

OCR4all is software that was developed for digital text indexing, primarily of works that were printed very early on. Since their print types and often complex layout concepts exceed the recognition capabilities of many other text recognition programs, it is necessary to be able to design the digitization process in a variable, corpus- and work-specific manner. The semi-automatic OCR workflow proposed in OCR4all, which is understandable and can be used independently, also addresses a dedicated non-IT user group and combines different work tools within a uniform user interface. The constant change between different programs is therefore no longer necessary. With **graduation** the second project phase of the BMBF-funded joint project **Kallimachos** the software is now being used at the **Center for Philology and Digitality** established by the university in order to make them permanently and freely available to as wide a user group as possible.

## workflow

From the pre-processing of the image files to be processed (so-called pre-processing) to the layout segmentation (so-called region segmentation with <https://github.com/OCR4all/LAREX>), line segmentation and text recognition (recognition with <https://github.com/Calamari-OCR>) to the correction of the recognized texts (Ground Truth Production) and the creation of plant-specific OCR models in a training module, OCR4all describes a fully-fledged OCR workflow.



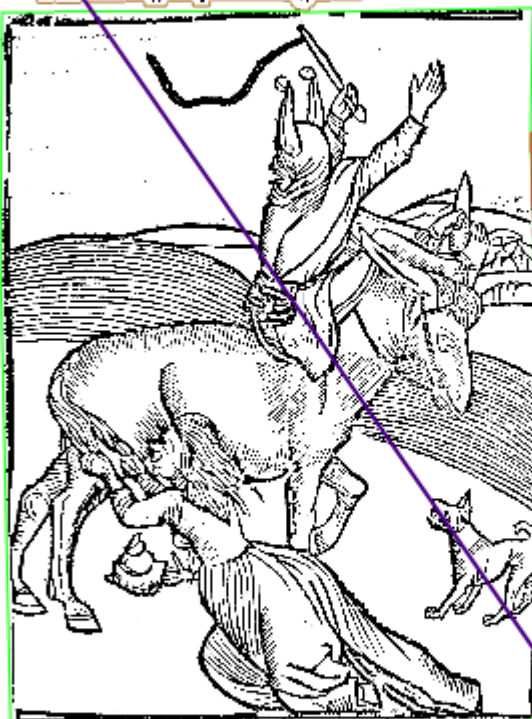
Above all, due to the possibility of producing and training plant-specific text recognition models, very good results can be achieved in digital text indexing with OCR4all for almost all printed texts.

7 pas du pain a manger & a l'adventure q  
gens/poures desues/sabres/auens  
dres compte devant celluy qui nous crea

Des conditiōs courroux et 2 grandes mauuaities des fēmes

Plusieurs asnes cheuancheroient  
Le nest que mo eust femme  
Au moyen de qu ne voudroient  
Monte dessus po euse diffame

Quilz ont fait & grief infame  
Au poure asne & grans tormēs  
Car luy ont toz tous les bons mēs



5 Ntendes folz e/  
flourdis & vous  
congnoistres les  
mauuaisties des  
fēmes. Aussi fēmes appro  
chez vo' & vo' oures bone ma  
tiere. Mes versetz ditez & es  
cripts Voultroiet des fēmes  
touliours dire grās louēges  
autāt que des anges bīuns/  
car de leurs natures sōt doul  
res/teiffors bīd sounēt se  
bāge penāt mauuaistie et  
sansce conditiōn Nō script  
ne poures differer/pourāt  
faulx il ēe escarpne des fem  
mes de l'apre/car de leur  
mauuaistie ie ne me parīan  
re Jay eyme & aymeray tant  
que ie l'auray lhōneur de tou  
tes bōnes fēmes/mais ptre  
les mauuaistes en tous mes  
escripts ie pleray. Cornelia  
la pāine mere de grāchus  
fut sage & pōtāt tā q sa grāt vtu & sagesse avec ses bōnes meurs redōit par tout  
les bōnes ie loueray/mais ie blasmeray les mauuaistes tā q au monde ie vi  
uray. La femme qui est alaitte de la māmelle de noblesse/ence/se elle est sa  
ge cest plus grāt tresor que d'auoir toutes les richesses du monde/elle ne se corēpt  
point facilement de vices ne de perchez. Le ceste fēme est courrouce a son mary  
facilemēt elle mitignera son courroux. Nous lisons du roy assuerus q auoit mā  
de mettre a occision les hebreux. Lors bester sceust ceste chose laquelle estoit regē  
te du pays/elle vint au roy & luy pria quil luy voulsist ottroyer sa priere. Il luy ot  
troya/et elle luy pna q l'enuocast sa sentence/laquelle chose fit/ & p ce moyen elle gar  
da le peuple de cruelle mort. Nabal estoit moult riche et vassal de dāuid/ cest roy

### Lucille

sp font desquelles ie me taie. O poures folz delistries qui de robes ceulz qui nōt  
pas du pain a manger & a l'adventure quilz nosent demander de honte les vieilles  
gens/poures desues/sabres/auengles/helas pense p/car certes vous en veni  
dres compte devant celluy qui nous crea

Des conditiōs courroux et grandes mauuaities des fēmes

Plusieurs asnes cheuancheroient  
Le nest que monte p eust femme  
Au moyen de quoy ne voudroient  
Monte dessus po euse diffame

Quilz ont fait & grief infame  
Au poure asne & grans tormēs  
Car luy ont toz tous les bons mēs  
Car luy ont toz tou les bons mēs



6 Ntendes folz e/  
flourdis & vous  
congnoistres les  
mauuaisties des  
fēmes. Aussi fēmes appro  
chez vo' & vo' oures bone ma  
tiere. Mes versetz ditez & es  
cripts Voultroiet des fēmes

Quilz ont fait & grief infame

Quilz ont fait 7 grief infame

Au poure asne & grans tormēs

Au poure asne 7 grans tormēs

Car luy ont toz tous les bons mēs

Car luy ont toz tou les bons mēs

E  
C

Ntendes folz e/

Ntendes folz e=

LOAD	SAVE
a b c d e f g h i j k l m n o p q r s t u v w x y z	
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32	
+	

## Cooperation with OCR-D

In the summer of 2020, a cooperation between OCR4all and the coordinated funding initiative for the further development of optical character recognition methods - <https://ocr-d.de> agreed upon. The main goal of the DFG-funded OCR-D project is the conceptual and technical preparation of the full text transformation of the prints published in the German-speaking area from the 16th to the 18th Century (VD16, VD17, VD18). For this purpose, the automatic full text recognition, analogous to the OCR4all approach, is broken down into individual process steps that can be reproduced in the open source OCR-D software, with the aim of creating optimal workflows for the old prints to be processed and thus scientifically usable full texts to generate.

The subject of the cooperation is, in addition to the ongoing exchange, above all about interfaces, scalable software implementations, creation and provision of GT and upcoming developments in the OCR area, a technical convergence of the two projects. OCR4all will implement the OCR-D specifications in its OCR solution and implement interfaces to OCR-D tools. If OCR4all relies internally on OCR-D solutions, OCR4all users benefit from the expanded selection of tools and the associated possibilities, while OCR-D has a greater range and, thanks to simplified access, new user groups both inside and outside of VD mass digitization achieved.

## Reporting (selection)

- einBlick: Recognizing **historical writings digitally**
- Radio interview SWR2 Impulse: Medieval manuscripts become text documents
- **tool presentation** as part of the DFG-funded project **forText**
- dr Johann Ramming: **ocr4all - OCR for Incunables**
- Der Tagesspiegel: Computer tool for old texts
- The Standard: **Reliable text recognition tool for historical publications**
- Augsburger Allgemeine: **Computer reads old texts**
- einBlick: **Modern tool for old texts**

## citation

If you use OCR4all, please cite the associated **paper**:

## funding

- DFG funding initiative "OCR-D" [↗](#), (phases 2 and 3)
- Center for Philology and Digitality (ZPD) [↗](#), University of Wuerzburg
- BMBF funded project Kallimachos [↗](#)
- Chair of Artificial Intelligence and Knowledge Systems [↗](#), University of Wuerzburg