



## STATE

## Assisted Text Transcription System

David Llorens, Andrés Marzal, Vicente Palazón, Federico Prat, Juan Miguel Vilar (UJI, Spain)

María José Castro, Salvador España, Joan Pastor, Francisco Zamora (UPV, Spain)

# The people



## **Automatic? Transcription**

- Optical Character Recognition (OCR) systems perform poorly:
  - On damaged documents.
  - On ancient documents.
  - On handwritten documents.
  - On text with non-estandar characters or fonts.
- It is always needed to supervise and correct the automatic transcription.

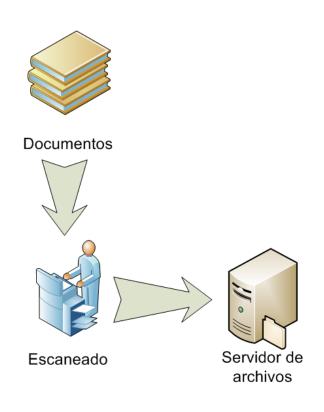
Delestylo.

aquella palabra, e den lo por fechor que puede ser justificado, segú dicho es Mas el querelloso no le deue matar: e si lo mata, deue ser dado por enemigo de los sus pariétes: e pechar el omezillo. Y esto es por razó q no gelo dio el alcalde por enemigo, segun dize eneltitulo delos omezillos: en la ley que comiença. Si aquelenel. &c. E todo otro hombre que matare su enemigo, maguer que lo aya desassado con dere

fuero, dar lo an por enemigo delos tes, e que salga dela villa, e del term sobre esto es a saber, que quando a manera de desendimiento se comi demandar la muerte segun el suero que todo lo que dize eneste suero de sazer, e de juzgar despues del di miento, q esso se ha de guarda, e den e juzgar: e no puede mudar la q rella demanda en orra manera si posserva

the Dowell finds it easier to take it cut of mothers, childrens and cick people than to take on this year the believing. It is Born commented icity, a let us have a full inquiry into the cost of drugs and the pharmaceutical bridgety. The health of children today owed much to the welfare food scheme. It was main tained during the was. Now in conditions of Tory offluence it seemed it could not be carried or.

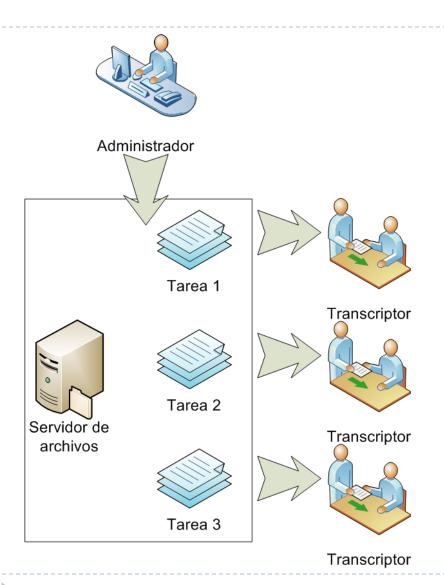
## Documentary resources digitization



Documents are scanned...

 ...and the obtained images are uploaded to a file server.

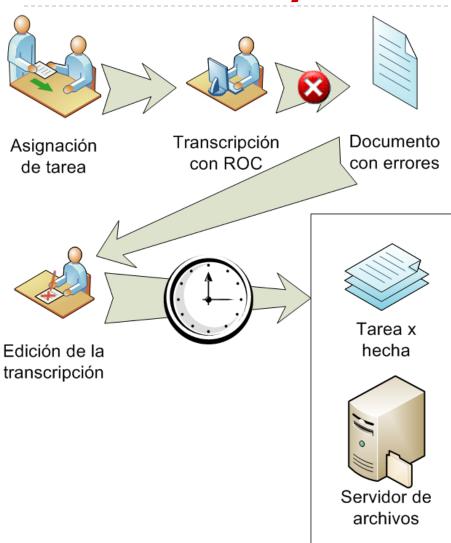
## **Documentary resources digitization**



- The administrator divides the documents in groups of pages (tasks).
- Each task is assigned to a human transcriber.

His dutie is to obtain a text file from the original document.

## Documentary resources digitization



- Each document is transcribed with an OCR sytem.
- The resulting text contains many errors.
- Text editing: the transcriber must manually correct the text.
- It is the most time consuming stage, where it is easy to introduce errors.
- Finally, the correct transcription is uploaded to the server.
- Now, the digital information can be exploited in many ways.

## How is text editing performed?

Delestylo.

Fol.7

aquella palabra, e den lo por fechor que puede ser justificado, segú dicho es. Mas el querelloso no le deue matar: e si lo mata, deue ser dado por enemigo de los sus parietes: e pechar el omezillo. Y esto es por razoq no gelo dio el alcalde por enemigo, segun dize eneltitulo delos omezillos: en la ley que comiença. Si aquelenel. &c. Etodo otro hombre que matare su enemigo, maguer que lo aya desassado con dere

fuero, dar lo an por enemigo delos parien tes, e que falga dela villa, e del termino. E fobre esto es a saber, que quando enesta manera de defendimiento se comiença a demandar la muerte segun el fuero viejo: que todo lo que dize eneste suero se sa de fazer, e de juzgar despues del desastamiento, se esto se de la desasta, e juzgar: e no puede mudar la se ella, nila demanda en arra manera si no segun lo co

aquella palabra, e den lo por fechor que .puede serj u [st]i[si]cado,seg[un] dicho es.Mas el quer.elloso no le deue matar: e si lo mata, deue ser dado por enemigo de los sus pari[en]tes:e pechar el omezillo. [ss] e[st]o es por raz[on][que] no gelo dio el alcalde por enemigo, segun dize enel titulo delos omezillos: en .l.a ley que comienca. bi aquel enel. ac. .E todo otro hombre que matare su enemi go, m.uguer quelo aya desa[si]ado con dere

- Comparing the page with its transcription is very uncomfortable and difficult.
- ▶ The human expert must
  - jump continuously from the image to the text,
  - find each error,
  - move with the mouse or keyboard,
  - > and type the correction.

#### Ancient documents

#### **Problems** get worse with ancient documents

Stains, handwritten annotations on the margins, bad conditions, cracks, patches, disappearing ink in some regions, and so on:

#### Text extraction is difficult for OCR systems.

Special fonts or ancient graphemes:

#### OCR systems do not recognize them.

Non-modern orthography or syntax and archaic language:

Language models of OCR systems are not suitable.

Humidity stains

Las leves

Hayacumbry

e de al rey el querelloso, o el acusado que diga que quiere ser oydo e librado por el, fi esto dixere ante que el pleyto sea contestado ante los alcaldes del lugar: en ronce suyo es de oyr e de librar estas cosas fobredichas, o puede los embiar el rey si quisiere estos pleytos a los alcaldes do fue te de los presos sueltos sobre fianças e el fuero delos Ancient chos. Pero fi c graphemes gun los fuere do tales fech..... destos en algunos destos fechos de muerte,o de tolimiento de miembro,o de echa miento de tierra, mas ay otra pena de dinero, o de al. Entoce tales pleytos, maguer vegan por querella ante el regidenen fer embiados a que los libren fus alcaldes de las villas do tales Echosacacícieron por la amino quebratado, maguer - 1- 1 peros fi querellaré al rev:

esta querella Archaic Orthography? biudas e de language Hyphenation?

- Ley. xey. Que el que no fu injuria, o delos fuyos no deue fer recebi do a acusacion, si no se obliga ala pena del

lalguno viene diziedo al alcalde, que fulano hombre q es ay enel lugar, que hizo algumalfecho que no atañe a el. wale el testimonio dela muger. Sife quifiere obligar a acufarle, e obligar a la pena que el otro deue auer, fi no gelo probare, deuelo oyr el alcalde: mas en otra guila no lodeue oyr: faluo fi mostrare carta,o alguna otra cosa que hiziesse alguna fe al alcalde, porque se ouiesse de mouer contra el aculado.

Ley.xciij. Como el marido no pue de matar al vno de los adulteros e dexar

Neltitulo delos adulterios, en la primeraley dize assi. Si muger casada fa ze adulterio, ambos sean en el poder del marido: e faga dellos lo q quifiere e de lo q an, assi que no pueda matar el vno dellos, e dexar al otro. Sobre estas palabras si acae sce q se vaya el vno, e prenden al otro: y el

preso es vécido d'adulterio por juyzio dar gelo an los al aldes en poder del marido: v el marido deuclo tener, mas no lo deuc matar fata q aya el otro y le veça por juy-zio, voro los mate ambos si quisiere.

xciiij. Que escriuanos ande

Os pleytos delos que estan presos, e delos q fueron enfiados ante el alcal de,o de escreuir la fiadus

rey que escriue concl al Handwritten tos de los fobredichos a annotations escreuir el escrivano de conclel alguazil en cafa uerrey.

Ley. xcv. Que manera terna el calde fi el acufado no viene a responde la acufacion.

Trofi, fi alguno acufa a otro que le quemo sus casas,o quele mato su pa e, o sobre otra cosa desguisada echo: y el alcalde lo fizo empla alos plazos que el fuero manre,entonce deue el alcalde fa-

ber del techo de que querello, fi fue fecho: mas no a de saber quien lo fizo. E si fallare que tal fecho es fecho, entoce lo deue dar porfechor of put det ma

Ley.xcvj. En que casos, e quando

Obre la ley que comiéça, toda muger, que es en el titulo delos testimonios: es a faber, que pueden las mugeres fer rescebidas en testimonio sobre las cosas, quier sea ceuiles, quier criminales q se fazen en tal lugar, q no es razo ni aguifado de fer, y hobres colas mugeres Farra f. fe screscibé las m Handwritten vedidas,y en la zer las mugere underlines leficios q acael uese por su dicho de mugere, en testimonio. E otrofi en la pesquisa que le faze de los yerros fechos de noche en yerra, fi ellas dan testimonio de vista, juzgen lo por prueua. E otrofi faze los sus dichos presum pcion para poder tormentar, mas en aque

#### **Our aims**

- Our transcription system integrates advanced and accurate tools
  - for image processing,
  - to detect page layout,
  - for text recognition.
- But also... it assists the human expert to
  - partially automatize the text editing process, and
  - attend to usability in order to ease the interactive text editing process.
- And the recognition system includes adaptive learning: it learns from samples of each new task.
- With all of this: we aim to drastically reduce the time devoted to text editing.

### **Multimodal interaction**

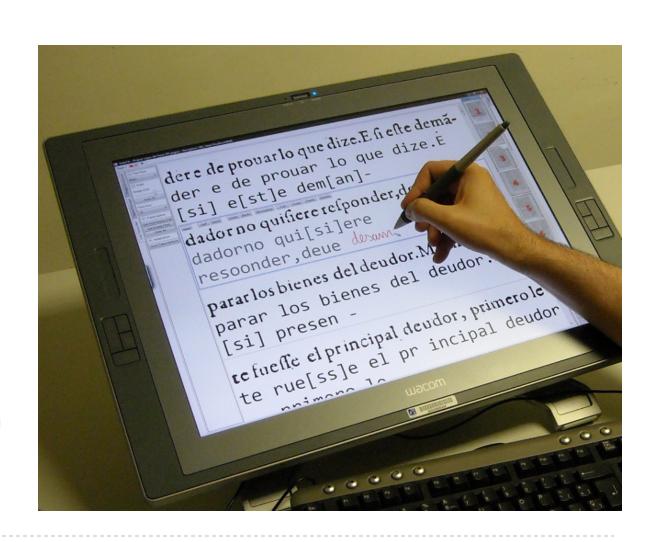
- Mouse, keyboard, and stylus.
- Voice



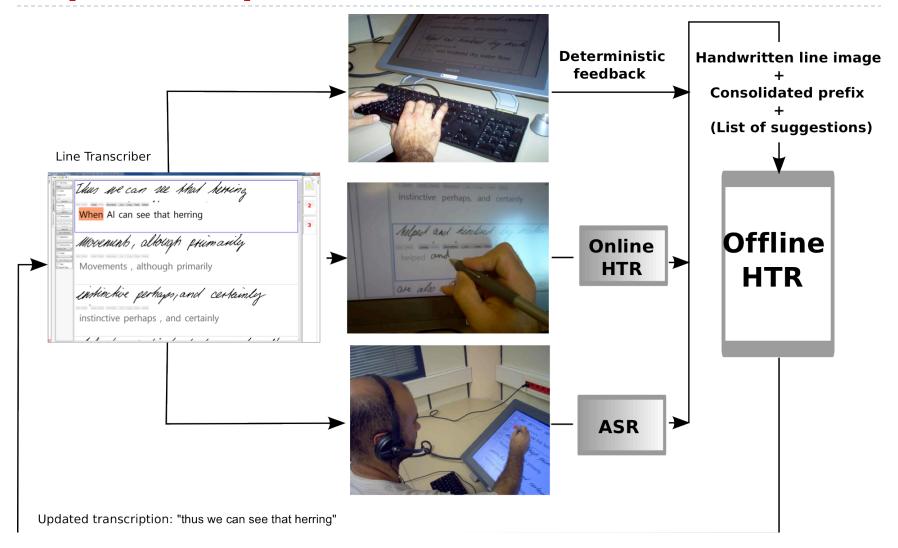


## STATE: An Assisted Text Transcription System

- ✓ It allows to edit text line by line
- ✓ Comfortable interaction
- ✓ Keyboard, stylus or speech



# User-interaction cycle with STATE via keyboard, stylus, or voice



## **Questions?**

