1. Record Nr. UNISA996490363503316

Titolo	Linking theory and practice of digital libraries : 26th International Conference on Theory and Practice of Digital Libraries, TPDL 2022, Padua, Italy, September 20-23, 2022, Proceedings / / edited by Gianmaria Silvello, [and six others]
Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2022] ©2022
ISBN	3-031-16802-X
Descrizione fisica	1 online resource (565 pages)
Collana	Lecture Notes in Computer Science ; ; v.13541
Disciplina	381
Soggetti	Digital libraries
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Intro -- Preface -- Organization -- Democratizing Data Access: What if We Could Just Talk to Our Data? -- Contents -- Keynote Talks -- Should We Preserve the World's Software History, And Can We? -- 1 Introduction -- 2 A Complex Task -- 3 A Universal Mission -- 4 Past, Present, Future: Much More Than an Archive -- 5 A Peek Under the Hood -- 6 Beyond Preservation, A Strategic Issue -- 7 For More Information -- References -- Full Papers -- Analyzing the Web: Are Top Websites Lists a Good Choice for Research? -- 1 Introduction -- 2 Related Work -- 3 Experiments and Data -- 3.1 Experimental Setup -- 3.2 Search -- 3.3 Top Sites -- 3.4 Common Crawl -- 3.5 Wikipedia -- 3.6 Twitter -- 4 Results -- 4.1 Completeness: Overlap of Data Sets -- 4.2 Search Results Versus Other Data Sets -- 4.3 Is a List of 1 Million Popular Sites Enough? -- 4.4 TLDs are Neither a Good Proxy for a Language nor for a Locale -- 5 Discussion and Conclusion -- References -- RDFtex: Knowledge Exchange Between LaTeX-Based Research Publications and Scientific Knowledge Graphs -- 1 Introduction -- 2 Foundations -- 2.1 Related Work -- 3 Concept -- 3.1 Import of Contributions -- 3.2 Export and Integration of Contributions -- 3.3 Prefix Syntax -- 4 Discussion -- 5 Conclusion -- References -- Enriching the Greek National Cultural Aggregator with Key Figures in Greek History and Culture: Challenges, Methodology, Tools

and Outputs -- 1 Introduction -- 2 Background: The Enrichment Scheme in SearchCulture.gr -- 3 Adapting the Enrichment Scheme for Person Entities -- 3.1 Two Kinds of Enrichments: Creators and Referred Persons -- 3.2 The Disambiguation Problem -- 3.3 The Scope -- 3.4 Developing a Vocabulary of Persons in Greek History and Culture -- 3.5 The Strategy for Enriching Aggregated Collections En Masse -- 3.6 Effort -- 4 Person-Driven Search and Browsing Functionality -- 5 Related Work.
6 Conclusions -- References -- Figure and Figure Caption Extraction for Mixed Raster and Vector PDFs: Digitization of Astronomical Literature with OCR Features -- 1 Introduction -- 2 The Data -- 3 Model Pipeline Development -- 3.1 Deep Learning Model and Feature Selection -- 3.2 Image Processing -- 3.3 Post-processing Pipeline -- 3.4 Feature Selection Ablation Experiments -- 4 Results -- 4.1 Benchmarks at High Levels of Localization (IOU = 0.9) -- 5 Discussion and Future Work -- References -- FAIROs: Towards FAIR Assessment in Research Objects -- 1 Introduction -- 2 Related Work -- 2.1 Adopting the FAIR Principles -- 2.2 Quality Assessment Models for Research Objects -- 3 FAIROs: A Research Object FAIR Assessment Framework -- 3.1 Modules for Individual Resource FAIR Assessment -- 3.2 Scoring Aggregations of Resources in Research Objects -- 3.3 Result Format and Visualization -- 4 Comparing FAIR Research Object Assessment Metrics -- 5 Discussion -- 6 Conclusions and Future Work -- References -- Searching Wartime Photograph Archive for Serious Leisure Purposes -- 1 Introduction -- 1.1 Image Searching in the History Domain -- 1.2 Search Goals and Image Searching -- 1.3 User Engagement -- 2 Research Setting -- 2.1 The Photograph Archive -- 2.2 The Questionnaire and Data Collection -- 3 Results -- 3.1 Respondents -- 3.2 Search Goals -- 3.3 Search Tactics -- 3.4 Looking at the Photographs -- 4 Discussion -- 5 Conclusion -- References -- Implementation and Evaluation of a Multilingual Search Pilot in the Europeana Digital Library -- 1 Introduction -- 2 State of the Art -- 3 Approach -- 4 Evaluation and Results -- 4.1 Real-Time Detection of the Query Language -- 4.2 Real-time Translation of the Query -- 4.3 Construction of the Multilingual Query -- 4.4 Impact of Typos, Ambiguity, and Non-translatable Entities (Q7).
4.5 Other Elements Affecting Relevance in the Top ten Positions (Q8) -- 5 Discussion -- 6 Conclusions and Future Work -- References -- DETEXA: Declarative Extensible Text Exploration and Analysis -- 1 Introduction -- 2 Related Work -- 2.1 Text Analysis Tools -- 2.2 Data Management Systems with Text Mining Opportunities -- 3 The DETEXA Framework -- 3.1 Function Library -- 3.2 Why YeSQL? -- 3.3 Interfaces -- 4 Evaluation -- 4.1 Experimental Setup -- 4.2 Experiments -- 5 Conclusions -- References -- The Way We Cite: Common Metadata Used Across Disciplines for Defining Bibliographic References -- 1 Introduction -- 2 Related Works -- 3 Materials and Methods -- 4 Results -- 5 Discussion and Lessons Learnt -- 6 Conclusions -- References -- Event Notifications in Value-Adding Networks -- 1 Introduction -- 2 Related Work -- 3 The Value-Adding Network -- 3.1 Design and Technology Considerations -- 3.2 A Network of Data Nodes and Service Nodes -- 3.3 Network Communication Patterns -- 3.4 Anatomy of a Notification Message -- 4 Experimental Investigation -- 4.1 Experiment -- 5 Summary and Outlook -- References -- A Chromium-Based Memento-Aware Web Browser -- 1 Introduction -- 2 Background -- 3 Related Work -- 4 Memento Detection and Icon Display -- 4.1 Detection of Root Page Memento -- 4.2 Detection of Mementos Embedded in a Live Page -- 4.3 Detection of Live Content Embedded in a Memento -- 5 Bookmark as Archive -- 6 Evaluation --

6.1 Memento Detection -- 6.2 Bookmark as Archive -- 7 Future Enhancements -- 8 Conclusion -- References -- Analysis of the Deletions of DOIs -- 1 Introduction -- 2 Related Work -- 3 Materials and Methods -- 3.1 Identifying Deleted DOIs -- 3.2 Analysis Methods -- 4 Results and Discussion -- 4.1 Basic Statistics of Deleted DOIs -- 4.2 Content Analysis Based on Crossref Metadata and DOI Links -- 4.3 Prefix Analysis -- 4.4 Suffix Analysis.
5 Conclusion -- References -- Analysing User Involvement in Open Government Data Initiatives -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 4 Results -- 5 Discussion -- 6 Conclusion -- References -- The Rise of GitHub in Scholarly Publications -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 4 Results -- 5 Discussion -- 6 Conclusions -- References -- Whois? Deep Author Name Disambiguation Using Bibliographic Data -- 1 Introduction -- 2 Related Work -- 2.1 Unsupervised-Based -- 2.2 Supervised-Based -- 2.3 Graph-Based -- 3 Approach: WhoIs -- 3.1 Model Architecture -- 3.2 Author Name Representation -- 3.3 Source and Title Embedding -- 3.4 Model Training -- 3.5 Prediction -- 4 Experiments -- 4.1 Dataset -- 4.2 Results -- 4.3 Limitations and Obstacles of WhoIs -- 5 Conclusion -- References -- Investigations on Meta Review Generation from Peer Review Texts Leveraging Relevant Sub-tasks in the Peer Review Pipeline -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 The Various Prediction Sub-tasks -- 3.2 Seq-to-Seq Meta-Review Generation: Main Task -- 4 Evaluation -- 4.1 Dataset -- 4.2 Baselines and Comparing System -- 5 Results and Analysis -- 5.1 Quantitative and Qualitative Analysis -- 5.2 Error Analysis -- 6 Conclusion -- References -- Improving Accessibility to Arabic ETDs Using Automatic Classification -- 1 Introduction -- 2 Related Work -- 2.1 Classical Machine Learning Approaches -- 2.2 Deep Learning Approach -- 3 Dataset -- 3.1 Mapping of AskZad Categories to ProQuest Categories -- 3.2 Data Preprocessing -- 4 Methodology -- 4.1 Classical Supervised Machine Learning -- 4.2 Deep Learning Using Pretrained Language Models -- 5 Results and Discussion -- 5.1 Classical Supervised Machine Learning -- 5.2 Deep Learning Using Pretrained Language Models -- 6 Conclusions -- 7 Future Work -- References -- Short Papers.
Robots Still Outnumber Humans in Web Archives, But Less Than Before -- 1 Introduction -- 2 Background and Related Work -- 3 Methodology -- 3.1 Dataset -- 3.2 Data Cleaning -- 3.3 Session Identification -- 3.4 Bot Identification -- 4 Results and Analysis -- 4.1 Robots vs. Humans -- 4.2 Discovering Access Patterns -- 4.3 Identifying Temporal Preferences -- 5 Future Work -- 6 Conclusions -- References -- Cui Bono? Cumulative Advantage in Open Access Publishing -- 1 Introduction -- 2 Related Work -- 3 Data Sources -- 4 Methodology -- 5 Results -- 5.1 Institutional Ranking and OA Consumption -- 5.2 Correlation Between Production/Consumption and Institutional Ranking -- 6 Conclusion -- References -- Overview Visualizations for Large Digitized Correspondence Collections: A Design Study -- 1 Introduction -- 2 Related Work -- 3 Collaborative Design -- 4 The Correspondence Matrix -- 5 Conclusion -- References -- Developing the EOSC-Pillar RDM Training and Support Catalogue -- 1 Introduction -- 1.1 Defining the Scope of the Catalogue -- 1.2 Accessing the EOSC-Pillar RDM Training and Support Catalogue -- 2 Catalogue Technical Specifications -- 2.1 Technology -- 2.2 Metadata Profile -- 3 Outcomes and Future Work -- References -- Automatic Generation of Coherent Image Galleries in Virtual Reality -- 1 Introduction -- 2 Related Work -- 3 System Architecture -- 4 Automatic Collection Generation -- 4.1 Self-Organizing Maps -- 4.2 From SOM to Exhibition -- 4.3 Features -- 5

Discussion and Conclusion -- References -- Extracting Funder Information from Scientific Papers - Experiences with Question Answering -- 1 Introduction -- 2 Related Work -- 3 Approach -- 3.1 Processing Pipeline -- 3.2 Data -- 4 Results -- 5 Discussion and Outlook -- References -- CDX Summary: Web Archival Collection Insights -- 1 Introduction -- 2 Implementation -- 2.1 Collection Overview.
2.2 Media Types and Status Codes.