

## Connecting the 'webs'. Building interoperability into online services for stemmatology

*Tara Andrews, Simo Linkola, Teemu Roos and Joris van Zundert*

In this paper we present a case study in interoperability and modular development for digital tools for humanities research. We demonstrate the tools themselves, discuss the constraints in each case for interoperability, and present an API devised for lightweight but robust and extensible interoperation of the tools for a generalized task.

In 2012 two online tools for text stemmatology were released independently. The first of these is Stemweb, developed at HIIT Helsinki Institute for Information Technology for the inference of phylogenetic trees based upon textual data using techniques such as RHM (Roos, Heikkilä, and Myllymäki 2006) and SemStem (Roos and Zou 2011), developed specifically for the case of recovering manuscript text stemmata. The second tool is Stemmaweb, developed by the Tree of Texts project at KU Leuven (now hosted at the University of Bern) in collaboration with members of the Interedition project for annotation, visualization, and regularization of text variants and analysis of variation against one or more stemma hypotheses using graph search methods (Andrews, Blockeel, Bogaerts et al. 2012). The hypotheses are provided by the scholar, who may define as many hypothetical stemmata as she or he would like to explore.

The value of interoperability between a tool that produces stemma hypotheses (or trees that serve as the basis of stemma hypotheses) and one that evaluates them is extremely clear, and in accordance with the principles of the Interedition project (Van Zundert et al. 2011) we wished to find a mode for interoperation that was as lightweight as possible, without the need to maintain extra infrastructure and with as minimal as possible a set of standards. Our approach was to define the individual components of the two services in as modular a way as possible (after the pattern of the Gothenburg model for text collation; see Dekker and Middell 2011) and to provide a lightweight RESTlike API for the abstract problem of phylogenetic tree inference for textual data; this allows not only interoperable functionality between the two tools, but will also allow for the seamless use of any algorithms developed independently in future.

In most microservice-based architectures, the different services are called through a user interface that ties them together; in this case a user interface is implemented by Stemmaweb, as well as various supporting services that may also be called independently as microservices. Stemweb likewise has its own user interface that can be used independently. Crucially, our architecture does not dictate that the services be accessed only through Stemmaweb—any present or future online service may also make use of the tools we provide that are not dependent on the specific GUI implementation.

### Acknowledgements

The authors would like to thank the European Association for Digital Humanities (EADH) for their support of this project through a Small Project Grant in spring 2013. Work on Stemmaweb was also supported by the COST Action IS0704 'Interedition' and the CREA fund of KU Leuven; work on Stemweb was supported by the Academy of Finland (CoE COIN).