Editorial for the TPDL 2015 special issue

Sarantos Kapidakis¹ · Cezary Mazurek² · Marcin Werla²

Published online: 26 July 2016
© Springer-Verlag Berlin Heidelberg 2016

The International Conference on Theory and Practice of Digital Libraries (TPDL) constitutes a leading scientific forum on digital libraries that brings together researchers, developers, content providers and users in the field of digital libraries. The advent of the technologies that enhance the exchange of information with rich semantics is of particular interest in the community. Information providers inter-link their metadata with user contributed data and offer new services outlooking to the development of a web of data and addressing the interoperability and long-term preservation challenges.

TPDL 2015, the nineteenth International Conference on Theory and Practice of Digital Libraries, held in Poznań, Poland during September 14–18, 2015, had the general theme “Connecting Digital Collections” and focused on four major themes: “Connecting digital libraries”, “Practice of digital libraries”, “Digital libraries in science” and “Users, communities, personal data”.

All submissions were independently reviewed on the basis of a triple peer review process, initially by four members of the Program Committee. A Senior Program Committee member subsequently coordinated a discussion among the reviewers. The selection stage that followed compared the paper evaluations and finalized the conference program. The ten papers with best TPDL 2015 review scores were selected as candidates for the Special Issue of the International Journal on Digital Libraries, asking the authors to introduce at least 30 % difference from the conference original paper. The new submissions were reviewed by four reviewers, common with the original TPDL 2015 reviewers if possible. At the end,

seven papers were accepted for publication to the Special Issue.

The paper “A Semantic Architecture for Preserving and Interpreting the Information Contained in Irish Historical Vital Records” is an extension of the conference paper “On a Linked Data Platform for Irish Historical Vital Records” by Christophe Debruyne, Oya Deniz Beyan, Rebecca Grant, Sandra Collins and Stefan Decker. It created two distinct ontologies and knowledge bases to support the clear separation of concerns that reflects the transcription and archival authenticity of the register pages and the interpretation of the transcribed data in the Irish Record Linkage 1864–1913 project. The advantage of this clear separation is the transcription of register pages resulted in a reusable dataset fit for other research purposes. These transcriptions were enriched with metadata according to the best practices in archiving for ingestion in suitable digital, long-term preservation platforms.

The paper “Using a File History Graph to Keep Track of Personal Resources across Devices and Services” is an extension of the conference paper “Memsy: Keeping Track of Personal Digital Resources across Devices and Services” by Matthias Geel and Moira Norrie. It introduces the concept of a file history graph that can be used to provide users with a global view of resource provenance and enable them to track specific versions across devices and services. It describes how this has been used to realize a version-aware environment, called Memsy, and reports on a laboratory study used to evaluate the proposed workflow. It also describes how reconciliation services can be used to fill in missing links in the file history graph and presents a detailed study for the case of images as a proof of concept.

The paper “Evaluating Unsupervised Thesaurus-based Labeling of Audiovisual Content in an Archive Production Environment” is an extension of the conference paper

✉ Sarantos Kapidakis
sarantos@ionio.gr

¹ Ionian University, Corfu, Greece

² Poznań Supercomputing and Networking Center, Poznań, Poland

“Practice-oriented Evaluation of Unsupervised Labeling of Audiovisual Content in an Archive Production Environment” by Victor de Boer, Roeland Ordelman and Josefiën Schuurman. It reports on a two-stage evaluation of unsupervised labeling of audiovisual content using collateral text data sources to investigate how such an approach can provide acceptable results given requirements with respect to archival quality, authority and service levels to external users. It concludes that with parameter settings that are optimized using a rigorous evaluation of precision and accuracy, the quality of automatic term-suggestion is sufficiently high.

The paper “Detecting Off-Topic Pages in Web Archives” is an extension of the conference paper having the same title by Yasmin Alnoamany, Michael Nelson and Michele Weigle. It addresses the problems of detecting off-topic pages in Web archive collections. It proposes different methods to detect when a page has gone off-topic relative to its first capture. Those predicted off-topic pages will be presented to the collection’s curator for possible elimination from the collection or cessation of crawling. It also created a gold standard data set from three Archive-It collections to evaluate the proposed methods at different thresholds.

The paper “Archive Profiling Through CDX Summarization” is an extension of the conference paper having the same title by Sawood Alam, Michael Nelson, Herbert Van de Sompel, Lyudmila L. Balakireva, Harihar Shankar and David Rosenthal. It shows how to generate profiles of the archives that summarize their holdings, using the CDX files produced after crawling, and can be used to inform routing of the Memento aggregator’s URI requests. It explores strategies between using full URIs (no false positives, but with large profiles) to using only top-level domains (TLDs)

(smaller profiles, but with many false positives). In the presented experiments, the registered domain profile doubled the routing precision with respect to the TLD-only profile, while complete hostname and one path segment gave a ten-fold increase in the routing precision.

The paper “Characteristics of Social Media Stories. What makes a good story?” is an extension of the conference paper “Characteristics of Social Media Stories” by Yasmin Alnoamany, Michele Weigle and Michael Nelson. It investigated 14,568 stories from Storify, comprising 1,251,160 individual resources. It also checked the population of Archive-It collections (3109 collections comprising 305,522 seed URIs) for clarifying the intended framework of its archival summaries characteristics. It found that the resources in human-generated stories are different from the resources in Archive-It collections. In summarizing a collection, we can only choose from what is archived (e.g., twitter.com is popular in Storify, but mostly it is missing in Archive-It). However, some other characteristics of human-generated stories will be applicable such as the number of resources.

The paper “What’s news? Encounters with news in everyday life: a study of behaviours and attitudes” is an extension of the conference paper “Digital news resources: An autoethnographic study of news encounters” by Sally Jo Cunningham, David Nichols, Annika Hinze and Judy Bowen. It analyzed a set of 35 autoethnographies of news encounters, created by students in New Zealand. These comprise rich descriptions of the news sources, modalities, topics of interest, and news ‘routines’ by which the students keep in touch with friends and maintain awareness of personal, local, national, and international events. It explores the implications of these insights into news behavior for digital news systems.