



Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

Connecting the 'webs': Building interoperability into online services for stemmatology

van Zundert, J.; Andrews, T.L.; Linkola, S.; Roos, Teemu

2014

document version

Early version, also known as pre-print

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

van Zundert, J., Andrews, T. L., Linkola, S., & Roos, T. (2014). *Connecting the 'webs': Building interoperability into online services for stemmatology*. Paper presented at Digital Humanities Benelux Conference 2014, The Hague, Netherlands.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl

Connecting the 'webs': Building interoperability into online services for stemmatology

Tara L Andrews¹, Simo Linkola², Teemu Roos², Joris van Zundert³

1) CH - University of Bern; 2) FI - Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki; 3) NL - Huygens ING

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,¹ developed at HIIT Helsinki Institute for Information Technology for the inference of phylogenetic trees based upon textual data using techniques such as RHM² and SemStem,³ developed specifically for the case of recovering manuscript text stemmata. The second tool is Stemmaweb,⁴ 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.⁵ 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 project⁶ 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 collation⁷) and to provide a lightweight

¹ Source code available at <https://github.com/Stemweb/Stemweb> [All links last accessed 1 March 2014].

² T. Roos, T. Heikkilä, and P. Myllymäki (2006), "A Compression-Based Method for Stemmatic Analysis". Proceedings of the 2006 European Conference on Artificial Intelligence (ECAI), Riva del Garda, 29 August–1 September 2006, pp. 805–6. [<http://cosco.hiit.fi/Articles/ecai06.pdf>].

³ T. Roos and Y. Zou (2011), "Analysis of Textual Variation by Latent Tree Structures". Proceedings of the 2011 IEEE International Conference on Data Mining, Vancouver, 11–14 December 2011. [<http://www.cs.helsinki.fi/u/ttonteri/pub/icdm2011.pdf>].

⁴ <http://stemmaweb.net/>; source code available at <https://github.com/tla/stemmaweb/>.

⁵ T. Andrews, H. Blockeel, B. Bogaerts et al. (2012), "Analyzing manuscript traditions using constraint-based data mining". Proceedings of CoCoMile 2012 - COmbining COnstraint solving with MIning and LEarning, Montpellier, 27 August 2012. [http://cocomile.disi.unitn.it/2012/papers/cocomile2012_manuscript.pdf].

⁶ For a discussion of these principles see J. van Zundert et al. (2011), "Interedition—Principles, Practice and Products of an Open Collaborative Development Model for Digital Scholarly Edition". *Digital Humanities 2011: Conference Abstracts*

[https://www.stanford.edu/group/dh2011/cgi-bin/wordpress/wp-content/uploads/2011/05/DH2011_BookOfAbs.pdf].

⁷ R. Dekker and G. Middell, "Computer-supported Collation with CollateX: Managing Textual Variance in an

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.

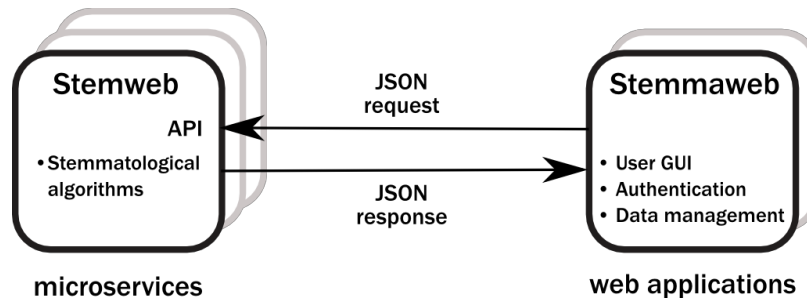


Figure 1: Microservice and client architecture

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).