

Benjamin Kiessling

CONTACT INFORMATION	27 rue Auvry 93300 Aubervilliers France	<i>mobile:</i> +33 7 67 53 59 58 <i>email:</i> benjamin.kiessling@psl.eu
PERSONAL INFORMATION	Date of Birth Nationality	31/12/1991 German
EDUCATION AND DEGREES	April 2021 October 2012 - October 2016 October 2009 - February 2013 August 2008 - July 2009	PhD, Computer Science, École Pratique des Hautes Études, Université PSL, <i>Advances in Optical Character Recognition for Historical Arabic Documents</i> Dissertation Committee: Marc Bui (Directeur de thèse) Gregory Crane (Rapporteur) Nachum Dershowitz (Rapporteur) Alicia Fornés Peter Stokes (Président) Daniel Stökl Ben Ezra M.Sc. Computer Science, University of Leipzig, <i>Median Strings for OCR Post-correction</i> B.Sc. Computer Science, University of Leipzig Academic Honors High School Diploma, Rockville Jr-Sr High School, USA
ACADEMIC AFFILIATIONS AND WORK EXPERIENCE	École Pratique des Hautes Études , Université PSL, Paris UMR 8546 Archéologie et Philologie d'Orient et d'Occident <i>Research Engineer</i>	May 2018 - <ul style="list-style-type: none">• Machine learning for paleographic and epigraphic applications• Principal involvement in the following projects:<ul style="list-style-type: none">• Biblissima+: Transformer-based document analysis, unsupervised pre-training of OCR models• ALTO: Evolution of the standard for better handwriting, non-Latin text, and complex layout encoding• DIM STCN: Conception of a high performance cluster for large-scale text recognition and ancillary tasks• RESILIENCE: Trainable document layout analysis methods for complex manuscripts• Tikkoun Sofrim and Sofer Mahir: OCR for medieval Hebrew manuscripts with crowdsourcing integration for postcorrection• eScripta in PSL Scripta: Adaptation of an OCR engine for handwritten text recognition as part of the eScriptorium platform; preparation of open datasets for text recognition and layout analysis
	Saxon Academy of Sciences , Leipzig <i>Researcher</i>	April 2018 <ul style="list-style-type: none">• Implementation of an OCR workflow for Arabic manuscript catalogues in the Bibliotheca Arabica project
	Faculty of Information Sciences and Knowledge Studies , University of Tehran (faculty sponsor: Prof. Fatima Fahimnia)	

Visiting Researcher

December 2017 - February 2018

- Development of a text line layout analysis method for a corpus of Iranian newspapers and periodicals
- Survey on characteristics of historical Persian language manuscript page layout
- Two introductory seminars on the principles of OCR for M.Sc. students of Information Science

Alexander-von-Humboldt-Chair of Digital Humanities, Department of Computer Science, University of Leipzig

Researcher

February 2017 - April 2018

- Methods of OCR based on artificial neural networks
- Page layout analysis using deep learning
- Training of OCR models for Arabic, Persian, Syriac, polytonic Greek, Hebrew, and style detection
- Evaluation of distributed training data creation using a crowdsourcing platform
- Integration of OCR functionality in a distributed corpus-building tool as part of **OpenITI/ShariaSource**

Research Assistant

February 2014 - January 2017

- Development of a distributed OCR pipeline
- Survey on dictionary-based OCR postcorrection for polytonic Greek texts
- **EUDAT** pilot project on the long-term archival of corpus data
- Training of neural networks for recognition of classical Arabic texts

Lehrstuhl für Rechnernetze und verteilte Systeme, Department of Computer Science, University of Leipzig

Research Assistant

August 2011 - February 2014

- Development of a wireless measurement infrastructure for IEEE 802.11n networks
- Port and enhancement of an ath9k patchset for the Linux kernel

EPAK GmbH, Leipzig

Network Administrator/Software Developer

July 2009 - August 2011

TEACHING

2022

Intelligence artificielle pour les SHS : initiation à la transcription automatique des documents (Artificial Intelligence for the Humanities: Text Recognition Tutorial), PSL-week, lectures and tutorials (8h), EPHE, PSL University

2020-2022

Fouille de données et apprentissage (Data Mining and Machine Learning), 2nd year Master Digital Humanities, lectures and tutorials (12h), University of Tours

2020-2021

Systèmes d'exploitation et Web (Operating Systems and Web), 1st year Master Digital Humanities, lectures, tutorials, and practical classes (12h), University of Tours

PUBLIC ENGAGEMENT	2021	<i>Paroles de chercheuses et chercheurs</i> , seminar, science communication for high school students, Lycée Gustave Monod, Enghien-les-Bains
SKILLS	Programming languages Software Formats and Standards	C, Python, Perl, Bash, Tcl pytorch, tensorflow, lxml, numpy, scipy METS, MODS, ALTO, hOCR, Page XML, MARCXML, TEI, Handle, Uni- code, protobuf, JSON, Arrow
LANGUAGES	German English French	native speaker bilingual intermediate/upper-intermediate
PHD THESIS	Benjamin Kiessling . “Advances in Optical Character Recognition for Historical Arabic Documents (Avancées en Reconnaissance Optique des Caractères pour les Documents Arabes Historiques)”. PhD thesis. École Pratique des Hautes Études, Université PSL, Paris, Apr. 2021	
JOURNAL ARTICLES	<p>P.A. Stokes and Benjamin Kiessling. “Sharing Data for Handwritten Text Recognition”. In: <i>DH in Practice</i> (2023). to appear</p> <p>Alan J. Wecker, Vered Raziel-Kretzmer, Benjamin Kiessling, Daniel Stökl Ben Ezra, Moshe Lavee, Tsvi Kuflik, Dror Elovits, Moshe Schorr, Uri Schor, and Pawel Jablonski. “Tikkoun Sofrim: Making Ancient Manuscripts Digitally Accessible: The Case of Midrash Tanhuma”. In: <i>J. Comput. Cult. Herit.</i> 15.2 (Apr. 2022). ISSN: 1556-4673. DOI: 10.1145/3476776</p> <p>Benjamin Kiessling, Gennady Kurin, Matthew Thomas Miller, and Kader Smal. “Advances and Limitations in Open Source Arabic-Script OCR: A Case Study”. In: <i>Digital Studies/Le champ numérique</i> 11.1 (2021). DOI: 10.16995/dscn.8094</p> <p>P.A. Stokes, Benjamin Kiessling, Daniel Stökl Ben Ezra, Robin Tissot, and El Hassane Gargem. “The eScriptorium VRE for Manuscript Cultures”. In: <i>Classics@ 18</i> (1 2021)</p> <p>Benjamin Kiessling, Matthew Thomas Miller, G Maxim, Sarah Bowen Savant, et al. “Important New Developments in Arabographic Optical Character Recognition (OCR)”. in: <i>Al-ʿUṣūr al-Wuṣṭā</i> 25 (2017), pp. 1–13. DOI: 10.48550/arXiv.1703.09550</p>	
CONFERENCE PAPERS	<p>Benjamin Kiessling. “CurT: End-to-End Text Line Detection in Historical Documents with Transformers”. In: <i>Frontiers in Handwriting Recognition</i>. Ed. by Utkarsh Porwal, Alicia Fornés, and Faisal Shafait. Cham: Springer International Publishing, 2022, pp. 34–48</p> <p>Daniel Stökl Ben Ezra, Bronson Brown-DeVost, Pawel Jablonski, Hayim Lapin, Benjamin Kiessling, and Elena Lolli. “BibLIA - a General Model for Medieval Hebrew Manuscripts and an Open Annotated Dataset”. In: <i>HIP@ICDAR 2021: The 6th International Workshop on Historical Document Imaging and Processing</i>,</p>	

Lausanne, Switzerland, September 5-6, 2021. ACM, 2021, pp. 61–66. DOI: 10.1145/3476887.3476896

Benjamin Kiessling. “A Modular Region and Text Line Layout Analysis System”. In: *17th International Conference on Frontiers in Handwriting Recognition, ICFHR 2020, Dortmund, Germany, September 8-10, 2020*. IEEE, 2020, pp. 313–318. DOI: 10.1109/ICFHR2020.2020.00064

Daniel Stökl Ben Ezra, Bronson Brown-DeVost, Nachum Dershowitz, Alexey Pechorin, and **Benjamin Kiessling.** “Transcription Alignment for Highly Fragmentary Historical Manuscripts: The Dead Sea Scrolls”. In: *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. 2020, pp. 361–366. DOI: 10.1109/ICFHR2020.2020.00072

Benjamin Kiessling, Daniel Stökl Ben Ezra, and Matthew Thomas Miller. “BADAM: A Public Dataset for Baseline Detection in Arabic-script Manuscripts”. In: *Proceedings of the 5th International Workshop on Historical Document Imaging and Processing, HIP@ICDAR 2019, Sydney, NSW, Australia, September 20-21, 2019*. ACM, 2019, pp. 13–18. DOI: 10.1145/3352631.3352648

Benjamin Kiessling, Daniel Stökl Ben Ezra, Rodney Ast, and Holger Essler. *Aligning Extant Transcriptions of Documentary and Literary Papyri with their Glyphs*. 29th International Congress of Papyrology (Lecce). 2019

Benjamin Kiessling, Robin Tissot, Daniel Stökl Ben Ezra, and Peter Anthony Stokes. *eScripta: A New Digital Platform for the Study of Historical Texts and Writing*. Digital Humanities 2019. July 2019

Benjamin Kiessling. “Kraken - A Universal Text Recognizer for the Humanities”. In: *Proceedings of the DH* (July 2019)

Benjamin Kiessling, Robin Tissot, Peter A. Stokes, and Daniel Stökl Ben Ezra. “eScriptorium: An Open Source Platform for Historical Document Analysis”. In: *2nd International Workshop on Open Services and Tools for Document Analysis, OST@ICDAR 2019, Sydney, Australia, September 22-25, 2019*. IEEE, 2019, p. 19. DOI: 10.1109/ICDARW.2019.10032

Benjamin Kiessling, Daniel Kinitz, Christoph Gümmer, and Parivash Mashhadi. “Script and Emphasis Detection using Recurrent Neural Networks”. READ2018: International Interdisciplinary Symposium on Reading Experience and Analysis of Documents. 2018

ORAL
COMMUNICATIONS
AND INVITED
TALKS

Benjamin Kiessling. “Why We don’t Postcorrect”. presented at the International Workshop on Error Correcting HTR, Ca’ Foscari University, Venice. Nov. 2022

Benjamin Kiessling. “New Developments in Kraken and eScriptorium”. presented at the Documents anciens et reconnaissance automatique des écritures manuscrites workshop, Ecole Nationale des Chartes, Paris. June 2022

Benjamin Kiessling. “eScriptorium : une plateforme d’applications web pour la transcription automatique de documents manuscrits ou imprimés”. presented at the Atelier Digit-Hum, ENS, Paris. Oct. 2021

Benjamin Kiessling. “Kraken - A universal text recognizer for the humanities”. presented at the Open Islamicate Texts Initiative Workshop, University of Maryland. Jan. 2020

Benjamin Kiessling. “The Limits to Digitization”. presented at the Dark Archives conference, Faculty of English, Oxford University. Sept. 2019

Benjamin Kiessling. “The intricacies of Arabic manuscript layout analysis”. presented at Digital Humanities Conference: Crowdsourcing and Citizen Science, University of Haifa. May 2019

Benjamin Kiessling. “The Kraken OCR engine for historical documents”. Workshop held at ManuSciences '19, Franco-German summer school. Mar. 2019

Benjamin Kiessling. “OCR-Processing in Leipzig’s OpenPhilology Project”. presented at 14th PhilTag meeting, KALLIMACHOS, University of Würzburg. 2017

Benjamin Kiessling. “Kraken and the Future of Arabic OCR”. Workshop held at the Genealogies of Knowledge group, University of Manchester. 2017

Benjamin Kiessling. “The State of Arabic Script OCR”. presented at meeting on the current state of Arabic OCR, Digital Library of the Eastern Mediterranean, Harvard University. 2017