

Հիմնարար գիտական գրադարան  
29, Մարտ, 2024 թ.



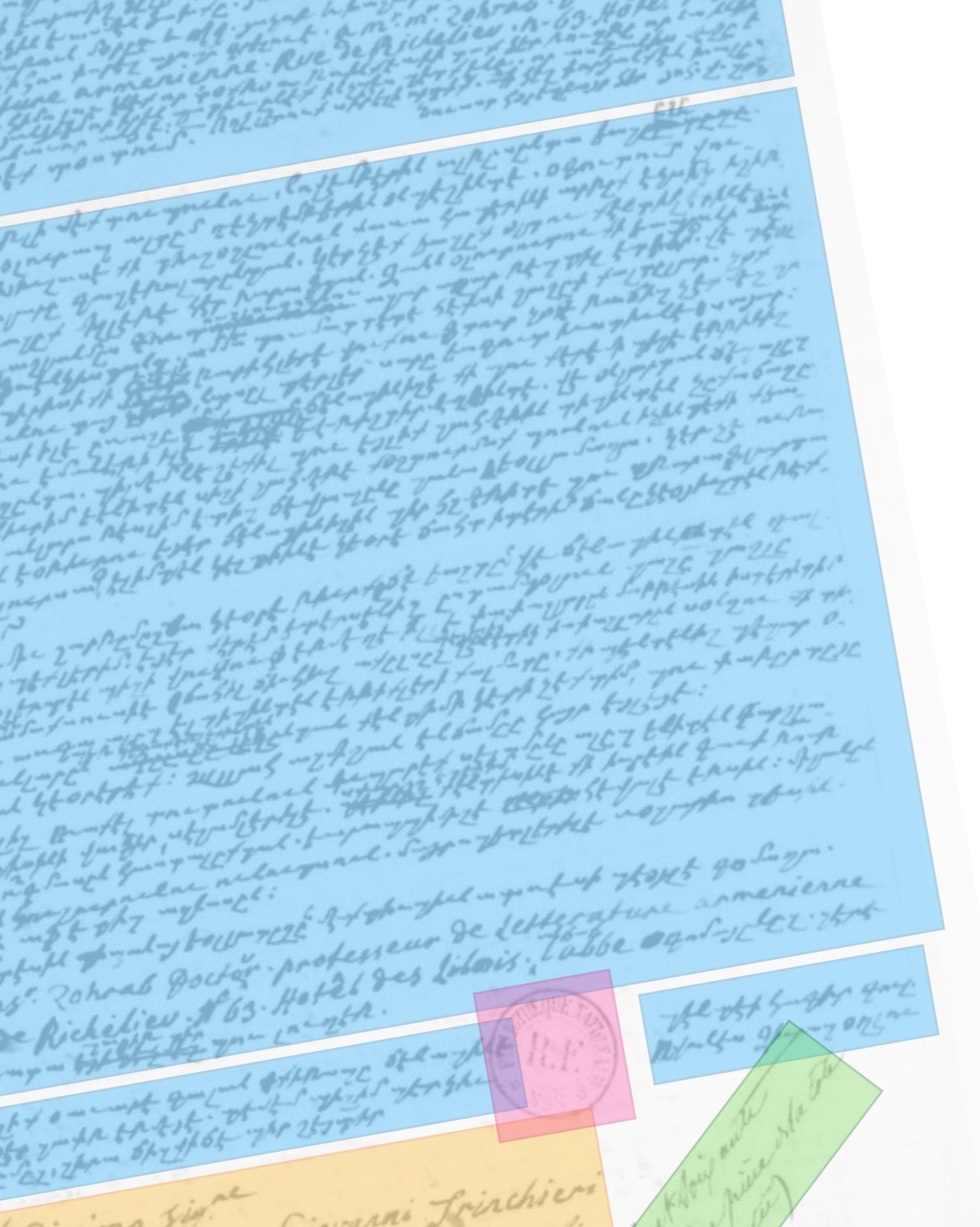
# Գրադարաններ և արհեստական բանականություն (AI)

## Օգտագործել AI-ն հայկական հավաքածուները զարգացնելու համար

Շահան Վիդալ-Գորեն  
Calfa



CALFA

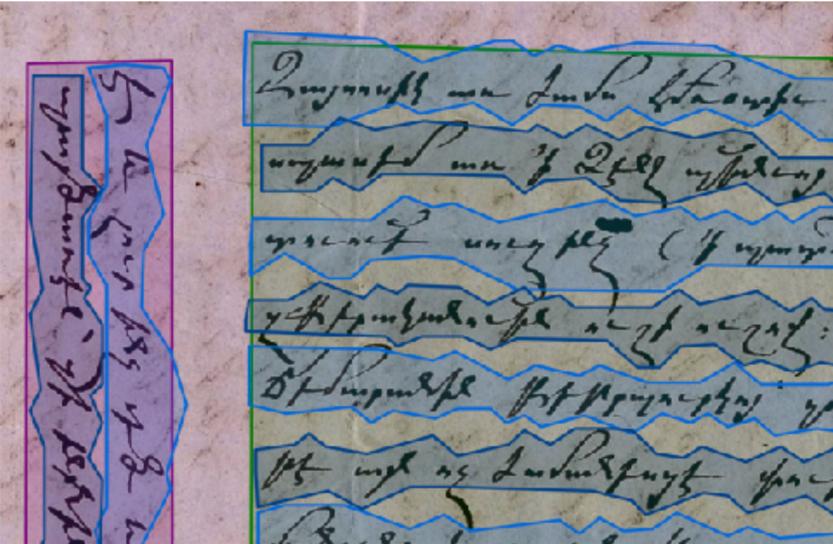




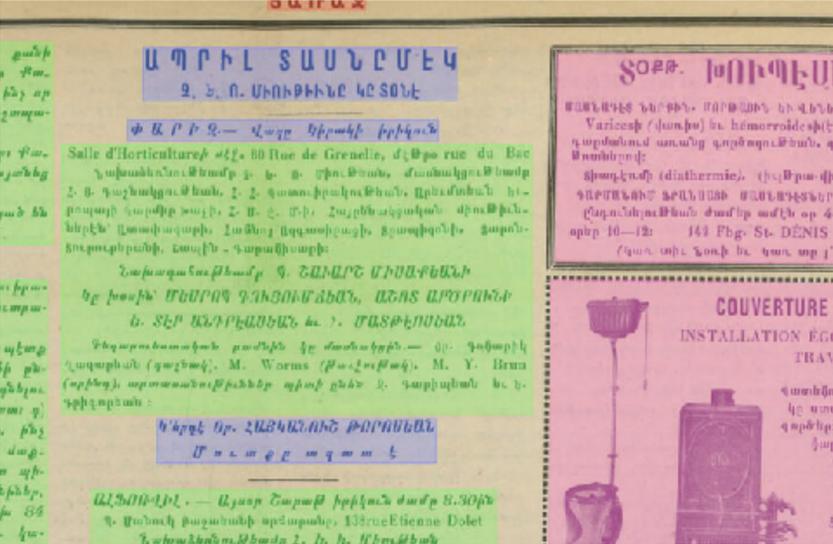
# CALFA

OCR-ի (Optical Character Recognition) և արհեստական բանականության (AI) ճարտարագիտությունները արելելեան լեզուների եւ հայերէնի բնագիրներու ճանաչման համար

- ➔ Նուիրուած ժառանգութեան պահպահման և զարգացման:
- ➔ Գլխաւոր նպատակ՝ դիրացնել հանրութեան և գիտաշխատողներու համար փաստաթղթերի և գիտելիքների հասանելիութիւնը:
- ➔ Մասնակցել թուայնացման նախագծերին և տուեալների շտեմարանի ծարգացման:
- ➔ Շարունակական մշակում՝ արաբական և հայկական հին ձեռագիրների, հայկական ժամանակակից ձեռագիր փաստաթղթերի և թերթերի:

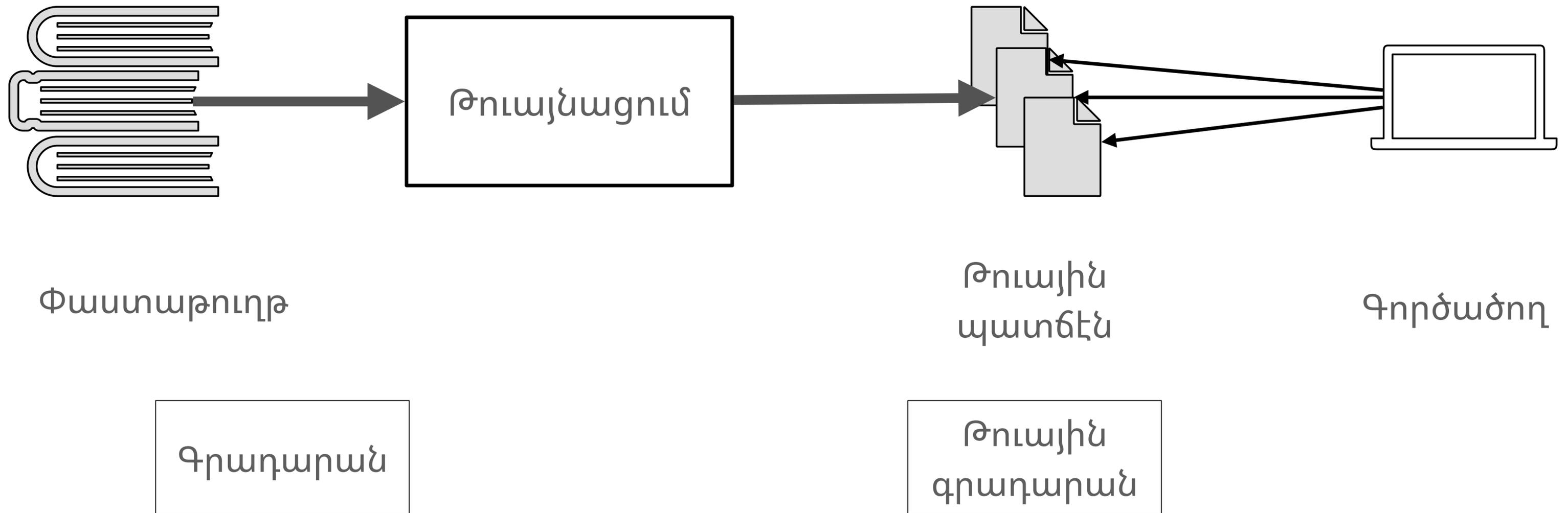


2022, Armenian, Mekhitarist Congregation



2021, Armenian, Fundamental Scientific Library

# Թուայնացնել, որպեսզի հայկական ժառանգությունը պահպանել եւ զարգացնել



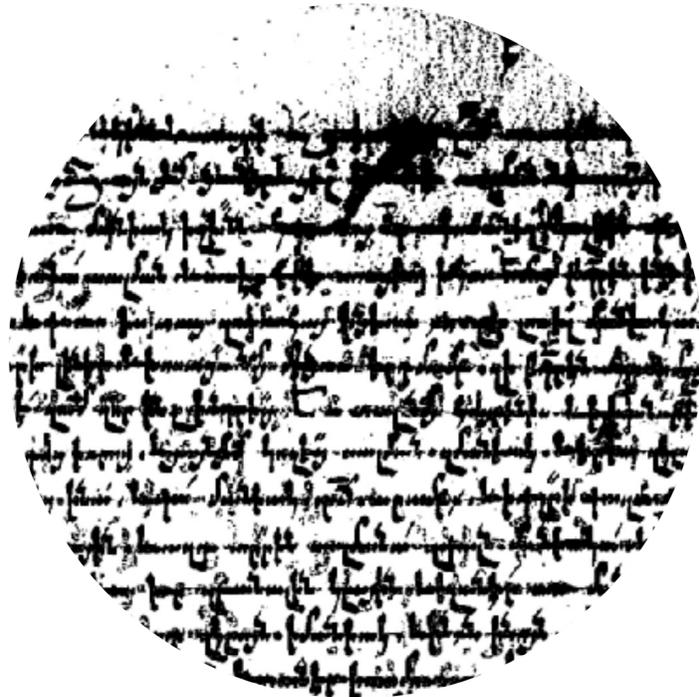
Օրինակի համար:

- Pan-Armenian Digital Library (FSL): <https://arar.sci.am/dlibra>

- Armenian newspapers (NLA): <http://tert.nla.am/>

- Armenian manuscripts, see the index of Digitized Armenian manuscripts (Calfa): <https://www.armenian-manuscripts-index.com/>

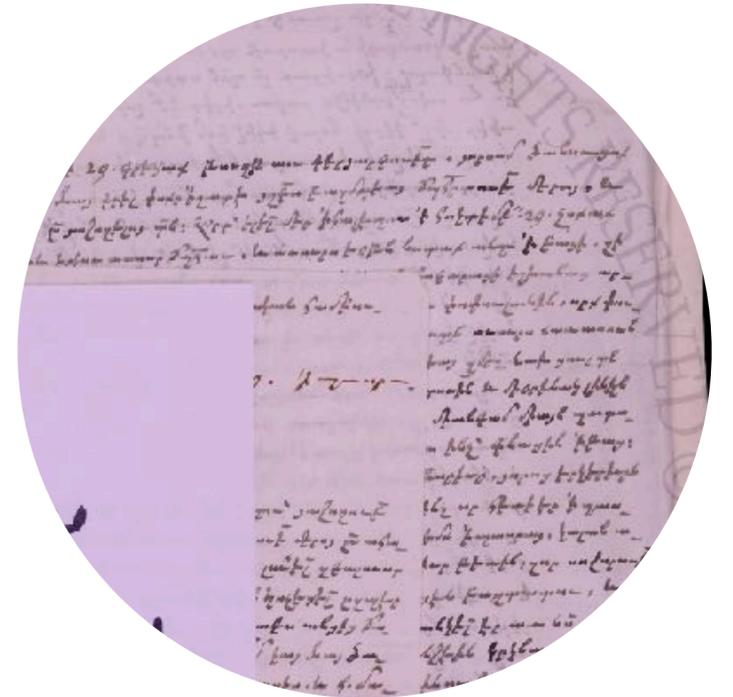
# Թուայնացնել, որպէսզի հայկական ժառանգութիւնը պահպանել եւ զարգացնել



Ձեռագիր փաստաթղթերը պահպանել պատճէներ օգտագործելով



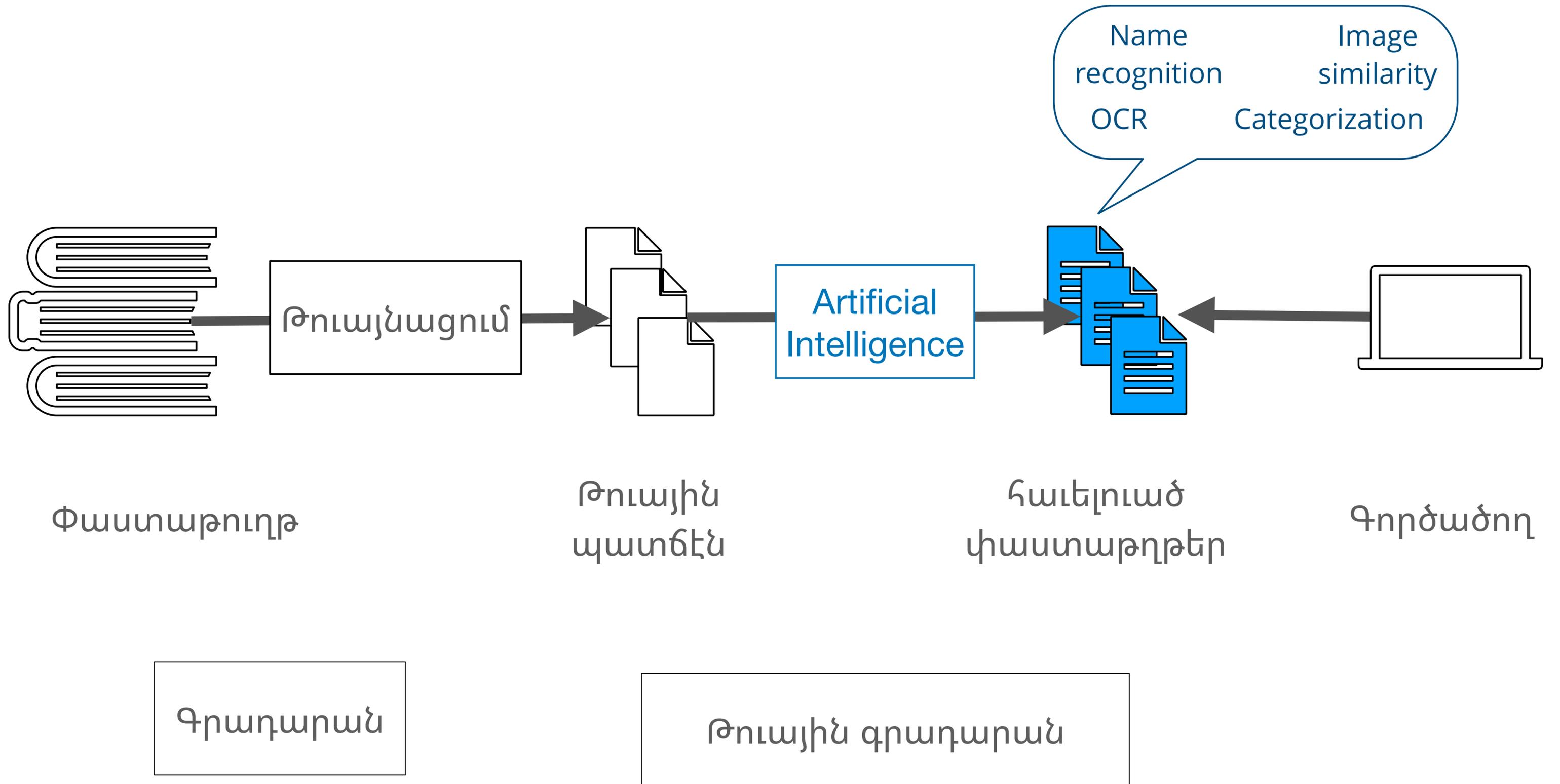
Ժառանգութիւնը մատչելի և տրամադրելի դարձնել կայքէջների միջոցով



Հայկական մշակույթը զարգացնել

Խնդիր՝ դիւրին չէ թուային պատկերի հետ աշխատել: Իսկ եթէ առարկան արուած է, լինի ձեռագիր կամ հին (գրաբար), կարդալը դժուար կարող է լինել (բացի եթէ ուսում ենք ստացել կարդալ իմանալու համար):

# Արհեստական բանականության ոգտագործել տեղեկությունները գտնելու համար



# Ներկայացման ժամանակ մտքին պահել՝

1

**Արհեստական բանականություն = որոշմանը օգնելու գործիք**

Արհեստական բանականությունը թուային հումանիտար գիտություններում միայն որոշում կայացնելու օգնութեան գործիք է, որն մեքենականացնում է որոշակի առաջադրանքներ, որոշակի ճշգրտութեամբ: **Երբեք կատարեալ չէ: Բայց արագ է:**

2

**Հետազոտողներն ու գրադարանավարները AI-ի նոր մոդելների մշակման և սահմանման գործընթացի հիմքում են**

Գրադարանավարները որոշում են, թե ինչպես ստեղծել և ինչպես օգտագործել AI-ն և ինչ կարիքների համար:

3

**Ընդհանուր մոդել չկայ, միայն մասնագիտացուած մոդելներ կան (արեւելեան գրություններ)**

Բայց շատ հեշտ է մոդելը մասնագիտացնել նոր առաջադրանքի, նոր ձեռագրի վրայ, և այլն:

# OCR (Optical Character Recognition)

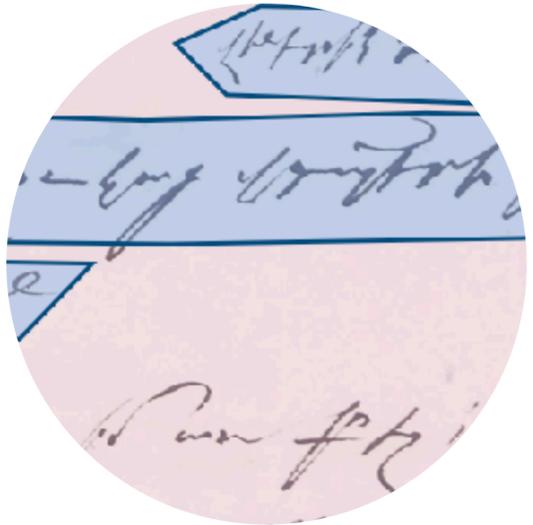
---

Գրանշանների օպտիկական ճանաչում

# Գրանշանների ճանաչում՝ ինչո՞ւ:



Մատչելի լինել փաստաթղթերի հասանելիությունը



Փաստաթղթերից յատուկ տեղեկություններ քաղել եւ գիտությունը զարգացնել



Գիտութեան պահպանումը եւ բացայայտումը ապահովել



# Տպագիր փաստաթղթեր - Գրանշանների օպտիկական ճանաչում (OCR)

Գրիգորիս Աղթամարցի, Տաղեր, p.109  
(Wikisource)

Ծաղիկ եմ տընկեր այգոյս,  
Կանաչ ու դեղին ծաղկոյս,  
Դեռ չեմ զհոտն առներ ծաղկոյս,  
Կասեն, թ' «Արե՛կ, ե՛լ այգոյս»:  
Այս [իմ նորաշէն տներոյս]:

Abby FineReader PDF (15.0.3)

Ծաղիկ եմ տընկեր այգոյս,  
Կանաչ ու դեղին ծաղկոյս,  
Դեռ չեմ զհոտն առներ ծաղկոյս,  
Կասեն, թ' «Արե՛կ, ե՛լ այգոյս»:  
Այս [իմ նորաշէն տներոյս]^:

CER: 2.25%

Մի քանի կետադրական նշանների բաց թողում

Tesseract (4.0)

Ծաղիկ եմ տընկեր աղգոյս,  
Կանաչ ու դեղին ծաղկոյս,  
Դեռ չեմ զոտն առներ ծաղկոյս,  
Կասեն, թ' «Արե՛կ, եղ ալգոյս»:  
Այս [իմ նորաշէն տներոյս]:

CER: 6.70%

Տառատեսակի նուազ ճանաչում (յ > լ)

Calfa (03/2023)

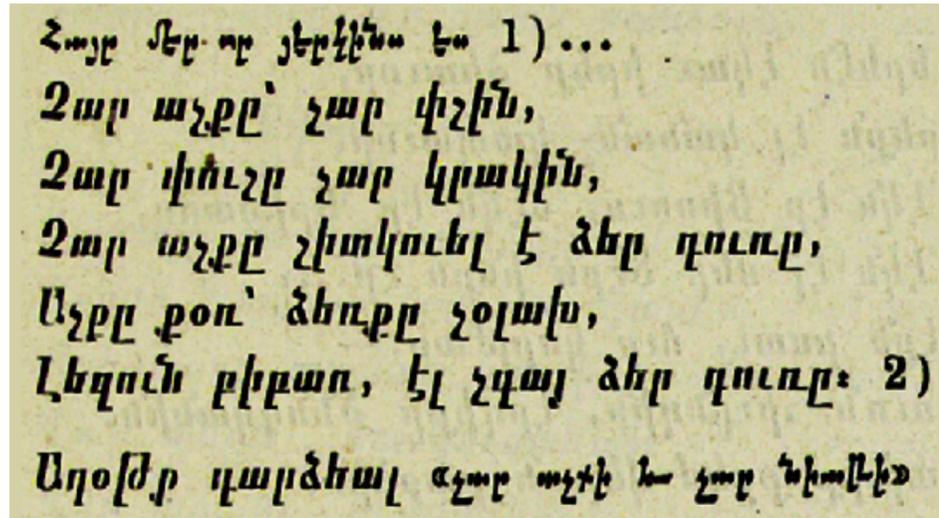
Ծաղիկ եմ տընկեր այգոյս,  
Կանաչ ու դեղին ծաղկոյս,  
Դեռ չեմ զհոտն առներ ծաղկոյս,  
Կասեն, թ<sup>1</sup> «Արե՛կ, ե՛լ այգոյս»:  
Այս [իմ նորաշէն տներոյս]]:

CER: 1.50%

Մի քանի կետադրական նշանների բաց թողում

# Տպագիր փաստաթղթեր - Գրանշանների օպտիկական ճանաչում (OCR)

Էմիլյան ազգագրական  
ժողովածու, 2, 1901, p. 183 (FSL of NAS RA)



## Abbyy FineReader PDF (15.0.3)

Հէք «ջ [redacted] է\* 1) ...  
Չար աչքը չար փշին,  
Չար փուշը չար կրակին,  
Չար աչքը շիտկուել է ձեր դուռը»  
Աչքը քօռ՝ ձեռքը չօլախ  
Լեզուն բիրառ, էլ չգայ ձեր գուռը: 2)  
Աղօթք դարձեալ [redacted] Ա

CER: 32.99%

- Անսովոր տառատեսակների և ոճերի թույլ ճանաչում

## Tesseract (4.0)

Հայր մէր որ յերկինս է» 1)...  
Չար աչքը՝ չար փշին,  
Չար փուշը չար կրակին,  
Չար աչքը շիտկուն է ձեր դուռը,  
Աչքը քօռ՝ ձեռքը չօլախ,  
Լեզուն բիրառ, էլ չգայ ձեր դուռը: 2)  
Աղօթք դարձեալ «որ «չի է չոր Ֆիանի»

CER: 14.72%

- Շատ զգայուն աղմուկի նկատմամբ
- Անսովոր տառատեսակների կամ ոճերի բաց թողում

## Calfa (03/2023)

Հայր մէր որ յերկինս է» 1)...  
Չար աչքը՝ չար փշին,  
Չար փուշը չար կրակին,  
Չար աչքը շիտկուել է ձեր դուռը,  
Աչքը քօռ՝ ձեռքը չօլախ,  
Լեզուն բիրառ, էլ չգայ ձեր դուռը: 2)  
Աղօթք դարձեալ «չոր աչքի և չար նիաթի»

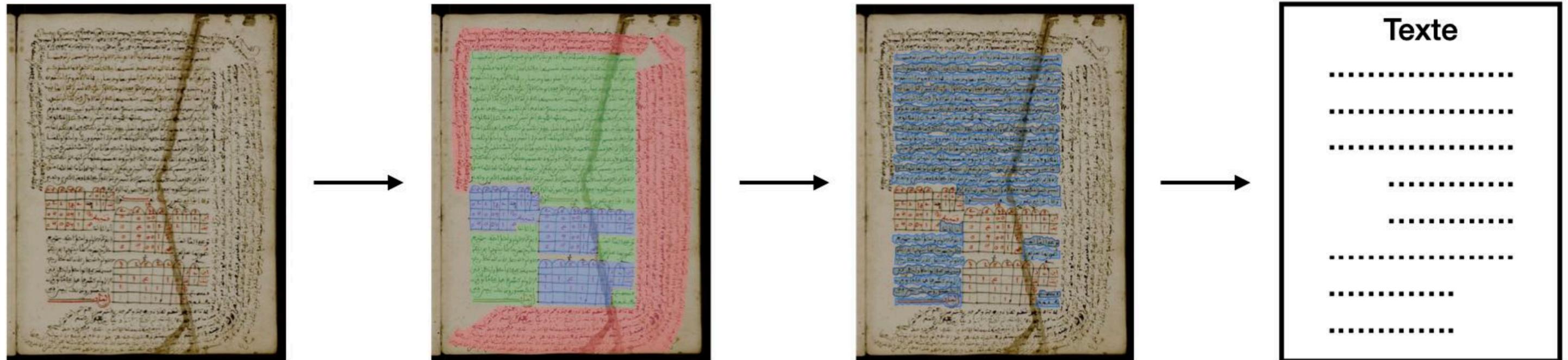
CER: 2.03%

# Ձեռագիր փաստաթղթեր - Ձեռագիր տեքստի ճանաչում (HTR)

## HTR. Ձեռագրերի և փաստաթղթերի ընկալման համար տեքստի ճանաչման համակարգ

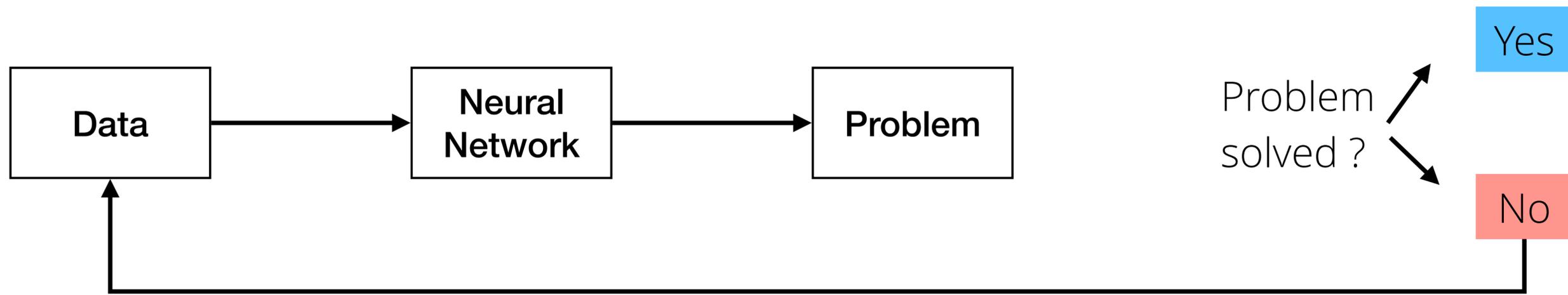
Ուսումնասիրության բաց դաշտ՝ արհեստական բանագիտության օգնությամբ (նեյրոնային ցանցեր) ժամանակակից համակարգեր, որոնք «հեշտությամբ» կարող են հասնել 5% CER-ի, յատկապես լատինատառ գրությունների համար

HTR-ն այժմ հասարակաց փուլ է ցանկացած թուային հումանիտար գիտությունների նախագծում լատինատառ գրությունների համար



# Ձեռագիր փաստաթղթեր - Ձեռագիր տեքստի ճանաչում (HTR)

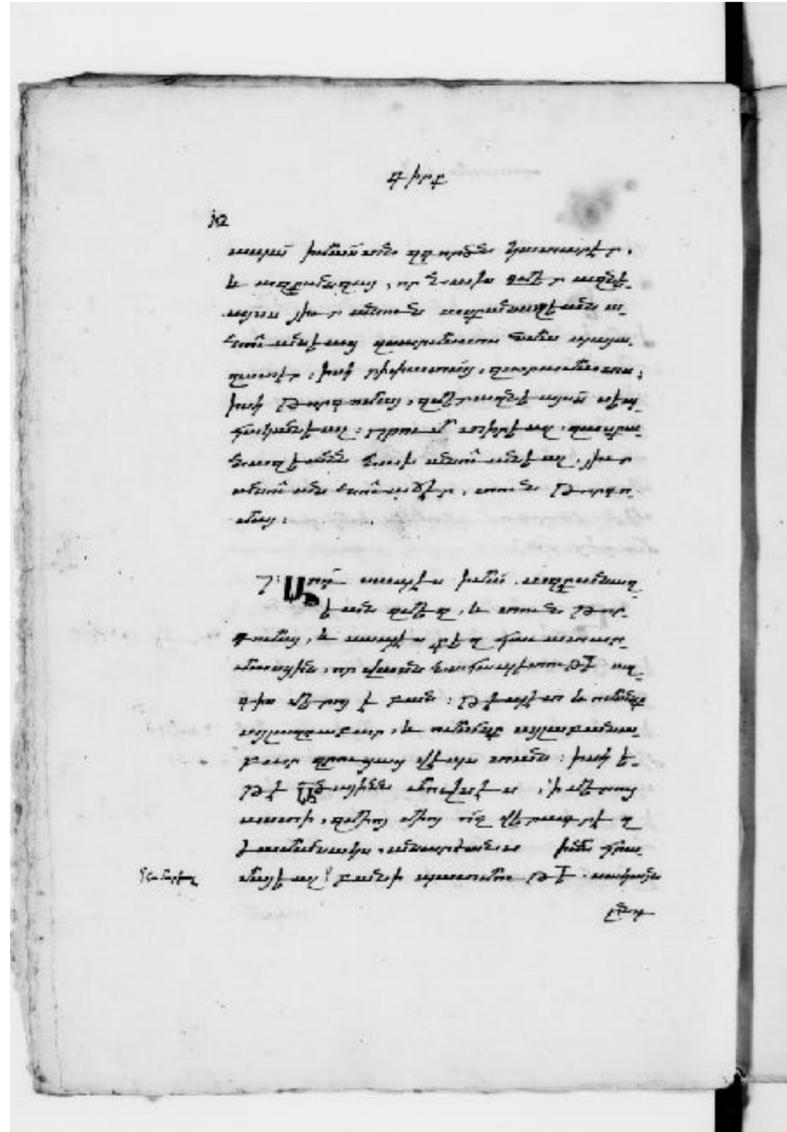
Արհեստական բանագիտությունը մշակել  
օրինակի միջոցով



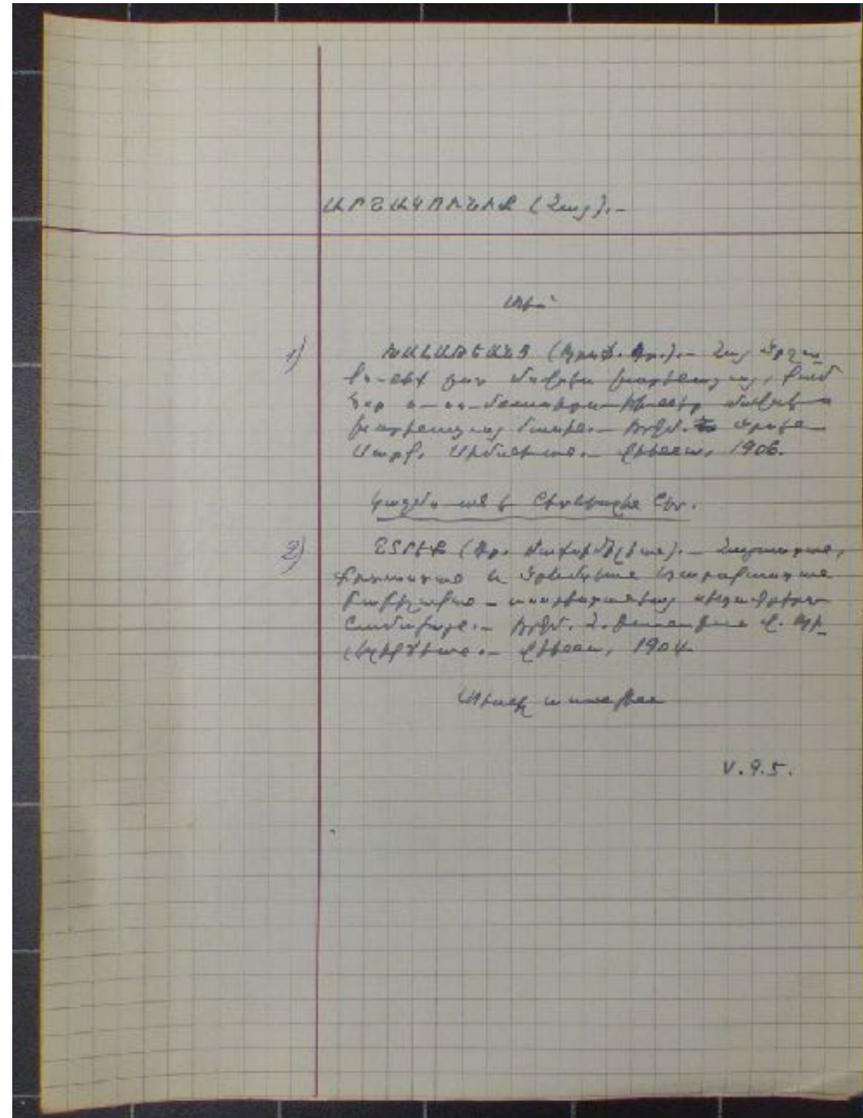
Ի՞նչ է պատահում, եթե իմ  
փաստաթուղթը վնասուած է / ձեռագիրը  
չափազանց յատուկ է / էջադրումը  
չափազանց բարդ է / եթե կարելի չէ  
ավելի շատ տուեալներ տրամադրել:

Ընդհանուր մօտեցումը  
պատշաճ չէ:  
Պահանջում է  
մասնագիտացուած  
մօտեցում:

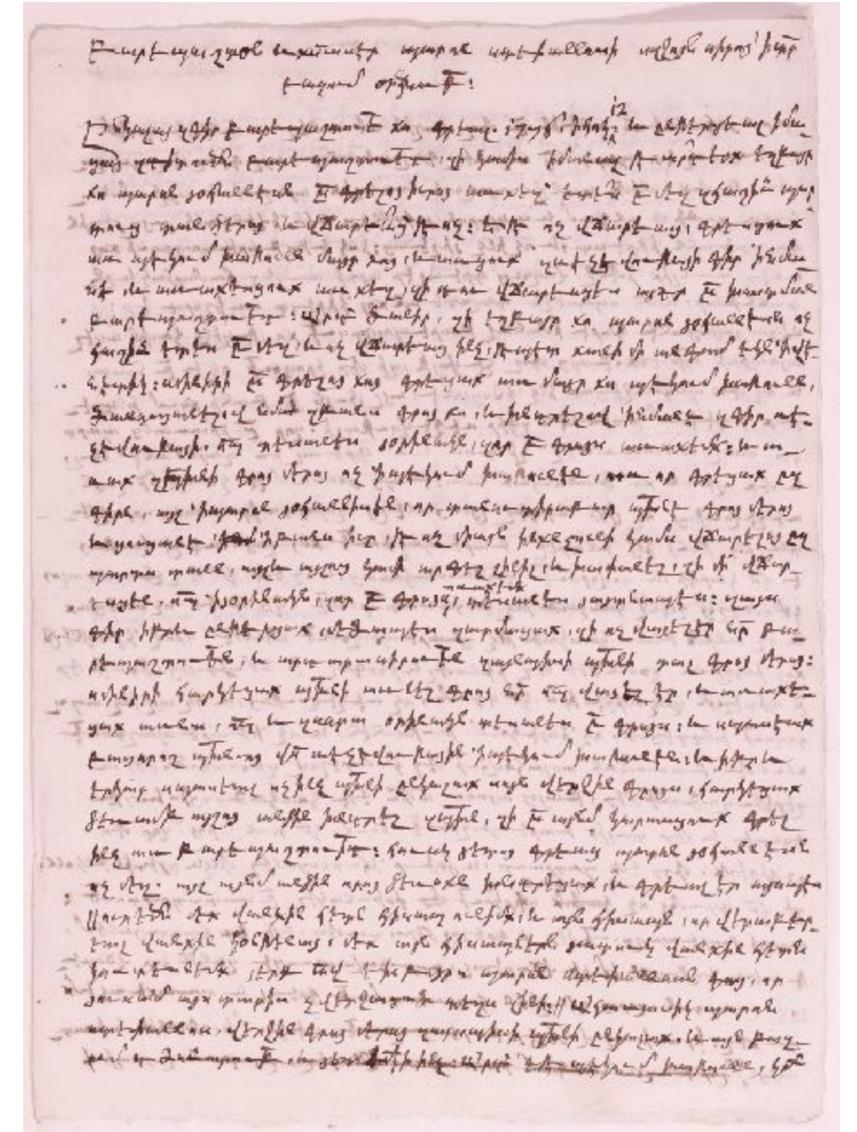
# Մեր ծրագիրները (2023-2024)



6.000 էջ  
Searchable on Gallica



60.000 էջ  
Searchable on Intranet

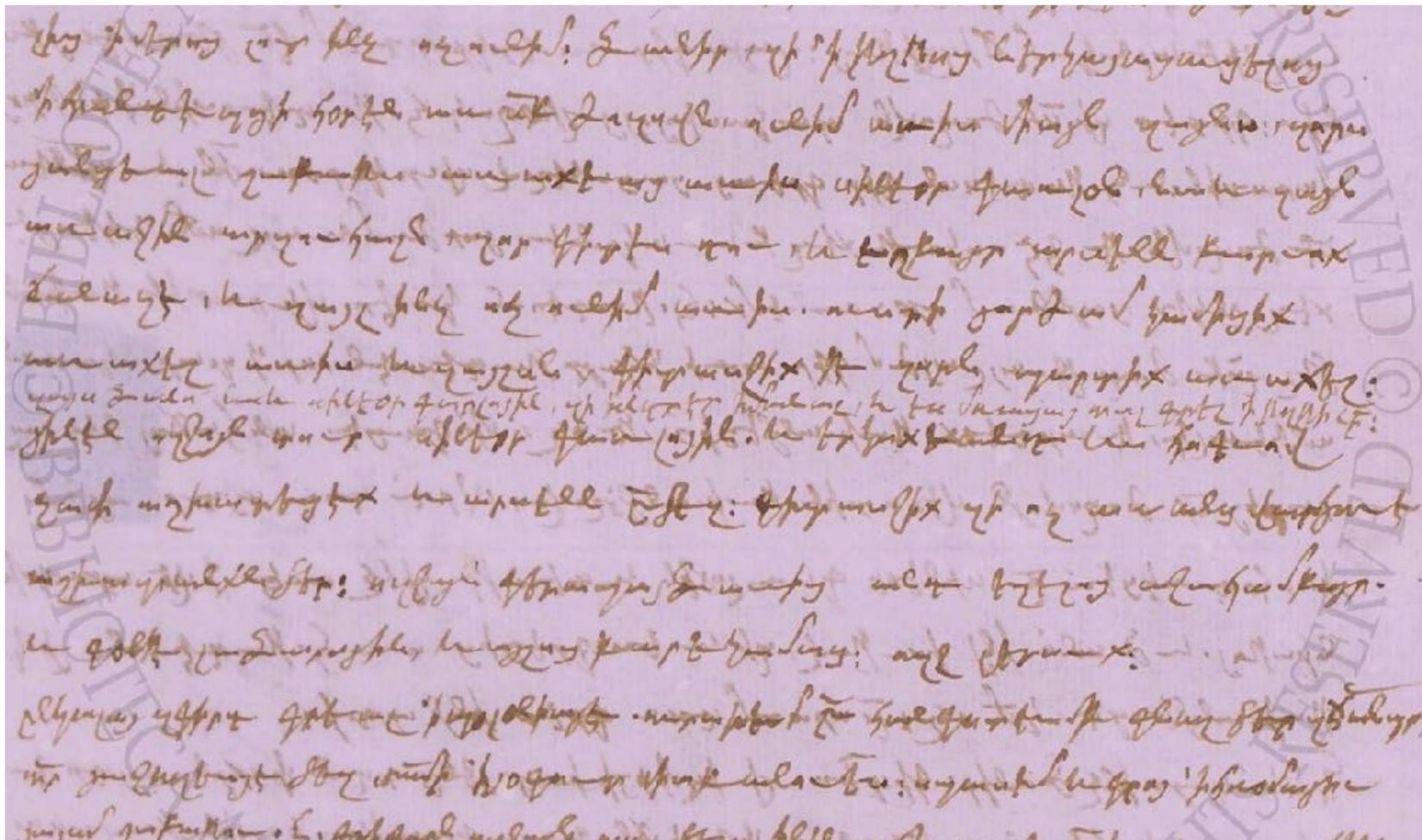


20.000 էջ  
Soon available



# Գրանշանների ճանաչում. օրինակներ

## Մխիթարեան միաբանութեան ձեռագիր նամակների մշակումը Հ. Վահանի և Մխիթարեան միաբանութեան գործակցութեամբ:



Melkonian letter, 1759, Mekhitarist Congregation

լից ի մերոց լուր ինչ ոչ ունիմ: Ծանիր, զի ի թղթոց ներկայացուցելոց  
հանցի հօրեն առ սբ ժողովն ունիմ առ իս միայն զայնս, զորս  
յանցել շաբաթս առաքեաց առ իս սինետր Գառլոն, նաև զայն  
առաջին արգուհլն, զոր գիտես դու, և եղբայր Արսենն բարուք  
ճանաչե, և զայլ ինչ ոչ ունիմ առ իս, ուստի յորժամ կամիցիք  
առաքել առ իս և զայլսն, գիտասցիք թե զորն պարտիք առաքեալ:  
զայս ժամու նաև սինետր Գարլոյին, զի խնդրեր խնանալ, և ես մոռացայ  
տալ գրել ի թղթի ի:

յինեն ողջոյն տուր սինետր Գառլոյին, և երկորքեանդ ևս հոգով  
չափ աշխատեցէք և Արսենն ըստ ձեզ: Գիտասցիք զի ոչ առանց վարձու է  
աշխատանքն ձեր: Ողջոյն գերապայծառից անդ եղելոց աջահամբայր,

Թանաքի  
թափանցիկություն

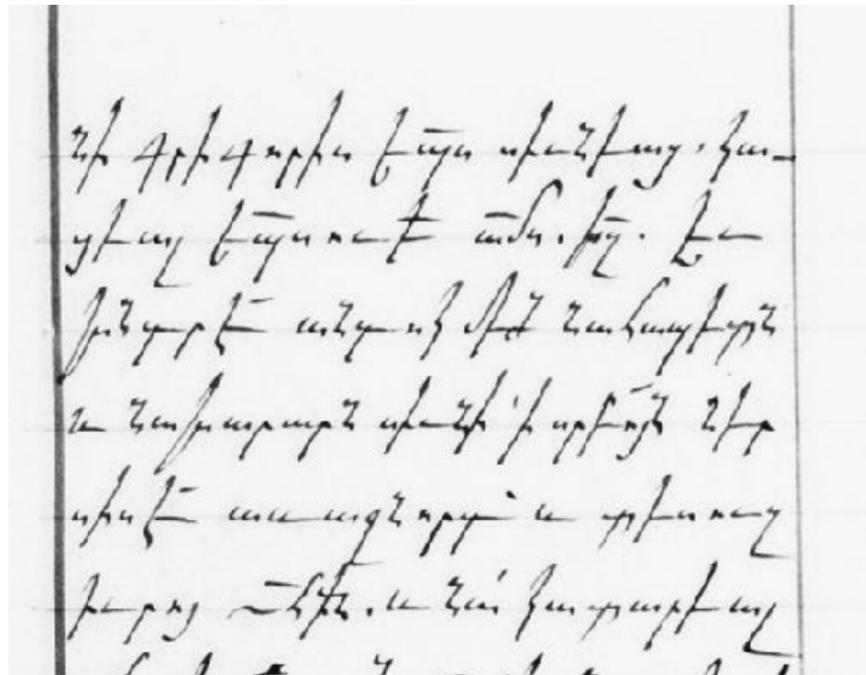
Հապալումները  
կարդացուում են

96.1%

Խնդիրը լուծուած է

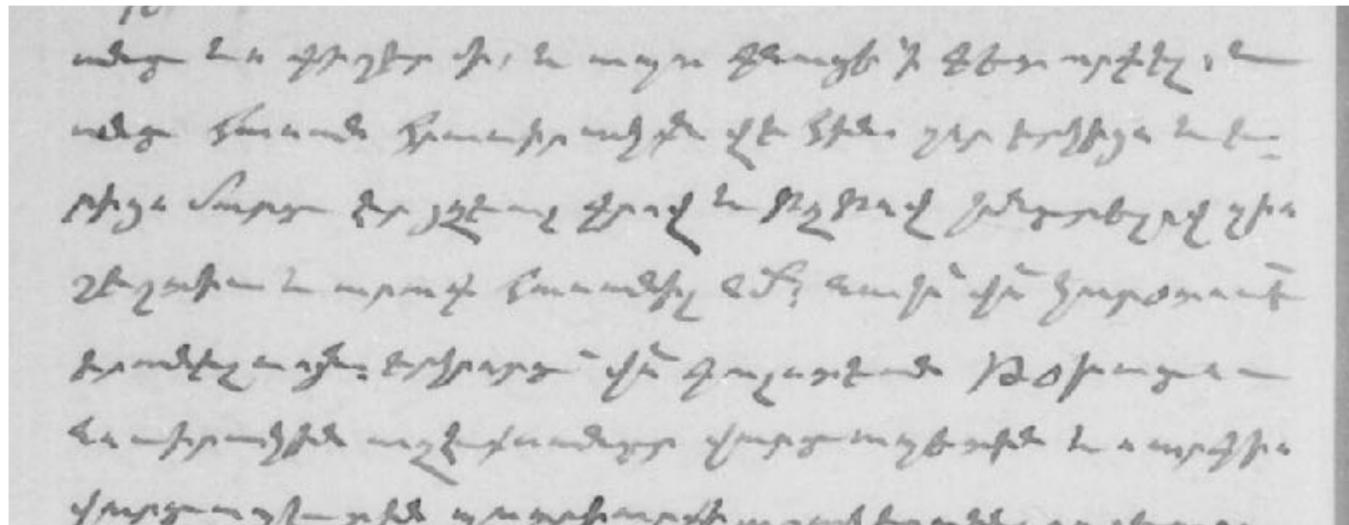
# Գրանշանների ճանաչում. օրինակներ

## Dulaurier նախագիծը Ֆրանսիայի Ազգային գրադարանի հետ



նի գրիգորիս եպ[իսկոպո]ս սինւեաց կա  
ցեալ եպ[իսկոպո]սու[թ]ե[ան] ամ[են]ս իս[ն]գ. Եւ  
խնդրէ անդոկ մեծ նահապետն  
և նախարարն սինի ի սրբոյն ներ  
սիսէ առաջնորդ և տեսուչ  
իւրոյ [աշխար]հին, և նա կատարեալ

98.75%



անդ ևս գրչեր մի, և ապա գնացի ի գետարկել: և  
անդ հասան հրաւիրակքն վեհին, զոր երկիցս և ե-  
րիցս մարդ էր յղեալ գրով և թղթով խնդրելով զիս  
շեշտիւ և արագ հասանիլ նմ: Նախ վ[ե]ր[ա]յ կարօտու[թ]ե[ան]  
երանելոյ ներկրորդ՝ վ[ե]ր[ա]յ գալստեան Թօխադու  
նուիրակին աղէքսանդր վարդապետին և սարգիս

Սկանաւորուած միկրոֆիլմեր

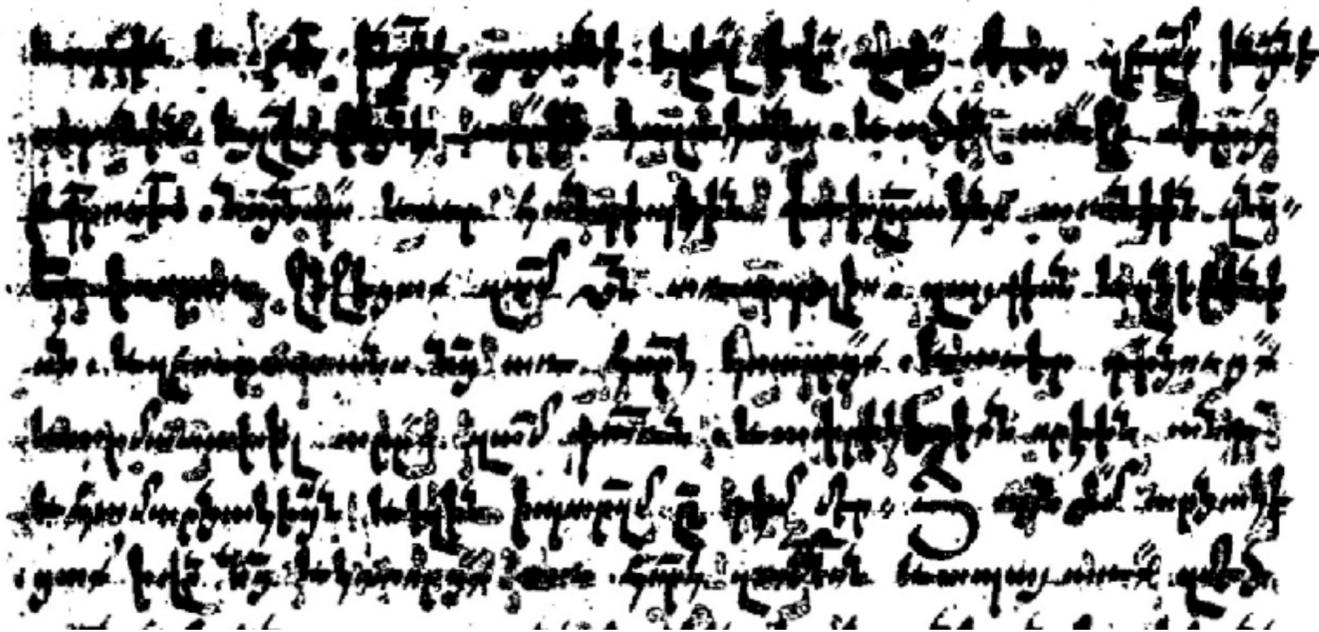
Հապաւումները կարդացուում են

96.12%

Խնդիրը լուծուած է

# Գրանշանների ճանաչում. օրինակներ

Ֆրանսիայի ազգային գրադարանի արուած հին ձեռագիրների (միկրոֆիլմերի) մշակում:



P191, Smbat Vardapet, French National Library

և որդիX և բժ. խնդնէ յաgտնի ելել ի վր զօջg մերոց զբզմս ի նցնէ սպանին: և յզվա ի նցնէ արրին կալկանg. և ածին առջի մերոյ թգրուեX նոյնպս և ուր հանդիպէին փախըXտկն առնէին զնս: Եւ իսպառ ջնջեցաք զամ ~ն տռապօլիս զայգիսն և զձիբէնի սն. և զբուրաստանս նց առ հարկ կոտորցք. և աւեր դիձուցք և արմատախիլ արրք զամ գւռն. և ափրիկեցիքն որ էին անդ և համարձակեցն և ելին ի պտրգմ ը դէմ մեր: Յայն ժմ արձակեցաք ի վր նց և կոտորցք XX հարկ զամնսն և ապայ առք զմեծ

Աղմուկը հաղթահարուած է

Հապաւումներ չկայ

95.1%

Խնդիրը լուծուած է

# Ընդհանուր ընդդեմ սովորական OCR

Իմ փաստաթուղթը  
պարզ է (կանոնաւոր  
փաստաթղթեր,  
հասարակ էջադրում -  
layout)



Պատրաստ է OCR-ի  
օգտագործման համար

Իմ փաստաթուղթը  
բարդ է  
(ձեռագիր, էջի  
ձեւաւորում - layout,  
թուայնացման որակ և  
այլն)



Պահանջվում է յատուկ OCR-ի  
մոդել

Տուեալ

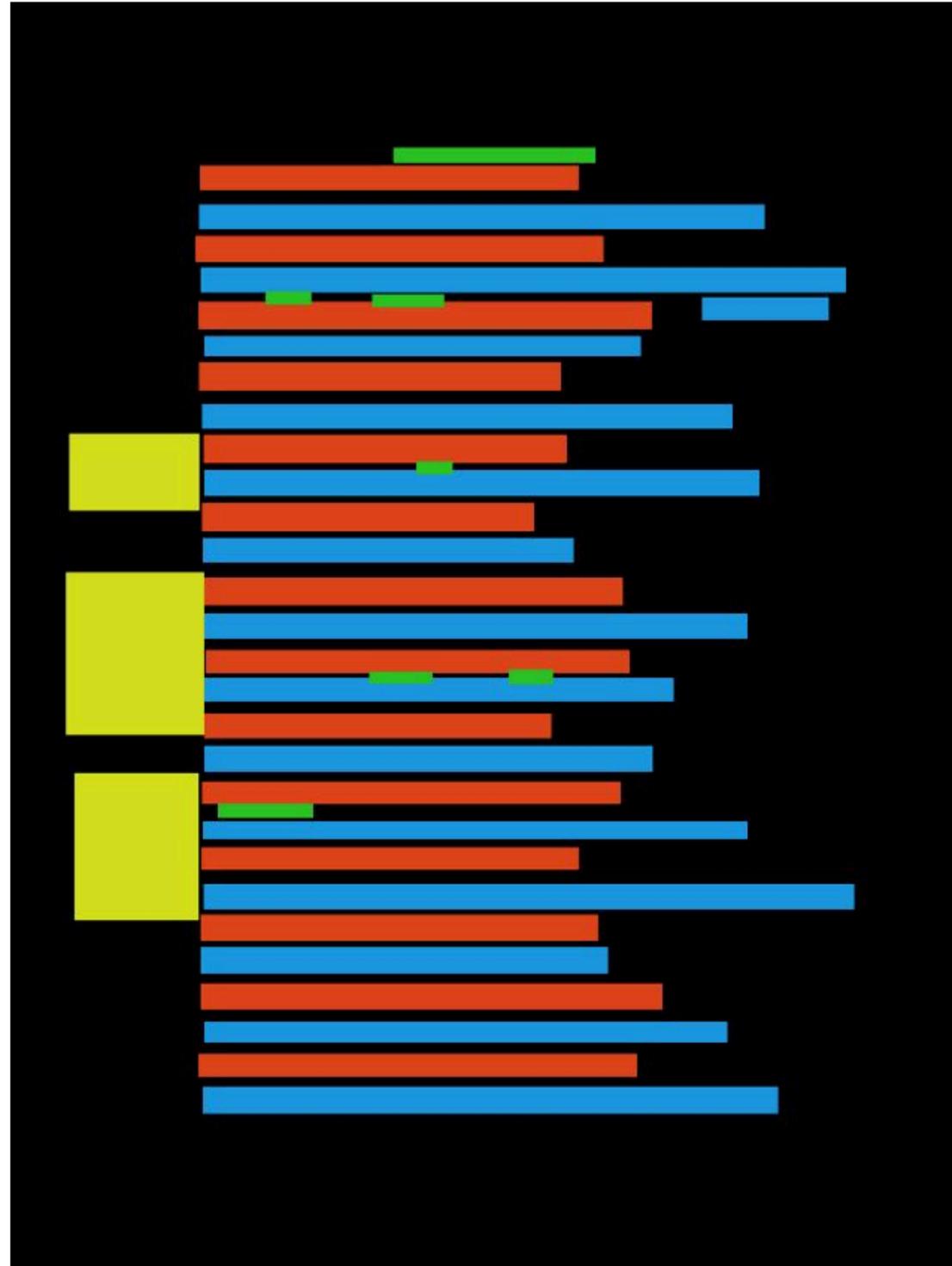
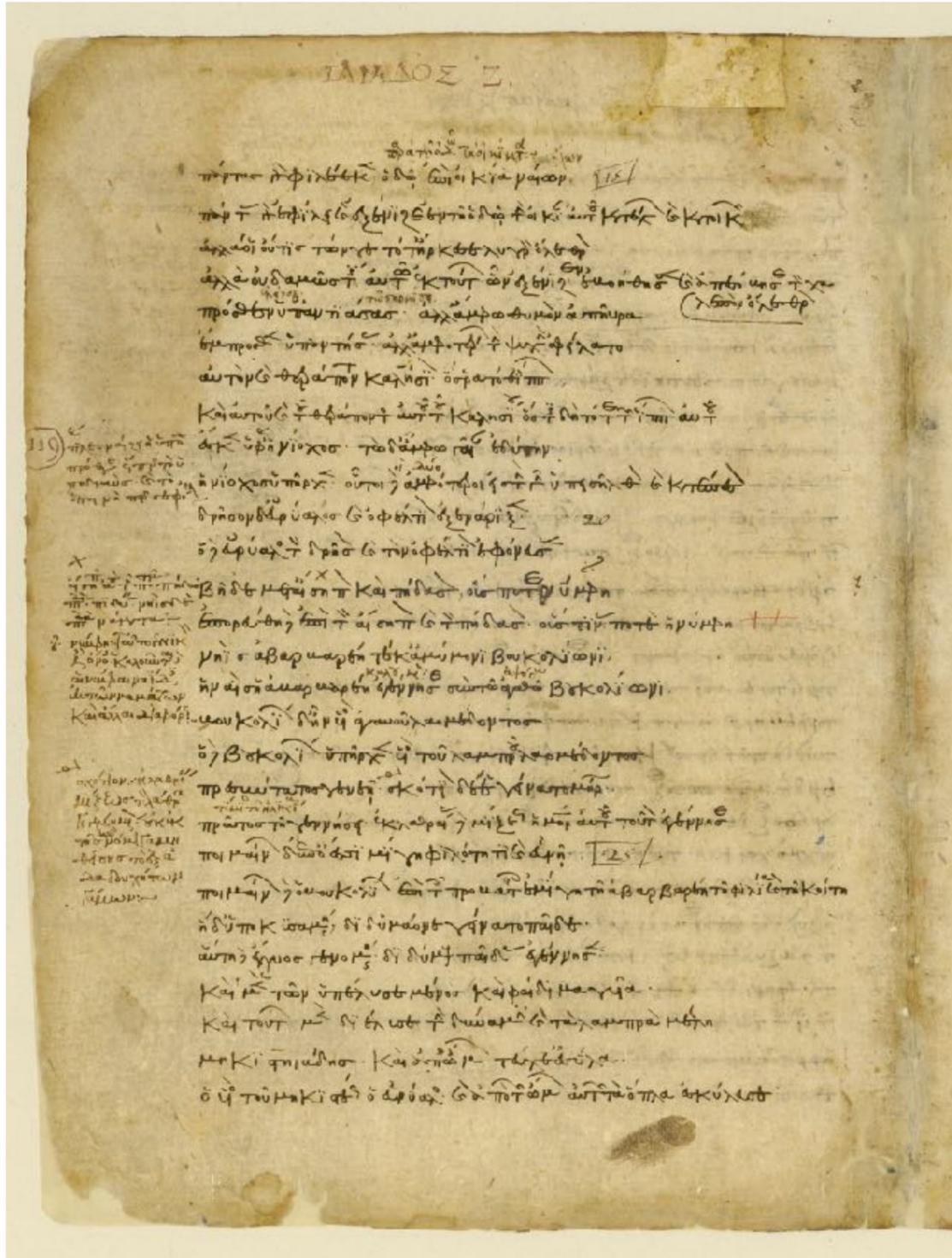
Վարժեցնել

- օգտագործողը
  - բաց աղբիւր
  - Քրաուդսորսինգ
  - e.g. Calfa
- e.g. Calfa

Արհեստական բանականութիւնը կարող է աշխատել միայն  
այն դեպքում, եթէ մենք գիտենք, թե ինչին ենք ուզում հասնել:  
Գրադարանավարը գտնուում է գործընթացի կենտրոնում,  
քանի որ նա է որոշում կարիքը:

# Φρασηαρωνների դերը - Էջի ձևաւորում (layout)

Ցանկացած նախագիծ, ներառեալ AI-ն և OCR-ը, սկսուում է կարիքների և խմբագրական ընտրութիւնների սահմանմամբ (մարդու որոշումն է)

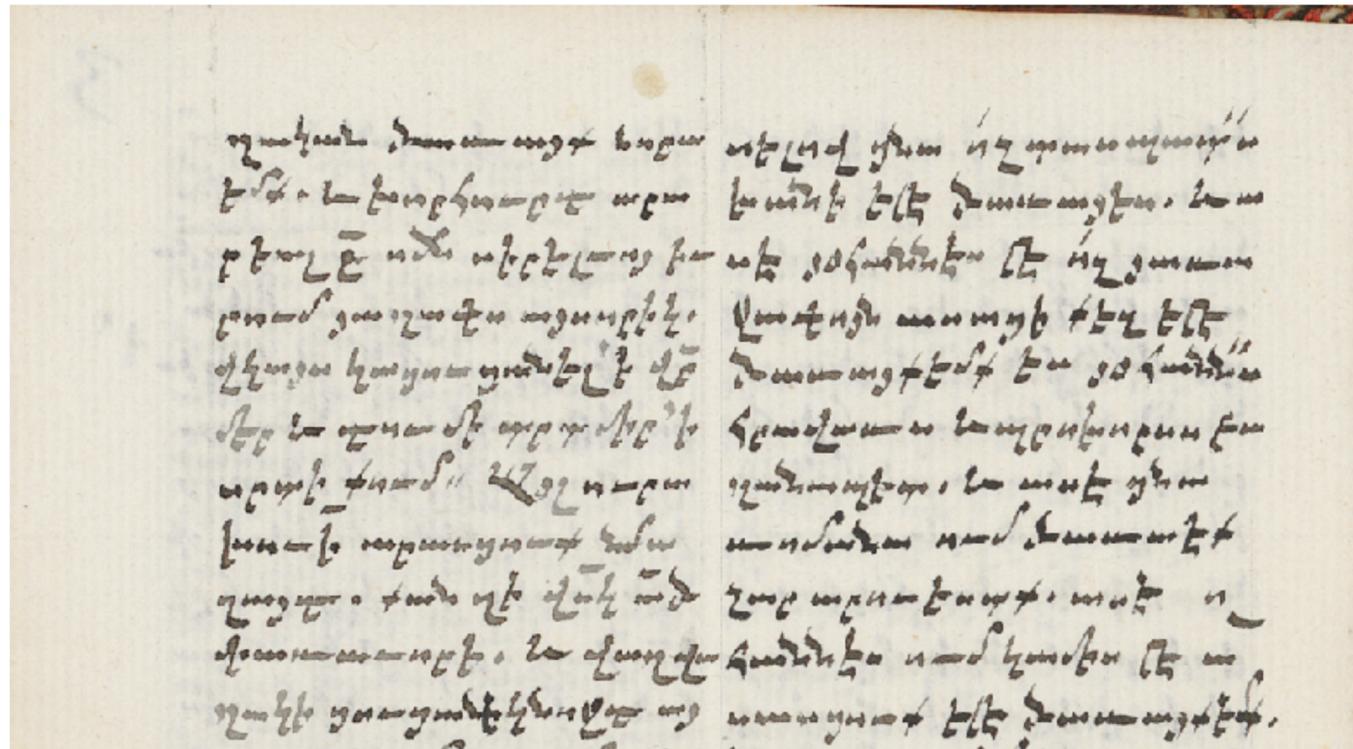


-  gloss
-  main text
-  commentary
-  interlinear text

<https://www2.unil.ch/iliade>

# Գրադարանների դերը - Տեքստ (text)

Ցանկացած նախագիծ, ներառեալ AI-ն և OCR-ը, սկսում է կարիքների և խմբագրական ընտրությունների սահմանամբ (մարդու որոշումն է)



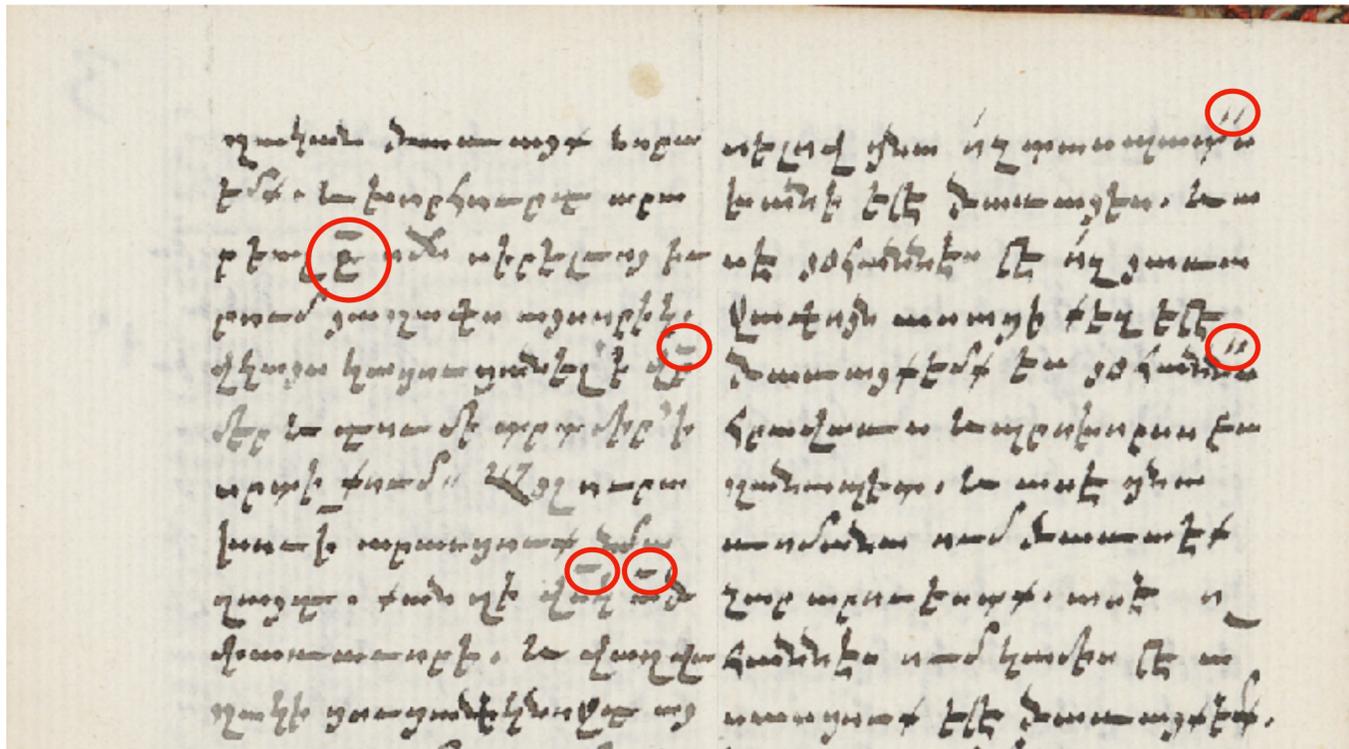
ms M5, f. 3b (Matenadaran, Erevan)

ղական ծառայքնորասելովցնառչտասպատս  
եմք.ևխորհուրդարախանիեթեծառայես.ևա  
րեալըոմնսիրելույիսէյօհաննէսթէոչյառա  
րումյաղագսայսորիկ.ջագոյնասացիքեզեթ  
վկայսկացուցանել'իվրծառայքեմքեսյօհաննս  
մերևդումիտրտմիր'իհրավառսևպրոխորսսբա  
սրտիքում..Այլուրաղանապետ.ևասէցնա  
խուիարասցուքնմառոմանաումծառաէք  
զայդ.քանզիվսկածչարարուեստք.ասէո  
միառաւորի.ևվաղվահաննէսումկամիսթէա  
ղակիցուցանէկնոջդայսասցուքեթծառայքեմք.

CER = 0%, բայց ինչ ասել էջի  
ձեւաւորման (layout), հապաւումների,  
տողադարձի և բառերի մասին:

# Գրադարանների դերը - Տեքստ (text)

Ցանկացած նախագիծ, ներառեալ AI-ն և OCR-ը, սկսում է կարիքների և խմբագրական ընտրությունների սահմանամբ (մարդու որոշումն է)



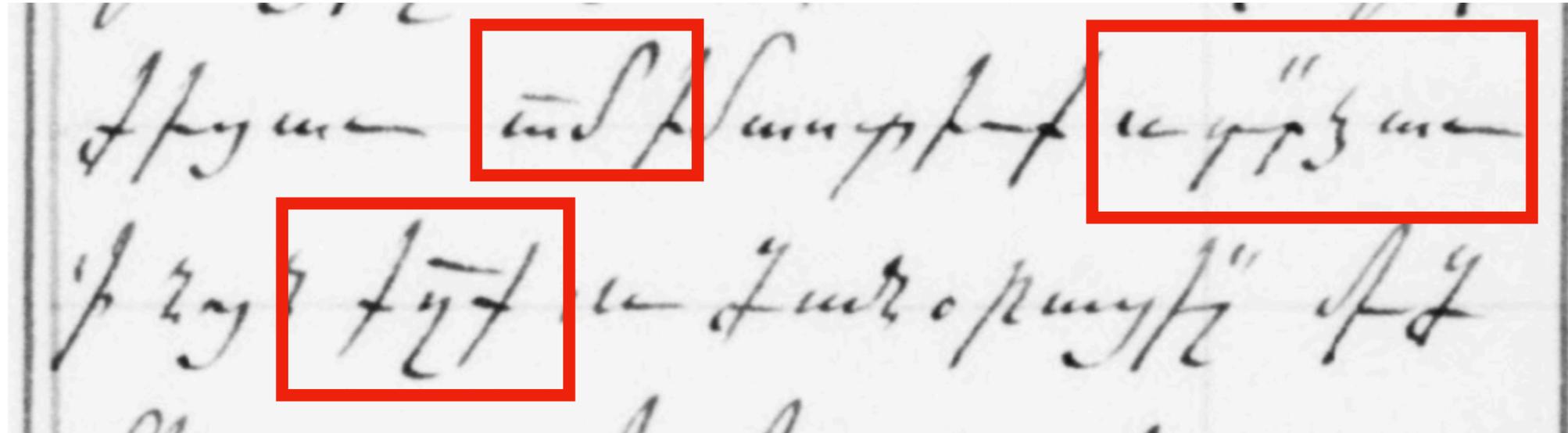
ms M5, f. 3b (Matenadaran, Erevan)

ղական ծառայքնորա  
ենք.ևխորհուրդարա  
րեալը[ստ]ոմնսիրելոյի  
րումյաղագսայտրիկ.  
վկայակացուցանել'իվ[ե]ր[այ]  
մերևդումիտրտմիր'ի  
սրտիքում..Այլուրա  
խու[թ]ի[ւն]արասցուքնմա  
զայդ.քանզիվ[ա]ս[ա]կա[ստուա]ծ  
միառաւորի.ևվաղվա  
ղակիցուցանէկնոջդայ

սելովցնառչտասպատ[ա]ս  
խանիեթէծառայես.ևա  
սէյօհաննէսթէոչյառա  
ջագոյնասացիքեզեթ  
ծառայքեմքեսյօհանն[է]ս  
հրավառսևարոխորսսբա  
ղանապետ.ևասէցնա  
ռոմանաումծառաէք  
չարարուեստք.ասէո  
հաննէսումկամիսթէա  
սասցուքեթէծառայքեմք.

## Գրադարանների դերը - Տեքստ (text)

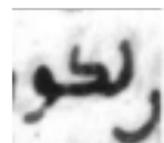
Ցանկացած նախագիծ, ներառեալ AI-ն և OCR-ը, սկսում է կարիքների և խմբագրական ընտրությունների սահմանամբ (մարդու որոշումն է)



- ▶ ամ [էնայն]
- ▶ դ [ա] ըձաւ
- ▶ ք [ա] ղ [ա] ք



▶ 1 տառադարձում = մի քանի տարբեր գրաֆիկական ձևեր



تكون ou يكون

▶ 1 գրաֆիկական ձև = մի քանի հնարաւոր տառադարձումներ

## Արհեստական բանագիտություն և փասթաթղթեր

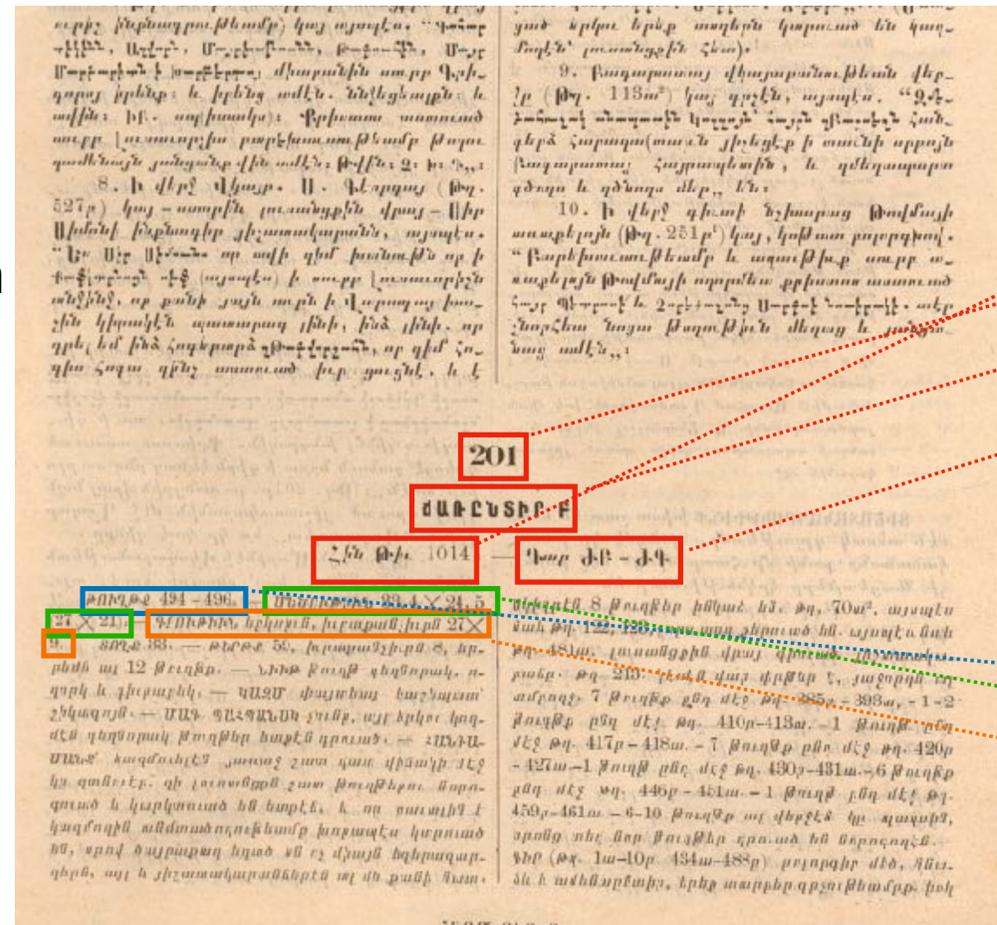
---

կատալոգներ • լուսանկարչական ֆոնդեր • գիտելիքների արդիւնահանում

# Գրանշանների ճանաչումի հնարաւորութիւնները ընդլայնել

## Սուրբ Ղազարի ձեռագիրների Քարտարան թուային տարբերակի ստեղծումը Միսիթարեան միաբանութեան համագործակցութեամբ:

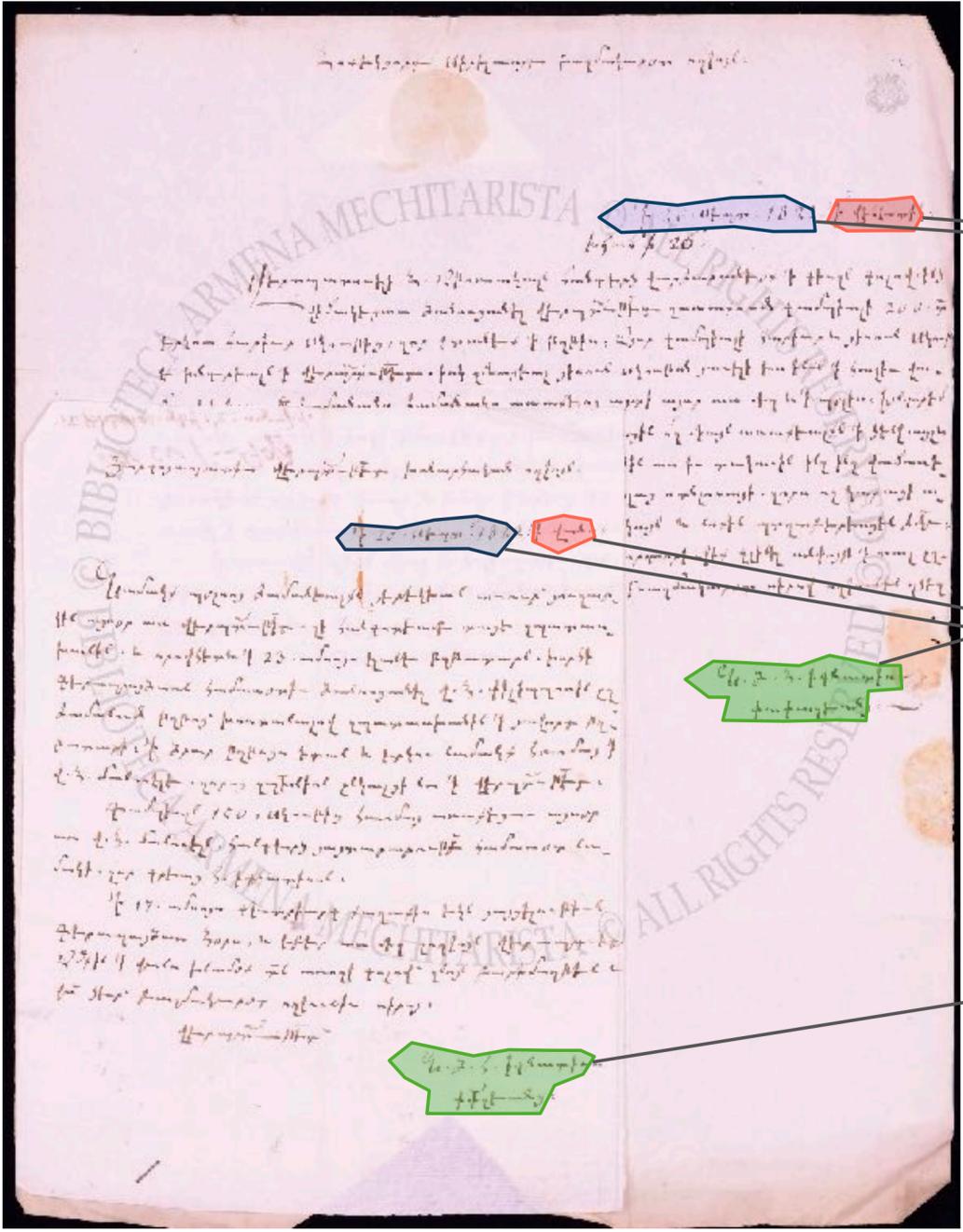
- ➔ Մասնագիտացուած Արհեստական բանականութիւն էջի ձեւաւորման, տառատեսակի և իմաստաբանութեան համար
- ➔ Միջին ճշգրտութիւնը՝ 99,5%
- ➔ Առցանց տուեալների շտեմարանի ինքնագործ լրացում



IDENTIFIANT	V1014
NUMÉRO	1814
NUMÉRO DE NOTICE	281
TITRE	ՃԱՌԸՆՏԻՐ Բ
DATE DE DÉBUT	XII
DATE DE FIN	XIII
DÉTAILS DE DATE	անյայտ, սակայն դատելով ներքին հանգամանքերէն հաւանօրէն գրուած առաջին կէսին, և կամ ժԳ. լոյին առաջին քառորդին. վասն զի վերջին սր յիշատակարաններն (տես Թղ. 123բ, 114բԵ). կը կրեն առաջինը՝ ՊԸԸ (14: -1377) թուականները, և որոնք համեմատութեամբ երկաթագրին՝ շատ նոյնն են:
LIEU DE COPIE	անյայտ, հաւանօրէն Ս. Աստուածածին կամ Ս. ԅովհաննէս վանքն, որ և կ'ընդգրկէ Ս. Մարտիրոսի վանքը:
RÉFÉRENCE	t. II, 1924, c. 35-66
GENRE	Մատենադարանի ձեռագիր
NOMBRE DE PAGES	494-496
DIMENSIONS	33,4X24,5 (27X21)
MISE EN PAGE	երկսիւն, իւրացանջիւրն 27X9
TYPE D'ÉCRITURE 1	Մատենադարանի ձեռագիր

# Գրանշանների ճանաչումի հնարաւորութիւնները ընդլայնել

Ցուցակագրել Մխիթարեան միաբանութեան նամակագրութիւնները  
Հ. Վահանի և Մխիթարեան միաբանութեան գործակցութեամբ:



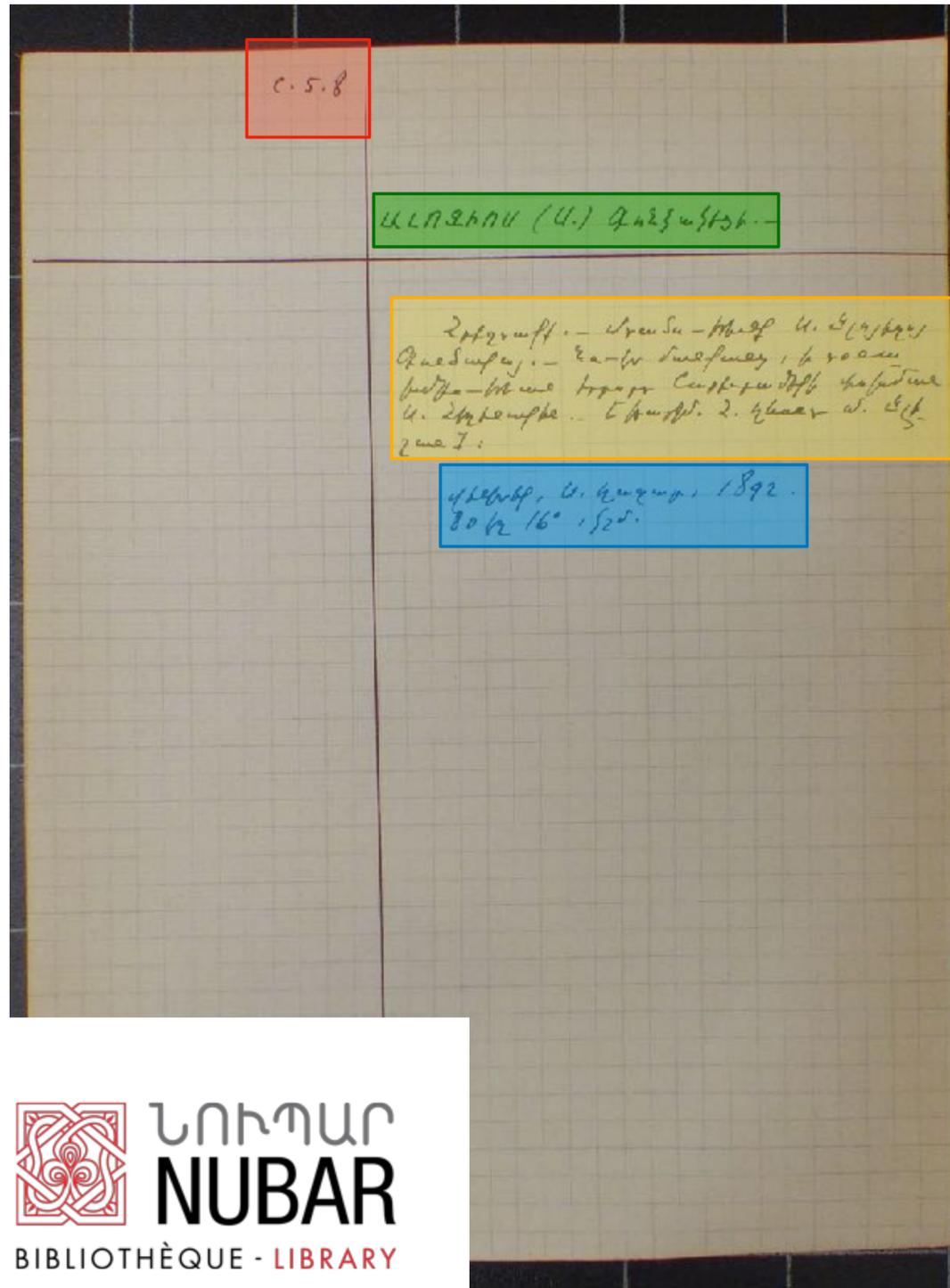
**Թուական:** 20 Սեպտեմբեր 1822  
**Վայր:** Վենետիկ  
**Հեղինակ:** Իգնատիոս Փափագեանց

**Թուական:** 20 Սեպտեմբեր 1822  
**Վայր:** Վանս  
**Հեղինակ:** Իգնատիոս Փափագեանց

Ongoing

# Գրանշանների ճանաչումի հնարաւորութիւնները ընդլայնել

## ՀԲԸՄ-ի նուբար գրադարանի (Փարիզ) մատենագիտական գրառումների վերածումը փնտրման համակարգի



**OCR**

<b>Catalog number:</b>	C.5.8
<b>Author Name:</b>	ԱՆՈՋԻՈՍ (Ս.) Գոնձակեցի
<b>Content:</b>	Հրեշտակք. Մտածութիւնք Ս. Ալոյեզոյ Գանձակայ. նուէր մանկանց, ի տօնա խմբութեան երրորդ հարիւրամեկի փոխման Ս. Հեղինակին. [թարգմ. Հ. Ղևոնդ Մ. Ալիշան]

**Text Analysis**

վենետիկ Ս. Ղազար 1892  
80 էջ 16°, կզմ

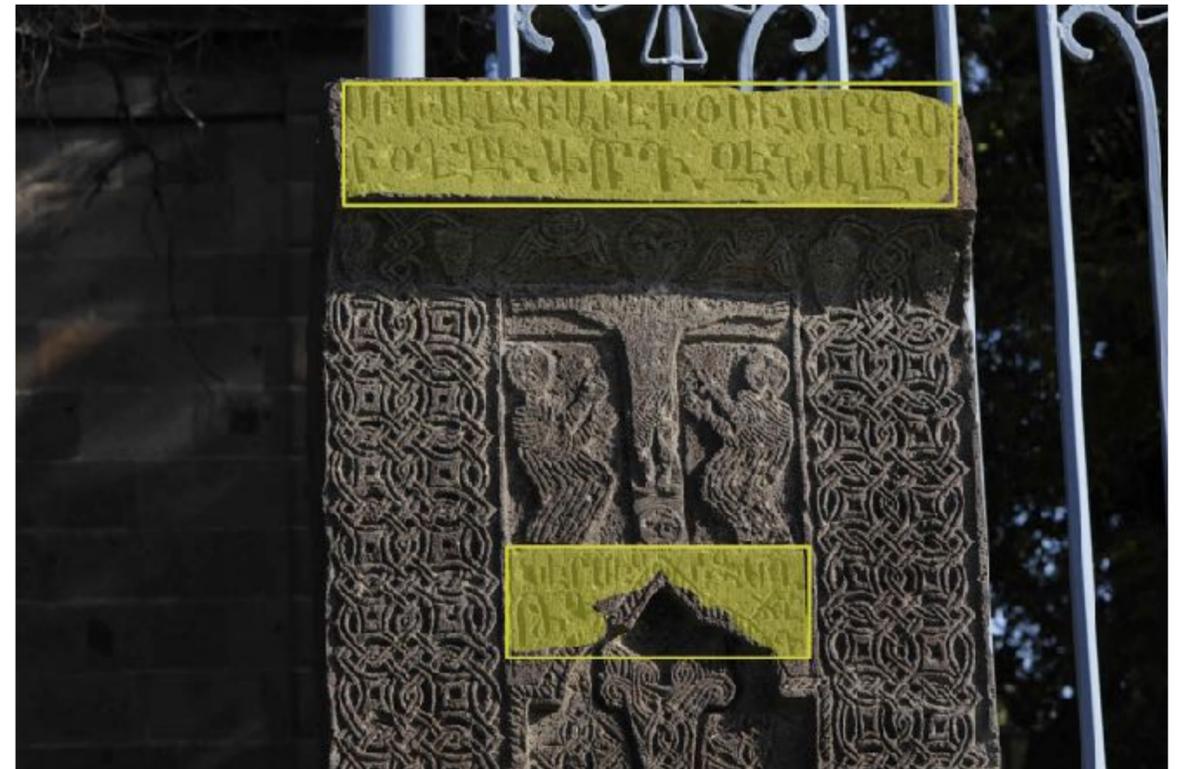


Latin catalog done in 2023, Armenian catalog < 1986 done Available in Nubar Library intranet

# Գրանշանների ճանաչումի հնարաւորութիւնները ընդլայնել

## Զարգացնել լուսանկարչական ֆոնդերի հասանելիութիւնը

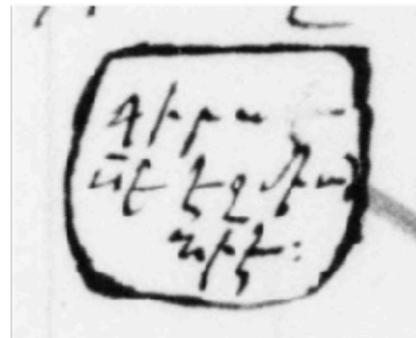
Ongoing



source image: Wikimedia Commons

# Գրանշանների ճանաչումի հնարաւորութիւնները ընդլայնել

Առարկաների ճանաչում ձեռագիր արխիւներում (Dulaurier, BnF Datalab նախագիծ, 2023)



{ BnF



CALFA

To know more: [calfa.fr/blog/39](https://calfa.fr/blog/39)





# Գրանշանների ճանաչումի հնարաւորութիւնները ընդլայնել

ԹՈՒՄՈ աշխատանոց (Մարտ 2024) - Երաժշտութիւն կարդալ և ծրագրաւորել AI-ի միջոցով

## My TUMO webapp

Select Model

chahan\_model.pt

select classes to predict

black x key x white x

tone x

Confidence Threshold

0.04

0.00 1.00

Choose a demo image

Image 1

Upload Image

Drag and drop file here

Limit 200MB per file • JPG, JPEG, PNG

Browse files

ՄԱՆՈՒՇԱԿԻ ՎԱՂԱՐՇԱՊԱՏԻ (Դափի ոճով)

Յոթ պար SEVEN DANCES երկրորդ խմբագրութիւն Second Edition

MANUSHAKI VAGHARSHAPAT (In the style of dap)

(Թեքև և սահուն) Leggiero ♩. = 52 - 55

Unloaded Image

ՄԱՆՈՒՇԱԿԻ ՎԱՂԱՐՇԱՊԱՏԻ (Դափի ոճով)

Յոթ պար SEVEN DANCES երկրորդ խմբագրութիւն Second Edition

MANUSHAKI VAGHARSHAPAT (In the style of dap)

(Թեքև և սահուն) Leggiero ♩. = 52 - 55

Detected Image

# Հեռանկարներ

1

OCR / HTR-ի ճշգրտություն հայերենի համար > 95% մասնագիտացուած մօտեցումներով

2

AI-ն որպէս որոշումների աջակցման/օգնութեան գործիք օգտագործողների և գրադարանավարների համար

3

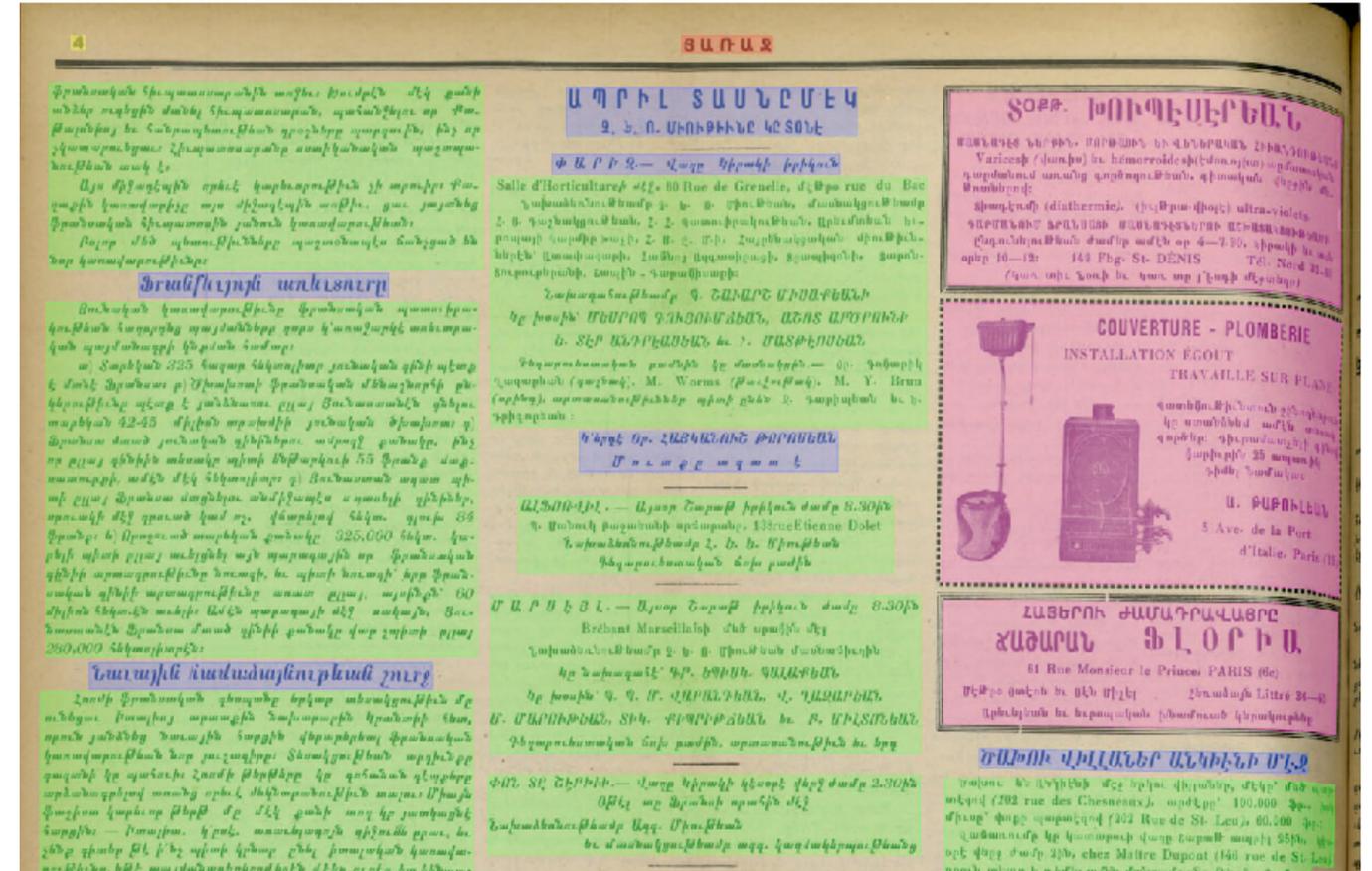
Հետազոտողներն ու գրադարանավարները AI-ի նոր մոդելների մշակման և սահմանման գործընթացի հիմքում են

4

Մասնագիտացուած մոդելներով տուեալների լայն հաւաքոյթի (datasets) մշակում, ճանապարհ բացելով տուեալների հետազոտման (Data Mining) համար

# Նոր նախագիծ 2024 թ. - Հ. Սամուելյան արևելյան գրախանուր

Հայկական թերթերի թուայնացում և OCR (~ 15k էջ)



FONDATION  
CALOUSTE  
GULBENKIAN





CALFA

Շնորհակալութիւն

To know more about Calfa and our solutions : <https://calfa.fr>

Contact us : [contact@calfa.fr](mailto:contact@calfa.fr)