IDCC15 | *Poster*

Scenario based decision-making on information encapsulation techniques

Anna Grit Eggers, Jens Ludwig
Göttingen State and University Library
Georg August Universität
37070 Göttingen, Germany

Fabio Corubolo
IPHS, University of Liverpool
Waterhouse Building, Brownlow St.,
Liverpool L69 3GL, UK

Efstratios Kontopoulos
Information Technologies Institute
Centre for Research & Technology – Hellas
6th km Xarilaou - Thermi, 57001, Thessaloniki, Greece

Abstract

The domain of information encapsulation encompasses techniques of packaging (METS, BagIt, ...) and information embedding (XMP, Watermarking, ...). However, given the abundance of available information encapsulation techniques, end users may have difficulty deciding which technique to apply to their scenario. In PERICLES we are developing a general purpose information encapsulation tool that integrates encapsulation techniques into an open source framework, supporting decision-making regarding the appropriate technique for a given user scenario. The decision mechanism suggests a specific technique, based on a comparison between its characteristics and the use scenario requirements. Our poster gives a brief overview of information encapsulation techniques, as well as an insight into the current research on our decision-making approach and the development of the framework.

Each information encapsulation technique has distinguishing features and characteristics, illustrated in figure 1, that make its application appropriate for certain scenarios and inappropriate for others. The issue is to evaluate these characteristics for all applicable techniques and to decide which one is the most suitable for a given scenario. Long term preservation scenarios emphasize preservation of encapsulated information in a standardised way. Therefore, the use of techniques that pack the information together with well-defined metadata schemes is preferred, such as METS (METS, 2002) and OAI-ORE (OAI-ORE, 2006). However, some embedding techniques also guarantee the bit-correct recovery of all information, and could be used for such scenarios. The key advantage of their use is that there is no need to unpack the information before it can be processed normally. An example is a technique that uses the

Submitted 14 November 2014

Correspondence should be addressed to Anna Eggers, Papendiek 14, 37070 Göttingen, Germany. Email: eggers@sub.uni-goettingen.de

The 10th International Digital Curation Conference takes place on 9–12 February 2014 in London. URL: http://www.dcc.ac.uk/events/idcc15/

file format feature of a PDF file to attach external files to the PDF. The recoverability of the carrier PDF, as well as the attached payload file, makes this technique applicable in preservation scenarios. Other embedding techniques, such as the F5 steganography algorithm (Westfeld, 2001) alter the carrier file making it impossible to restore precisely. Our framework includes such techniques as they could be applied in many scenarios, e.g. scenarios which focus on data hiding or fingerprinting.

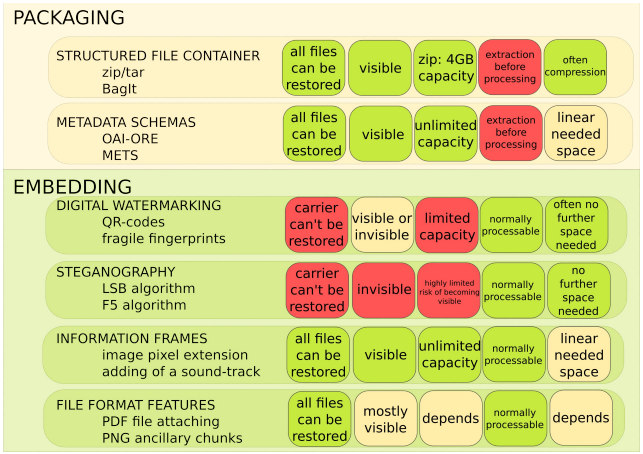

Figure 1. Overview of information encapsulation techniques and distinctions

We explored various decision-making techniques for the integration into our framework, and came to the conclusion that such decision can be made by measuring the distance from the available techniques to the use scenario. We plan to use this calculation to measure how well the features of each technique fulfil the scenario requirements giving the user option to provide importance weighting of the parameters.

The framework is designed to collaborate with the PERICLES Extraction Tool (Corubolo et al., 2014). It is planned to develop an adapter for the tools, to execute them together in a sheer curation scenario in which information is extracted from a working environment and directly encapsulated with the digital object. This approach prevents information loss in case the object is shipped away from its environment, and supports the digital object reuse.

Acknowledgements

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7- 601138 PERICLES.

Submitted 14 November 2014

Correspondence should be addressed to Anna Eggers, Papendiek 14, 37070 Göttingen, Germany. Email: eggers@sub.uni-goettingen.de

The 10th International Digital Curation Conference takes place on 9–12 February 2014 in London. URL: http://www.dcc.ac.uk/events/idcc15/

References

[proceedings] Corubolo, F., Eggers, A. G., Hasan, A., Hedges, M., Waddington, S., and Ludwig, J. 2014. A pragmatic approach to signifcant environment information collection to support object reuse, in IPRES 2014 proceedings.

[website] (2002). METS - Library of Congress. Retrieved November 10, 2014, from http://www.loc.gov/standards/mets/.

[website] (2006). OAI-ORE - Open Archives Initiative. Retrieved November 5, 2014, from http://www.openarchives.org/ore/.

[website] (2013). PERICLES – FP7 Digital Preservation. Retrieved November 5, 2014, from http://pericles-project.eu/.

[proceedings] Westfeld, A. (2001, January). F5—a steganographic algorithm. In *Information hiding* (pp. 289-302). Springer Berlin Heidelberg.

Submitted 14 November 2014

Correspondence should be addressed to Anna Eggers, Papendiek 14, 37070 Göttingen, Germany. Email: eggers@sub.uni-goettingen.de

The 10th International Digital Curation Conference takes place on 9–12 February 2014 in London. URL: http://www.dcc.ac.uk/events/idcc15/