

# Simtho: The Syriac Thesaurus

## LAUNCH OF A SYRIAC TEXTUAL CORPUS PORTAL

April 23  
12–2 PM EST  
over Zoom

Hosted by  
Sabine Schmidtke and  
George A. Kiraz

Institute for Advanced Study,  
Princeton



Beth Mardutho: The Syriac Institute will be launching a new Syriac corpus portal over Zoom co-hosted with DigitalScholarship@IAS (Institute for Advanced Study, Princeton). The event is free and open to the public. Please register [here](#).

Simtho [simtho.bethmardutho.org] is a Syriac corpus search engine with a textual database spanning almost two millennia. A Beta version was revealed at the 2018 AAR/SBL meeting in San Diego with over 6 million words. The upcoming Beta II zLaunch will uncover a textual database of over 13 million words with a new responsive and more attractive user interface.

Team members will discuss various digital humanities and computational linguistics techniques including corpus building, the power of regular expressions, building OCR and HTR models, metadata, and part-of-speech tagging. While these techniques are applied to Syriac, they can be easily transferable to other (especially Semitic) languages.

### PROGRAM

- Welcome
- Simtho Overview
- Initial Designs, First Attempts
- Building OCR/HTR Models
- Metadata Considerations
- Part-of-Speech Tagging
- Reactions

- Q&A
- Closing & zToast

- Sabine Schmidtke and Maria Tuya (IAS)
- George A. Kiraz (IAS & Beth Mardutho: The Syriac Institute)
- Johan Lundberg (Cambridge University)
- Abigail Pearson (Exeter University) & Brandon Allen (BM)
- Shelby Loster (Beth Mardutho: The Syriac Institute)
- Slavomír Čéplö (Austrian Academy of Sciences)
- Sebastian Brock (University of Oxford)
- Jack Tannous (Princeton University)

- Sabine Schmidtke and Maria Tuya (IAS)
- Open your own champagne 🍾