation of the Synthetic Dataset -- 4.2 The Proposed RAG-Based Techniques. -- 5 Experiments with an RW-RDB -- 5.1 Experimental Setup -- 5.2 Results -- 6 Experiments with Mondial -- 7 Conclusions -- References -- QPSEncoder: A Database Workload Encoder with Deep Learning -- 1

Introduction -- 2 Related Work -- 3 QPSEncoder Framework -- 3.1 Physical Plan Encoding -- 3.2 SQL Query Encoding -- 3.3 Database Schema Encoding -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Effectiveness of QPSEncoder on Numeric Predicates. -- 4.3 Effectiveness of QPSEncoder on Mixed Predicates. -- 4.4 Effectiveness of Model Components -- 5 Conclusion -- References -- Efficient Random Sampling from Very Large Databases -- 1 Introduction -- 2 Related Work -- 3 The Proposed Algorithms -- 3.1 Random Sampling in B+Tree of Height Three -- 3.2 Random Sampling in B+Tree of Height Four -- 3.3 Generalization to Any B+Tree Height -- 4 Analysis -- 5 Simulation Study -- 5.1 Experiments Framework -- 5.2 Experiments Setup -- 5.3 Implementation Details -- 5.4 Results -- 6 Summary and Future Work -- References -- SQL-to-Schema Enhances Schema Linking in Text-to-SQL -- 1 Introduction -- 2 Related Work -- 2.1 Customized Machine Learning Fine-Tuned Methods -- 2.2 Stimulating General LLM with Prompting -- 3 Methodology -- 3.1 Evaluation Metrics -- 3.2 Introduction to Each Module -- 4 Experiments and Analysis -- 4.1 Experiment One -- 4.2 Experiment Two -- 4.3 Experiment Three -- 5 Conclusion -- References -- Efficient Algorithms for Top-k Stabbing Queries on Weighted Interval Data -- 1 Introduction -- 2 Preliminary -- 3 Algorithm Based on Interval Forest -- 3.1 Data Structure and Construction -- 3.2 Query Processing Algorithm -- 4 Algorithm Based on a Variant of Segment Tree -- 4.1 Variant of Segment Tree and Its Construction -- 4.2 Query Processing Algorithm -- 5 Conclusion -- References -- A Hierarchical Storage Mechanism for Hot and Cold Data Based on Temperature Model. 1 Introduction -- 2 Related Work -- 3 System Architecture -- 3.1 Data Temperature Model -- 3.2 Data Hierarchical Storage Mechanism -- 4 Experimentation and Analysis -- 4.1 Experimental Environment and Configuration -- 4.2 Experiment on Local Hot and Cold Data Migration Management -- 4.3 Experiment on Local and Remote Data Migration Management -- 5 Conclusion -- References -- Machine Learning and Large Language Models -- A Pre-trained Knowledge Tracing Model with Limited Data -- 1 Introduction -- 2 Related Work -- 3 Method -- 3.1 Transformer-Based Knowledge Tracing Model -- 3.2 BERT-Based Knowledge Tracing Model -- 4 Experiments -- 4.1 Datasets -- 4.2 Experimental Setup -- 4.3 Experiment Results and Analysis -- 5 Conclusions and Future Work -- References -- Chorus: More Efficient Machine Learning on Serverless Platform -- 1 Introduction -- 2 Background and Motivation -- 3 Lambda Synchronous Parallel (LSP) Model -- 4 Buffering in Parameter Server Model -- 5 Design of Chorus -- 5.1 Architecture Overview -- 6 Evaluation -- 6.1 Methodology -- 6.2 Lambda Synchronous Parallel (LSP) Model -- 6.3 Buffering System -- 6.4 Comparison -- 7 Conclusion -- References -- Evaluating Performance of LLaMA2 Large Language Model Enhanced by QLoRA Fine-Tuning for English Grammatical Error Correction -- 1 Introduction -- 2 Related Work -- 2.1 Grammar Error Correction -- 2.2 LLaMA2 Large Language Model -- 3 Methodology -- 3.1 Data Preparation -- 3.2 Prompts and Learning Strategies -- 3.3 Parameter Efficient Fine-Tuning via QLoRA -- 4 Experiments -- 4.1 Text Generation Settings -- 4.2 Parameter Settings of Training -- 4.3 Experimental Results -- 4.4 Error Type Analysis -- 4.5 Performance Evaluation of Open Source Language Tools -- 5 Conclusion -- References -- A Label Embedding Algorithm Based on Maximizing Normalized Cross-Covariance Operator -- 1 Introduction. 2 Preliminaries -- 3 The Proposed Method -- 4 Experiments -- 4.1 Benchmark Data Sets and Evaluation Metrics -- 4.2 Compared Methods and Experimental Settings -- 4.3 Performance Evaluation and Analysis

-- 5 Conclusion -- References -- Analyzing the Efficacy of Large Language Models: A Comparative Study -- 1 Introduction -- 2 Literature Review -- 3 Datasets and Preprocessing -- 3.1 Dataset Construction by Question-Answer Pair Generation -- 3.2 Fine-Tuning the LLM -- 3.3 Document Parsing and Text Comprehension -- 4 Methodology -- 4.1 Integrated Evaluation: BLEU, ROUGE, and Cosine Similarity -- 4.2 Accuracy Assessment Through Z-Score Outlier Detection -- 5 Observations and Results -- 5.1 Classification of Errors -- 5.2 Comparison of Performance of Our Framework -- 6 Conclusion and Future Directions -- References -- Leveraging Large Language Models for Flexible and Robust Table-to-Text Generation -- 1 Introduction -- 2 Related Work -- 3 Methods and Experiment Settings -- 4 Experiments Results -- 5 Conclusion -- References -- Recommender Systems and Personalization -- Collaborative Filtering for the Imputation of Patient Reported Outcomes -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 MDASI-HN Data -- 3.2 Collaborative Filtering (CF) for MDASI-HN -- 4 Evaluation -- 4.1 Evaluation Metrics: -- 5 Experimental Results -- 5.1 Experimental Setup -- 5.2 Data Statistics -- 5.3 CF Techniques Comparison -- 5.4 Effect of k on CF-SYM-PCC Imputation -- 5.5 Comparing CF-SYM-PCC Against Other Methods -- 5.6 Comparing CF-SYM-PCC and LI Techniques Per Symptom -- 5.7 PCC Correlation Symptom Clusters -- 6 Conclusion -- References -- Category-Aware Sequential Recommendation with Time Intervals of Purchases -- 1 Introduction -- 2 Method -- 2.1 Problem Setup -- 2.2 Category and Time Interval Aware Sequence Recommendation -- 2.3 Framework of the Dual Model. 3 Performance Evaluation.

---

## Sommario/riassunto

The two-volume set LNCS 14910 and 14911 constitutes the proceedings of the 35th International Conference on Database and Expert Systems Applications, DEXA 2024, which took place in Naples, Italy, in August 2024. The 27 full and 20 short papers included in the proceedings set were carefully reviewed and selected from 102 submissions. They were organized in topical sections as follows: Part I: Financial and economic data analysis; graph theory and network analysis; database management and query optimization; machine learning and large language models; recommender systems and personalization; Part II: Blockchain and supply management; data mining and knowledge discovery; spatiotemporal data and mobility analysis; computer vision and image processing; data security and privacy; database indexing and query processing; specialized applications and case studies. .

---