

Транскрибус колаборатива

задружним радом до
рашчитаних рукописа
Европе



Transkribus

Развој Транскрибуса

- Транскрибус је осмишљен и покренут 2015. године у оквиру пројекта под називом tranScriptorium (2013–2015) уз финансијску подршку Европске комисије
- имплементација и развој Транскрибуса су били у фокусу следећег, наследног пројекта READ (Recognition and Enrichment of Archival Documents), тј. Пројекта за препознавање и обогаћивање архивских докумената, финансираног од стране Европске комисија у периоду 2016-2019.

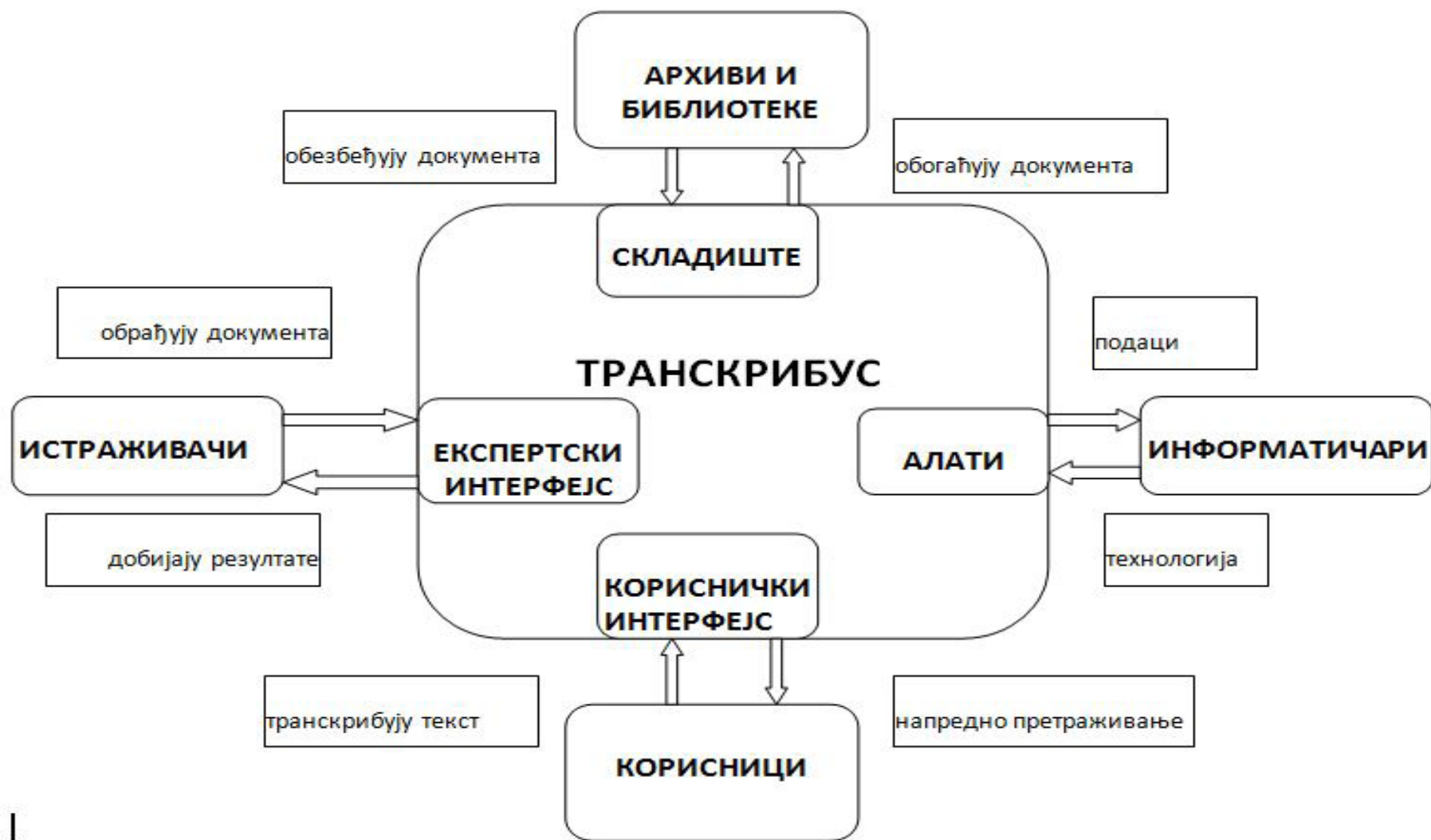
Циљ READ пројекта

- одржавање,
- развој и
- промоција Транскрибуса
- Транскрибус = функционална мрежна истраживачка инфраструктура у којој нове технологије могу да подстакну иновације у архивским истраживањима.

READ пројекат

- координатор: Група за дигитализацију и дигиталну заштиту (нем. Digitalisierung & Elektronische Archivierung - DEA) са Универзитета у Инсбруку коју је предводио др Гинтер Милбергер
- сарадници четрнаест европских универзитета истраживачких институција.

Главни актери, њихов допринос и могући бенефити



READ-COOP SCE

- Завршетком READ пројекта омогућено је неметано коришћење платформе Транскрибус преко новоустановљене организације
 - READ-COOP SCE или, краће речено READ колаборатива.
- READ колаборатива је основана је 9. новембра 2019, као европско задружно друштво са седиштем у Инсбруку у Аустрији на иницијативу неколико чланова READ пројекта ради одржања и даљег развоја платформе Транскрибус.

Укључивање библиотека и архива у Транскрибус колаборативу

- Организација је отворена је све који су заинтересовани за дигитализацију, транскрипцију, препознавање текста и претрагу историјских докумената.
- Чланство у READ колаборативи се стиче попуњавањем Изјаве о чланству (енгл. Declaration of membership), а све неопходне информације могу се пронаћи на адреси: <https://readcoop.org/join>

Деонице и чланарина

- Попуњавањем приступнице чланови аутоматски постају и деоничари.
- Деонице READ колаборативе се добијају једнократном уплатом и представљају обавезан улог сваког члана, који може бити враћен ако члан одлучи да напусти колаборативу.

Потребан износ за правна лица

- 1000 ЕУР (минимум);
 - ова уплата доноси 1 глас на Генералној скупштини.
- 2000 ЕУР;
 - ова уплата доноси 2 гласа на Генералној скупштини.
- 3000 ЕУР;
 - ова уплата доноси 3 гласа на Генералној скупштини.
- 4000 ЕУР;
 - ова уплата доноси 4 гласа на Генералној скупштини.
- 5000 ЕУР (максимум);
 - ова уплата доноси 5 гласова на Генералној скупштини.

Потребан износ за физичка лица:

- 250 ЕУР;
 - ова уплата доноси 1 глас на Генералној скупштини.
- **ВАЖНО**
- Након истека прве године, плаћа се годишња чланарина која износи 25% положеног улога (тј. 62,5,00 ЕУР за приватна лица и минимално 250,00 ЕУР за правна лица на годишњем нивоу).

Стипендије

<https://www.transkribus.org/scholarship>

For Students

With a Transkribus Scholarship you can get free credit packages for Transkribus handwritten text recognition for your thesis project. Just send us a short description of the bachelor, master or PhD project you are working on and provide some background information about yourself.

Apply Now

For Teachers

To help the next generation of digital scholars as much as we can, we also aim to support those who teach them with free credits. Feel free to send us a short description of your current or upcoming course, webinar, programme or workshop, regardless of the kind of school, university or event, and provide some background information about yourself.

Apply Now

Нови модел финансирања платформе Транскрибус

- У складу са одлуком Генералне скупштине READ колаборативе, од 19. октобра 2020. године поједине услуге на Транскрибус платформи се наплаћују будући да је пројектно финансирање окончано, а неопходна су новчана средства за одржавање и унапређење коришћене технологије.
- <https://www.transkribus.org/plans>
- Модел финансирања се заснива на кредитима који се могу купити директно на онлајн продавници. Ови кредити представљају валуту плаћања за услуге на Транскрибус платформи.

Нови модел финансирања платформе Транскрибус

Individual

0 €

Ideal for Genealogists & Students
/month incl. 20% VAT*
Credits available **per month**

- ✓ AI Text Recognition
- ✓ Custom AI Model Training
- ✓ Powerful Document Editor

Start for free

Scholar

14.9 €

Tailored for Professionals
/month incl. 20% VAT*
Credits available **per month**

- ✓ Advanced processing speed
- ✓ Advanced AI Tools
- ✓ Transkribus Sites (1000 pages)

Choose

Organisation

—

For Research & Cultural Institutions
Tailored to your needs

- ✓ User Seats & Management
- ✓ API Access
- ✓ Success Team

Get in Touch

Нови модел финансирања платформе Транскрибус

- Наплата се врши само за аутоматско генерисање транскрипта на приложеним документима, док су све остале услуге попут аутоматске сегментације и израде модела бесплатне.
- Омогућено је и бесплатно рашчитавање стотињак страница додељивањем 100 кредита месечно сваком кориснику понаособ.

Features & Allowances

AI Text Recognition



Custom AI Model Training



Powerful Document Editor



Fulltext Search



Број регистрованих корисника платформе Транскрибус

150 000+

Registered Users

20 000+

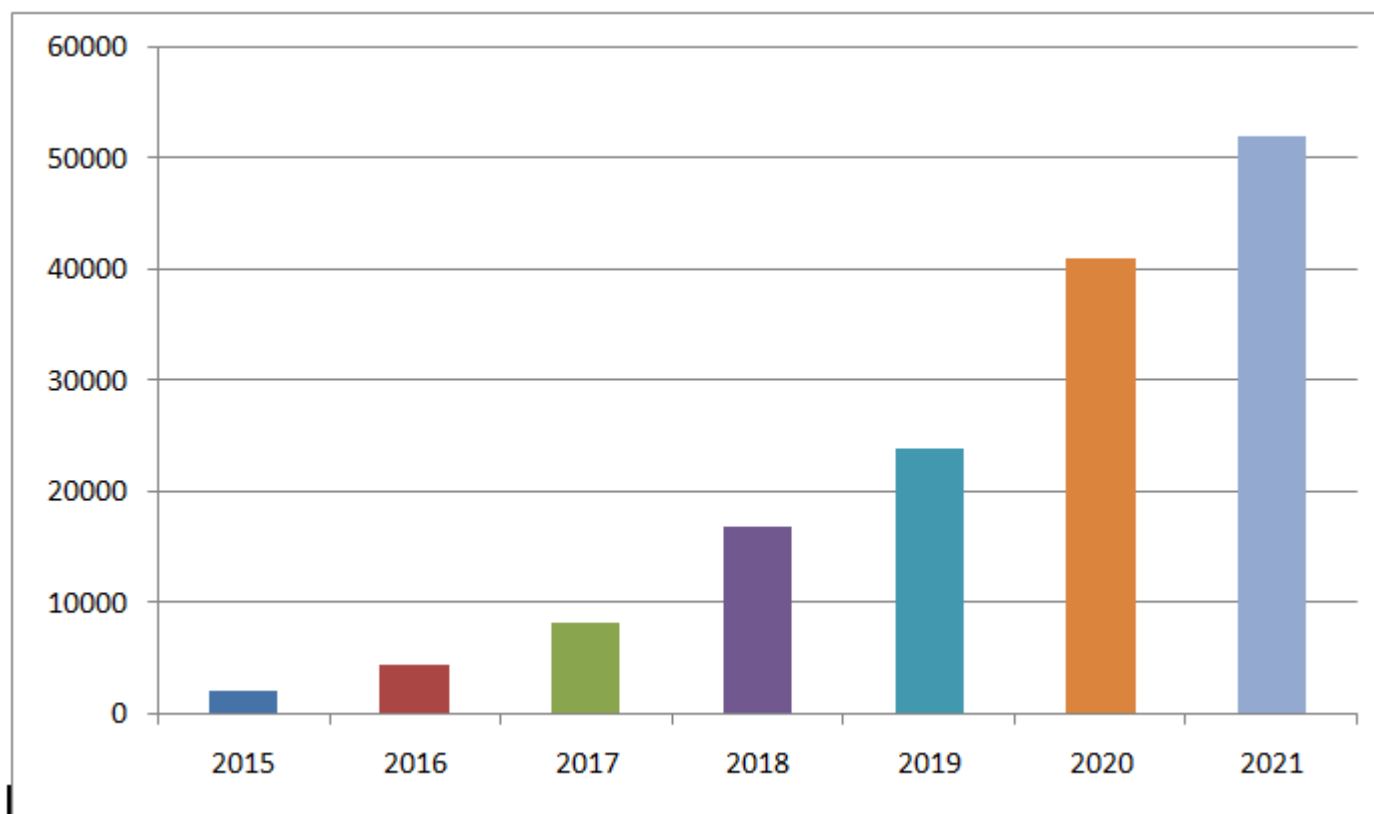
HTR AI Models trained

50+ Million

Pages processed

150+

Free Public AI Models



Дигитализација путем мобилног телефона

- истраживачи у лабораторији [Computer Vision Lab](#) на Техничком универзитету у Бечу као део Н2020 пројекта READ развијају:
 - мобилну апликацију DocScan и
 - преносни уређај ScanTent

DocScan

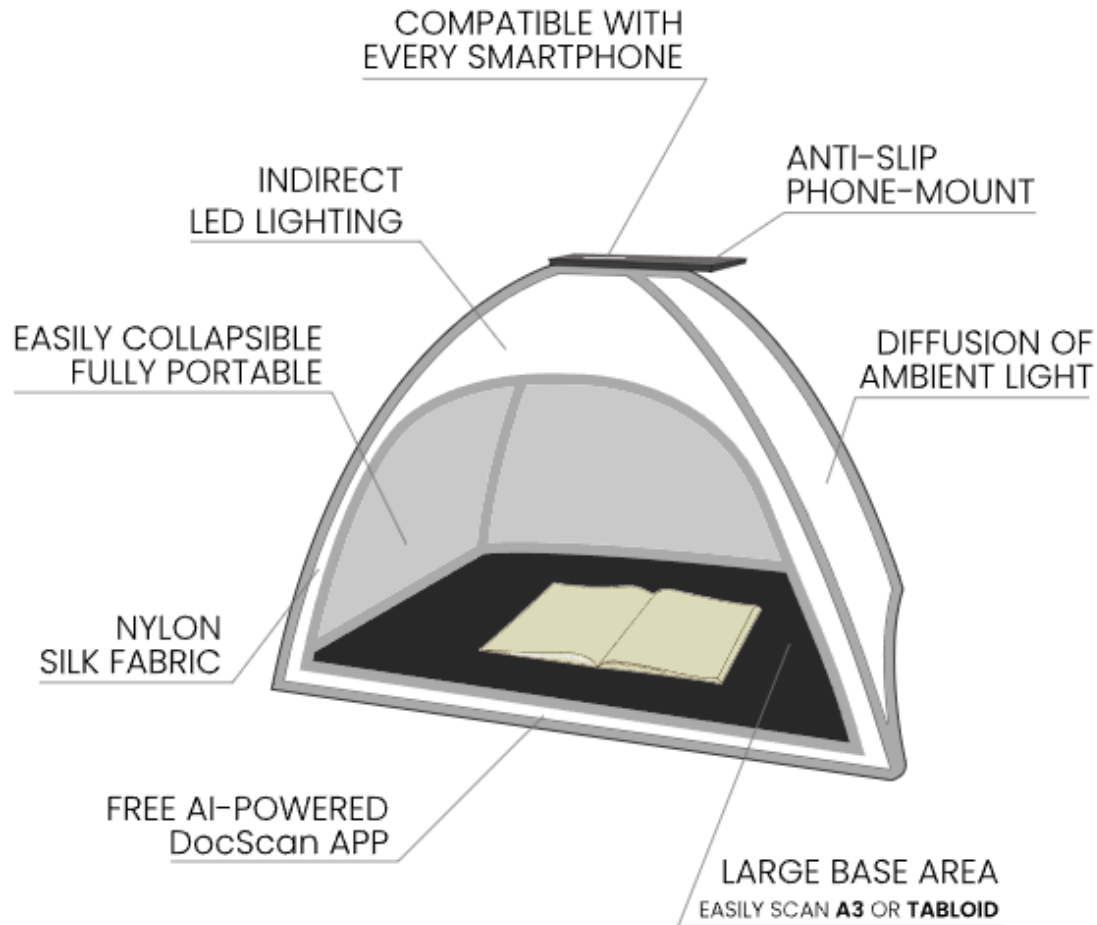
- DocScan је мобилна апликација за скенирање докумената дизајнирана за ScanTent
- Садржи приказ уживо и детектује страницу документа у реалном времену и врши процену квалитета снимка (прилагођавање фокуса)
- Омогућен је режим рада у серији, тј. аутоматско фотографисање приликом окретања странице
- Апликација је отвореног кода и доступна је на платформи GitHub (<https://github.com/TUWien/DocScan>), као и на Google Play продавници (<https://play.google.com/store/apps/details?id=at.ac.tuwien.caa.docscan>)

ScanTent

- Преносни уређај ScanTent:
 - блокира амбијентално светло,
 - обезбеђује недеструктивно лед светло и
 - функционише као постоље за мобилни телефон када је у режиму рада у серији
- Корисници једноставно треба само да прелиставају страницу по страницу књиге како би је дигитализовали, без притискања разних команди за преузимање, сетовање или меморисање страница.

ScanTent

Мана: цена
239,90 Еура



Транскрибус

- Три главне фазе у рашчитавања документа:
 - Постављање дигитализованих слика на платформу у форматима PDF, JPEG или PNG.
 - Коришћење алата за сегментацију дигитализованих слика у редове
 - Тачно преписивање текста који се налази у сваком од сегментираних редова на дигитализованој слици.

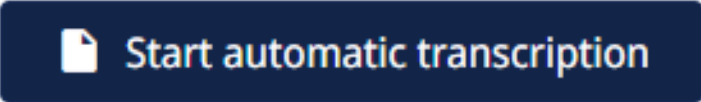
Фаза 1: Отпремање слика на Транскрибус

- Увоз докумената са рачунара
- Дозвољени формати:
 - JPG
 - PNG
 - PDF

Мах 200 MB

Напомена: слике остају приватне, везане су искључиво за кориснички налог и нису јавно доступне.

Фаза 2: Сегментација

A dark blue rectangular button with a white document icon on the left and the text "Start automatic transcription" in white.A dark blue rectangular button with a white rounded rectangle in the center containing the text "Text Recognition".A dark blue rectangular button with a light blue rounded rectangle in the center containing the text "Layout".

- Коришћењем алата за сегментацију дигитализованих слика у редове тзв. Layout Analysis стварају се :
 - текстуални блокови који служе да означе сваки део или пасус текста,
 - област линија око сваког реда текста и
 - основне линије које се протежу при дну сваког реда текста.

Одабрати опцију:

Horizontal Line Orientation

Фаза 3: Транскрипција

- Основна правила транскрипције су:
 - текст се прекуцава доследно, пратећи дати редослед, онако како се види на страници;
 - нема исправљања правописних и других грешака;
 - велико слово остаје велико, специјални карактер као специјала.

Сегментација и транскрипција докумената

- *Сегментација* – свака слика мора бити подељена (сегментирана) на текстуалне блокове, на блокове редова и на основне линије. Изводи се аутоматски уз помоћ алата Layout Analysis
- *Транскрипција* – текст се прекуцава доследно, пратећи дати редослед, онако како се види на страници без исправљања правописних и других грешака.
- Сегментација и транскрипција омогућавају обучавање AI модела за аутоматско читање историјских докумената.

Референтни скуп података – ground truth

Урађена транскрипција референтних података
погодна је за:

- креирање сета података који омогућава систему за препознавање рукописног текста (AI) да научи да дешифрује рукописна документа;
- покретање AI модела на приложеним документима и добијање аутоматски генерисаног транскрипта;
- претраживања пуног текста коришћењем методе “проналажење кључних речи” која даје неупоредиво боље резултате у односу на стандардне претраге пуног текста;
- припрему докумената за научна издања.

Коришћење постојећих модела

- <https://www.transkribus.org/public-models>
- 100 доступних модела за различите језике

Демократизација дигитализације културног наслеђа

- Од 2018. године одржан је велики број радионица за библиотекарe и архивистe
- Више на:

<http://unilib.rs/transkribus/>

Радионица

- <https://www.transkribus.org/>
- <http://arhiva.unilib.rs/cirilica/zbirka/knjiga/rukopisi-isidore-sekulic>
- <https://pretraziva.rs/pretraga>

Хвала на пажњи!!!

доц. др Адам Софронијевић

sofronijevic@unilib.rs

др Наташа Дакић

dakic@unilib.rs

др Александра Тртовац

aleksandra@unilib.rs