

**Santy Chumbe**

*Heriot Watt University, UK*

***ticTOCron: An automatic solution for propagating quality metadata to scholarly RSS TOC feed metadata***

Keeping up-to-date with the latest publications is at the heart of much research activity. Online current awareness services are important to virtually all librarians and researchers. As part of this process, users have a growing interest in the RSS feeds that publishers are providing for an increasing number of scholarly journal Table of Contents (TOCs). Publishers and users stand to benefit from the facilitation of more widespread syndication and easier access to RSS TOC feeds; however, many TOC RSS feeds are at present being produced with erroneous, poor or incomplete metadata. This can hamper the usefulness of scholarly current awareness services. This is exactly what the ticTOCron software toolkit is aiming to overcome. The ticTOCron toolkit automatically enhances poor, heterogeneous and incomplete metadata found in TOC RSS feeds by making use of a pre-defined "Best Practice" metadata set suitable for scholarly journals. The toolkit has been developed as part of the JISC-funded ticTOCs Project which is prototyping a journal tables of contents service. The toolkit works as an "invisible infrastructure" for the ticTOCs public service prototype. ticTOCron itself is a set of software tools managed by a computer system utility (CRON) found in most UNIX systems. In this work we depict the main issues and "bad practices" found in RSS TOC metadata harvested from more than 300 scholarly publishers. We also describe the implemented solutions. Some references are made to the algorithms for generating semantic relations within, between and from the harvested TOCs and to the mechanisms for propagating "metadata associations" from a previously crawled metadata-rich reference set. However, an effort is made to avoid technical jargon and to replace complex technical descriptions with samples and simple comparisons. The determination of the "Best Practices metadata set" for scholarly papers is also discussed. We also present the results produced by ticTOCron when it was used for enhancing and normalizing the TOC RSS feeds harvested by the ticTOCs Project for more than 12,000 journals. Finally we propose a sustainable and scalable computational scenario whereby the automatic solution is complemented and fine-tuned by a cost-effective human cross-validation process.

---

*Dr. Chumbe is a Research Associate at the Heriot-Watt University in UK, where he works at the Institute for Computer Based Learning (ICBL), School of Mathematical and Computer Sciences. His interests are in the area of RSS syndication, digital libraries, federated searching and e-journal publishing. He has been the Principal Investigator of the ticTOCs Project. He is the technical manager of the TechXtra service and a member of Europeana Project.*