

Transcriptor: Sistema colaborativo para la transcripción de manuscritos históricos

Christian E. Bracco¹, Lucas E. Cuevas², Diego Torres³, Gimena Del Rio Riande⁴

¹ LÍFIA – CICPBA, Fac. de informática, UNLP, La Plata, Argentina

² LÍFIA – CICPBA, Fac. de informática, UNLP, La Plata, Argentina

³ LÍFIA – CICPBA, Fac. de *informática*, UNLP, La Plata, Argentina, Dept. de CyT, UNQ

⁴ CONICET, Argentina

bracco66.cb@gmail.com, lucascuevas1992@gmail.com, diego.torres@lifa.unlp.edu.ar, gdelrio.riande@gmail.com

Introducción

Transcriptor es una herramienta de transcripción de manuscritos de colaboración abierta (crowdsourcing) presentado como un sitio web el cual busca facilitarle el trabajo a aquellas personas interesadas en compartir y perpetuar el mensaje de documentos históricos, brindando herramientas simples para realizar la transcripción [1] de los mismos a formato digital apoyándose en una comunidad que respalda la fiabilidad de los datos recabados y garantizando un medio en el que fácilmente pueden guardarse y consultarse el contenido de las obras trabajadas.

Motivación

A menudo personas interesadas en manuscritos históricos, como pueden ser los diarios de un pariente, periódicos o libros antiguos, encuentran difícil interpretar lo que dicen ya sea porque no comprenden el idioma, usan palabras que desconocen o simplemente porque el contenido no es fácil de leer, lo que puede llevar a problemas de comprensión del contenido de los mismos. Otras cuestiones involucran también las limitaciones que existen al procesar toda esta información en un formato de texto digital, siendo que el mismo no abarca todos los aspectos visuales del documento, resultando en la pérdida de importantes datos sobre la composición de dicho manuscrito y el contexto en el que fue escrito (correcciones, borrones, tachaduras, llamados, sellos, marcas, etc.).

Adicionalmente, el día de hoy existen herramientas que nos permiten realizar una labor más automatizada en la transcripción de textos [2] (OCR y herramientas de corrección automática) pero los resultados, aún ante la optimización de estos procesos, siguen siendo dependientes de la supervisión humana para garantizar la calidad y la fiabilidad en los textos obtenidos [3] [4].

Enfocándose en la ardua labor que conlleva esta interpretación y curación de datos, surge la propuesta de Transcriptor.

La Plataforma

Transcriptor busca alentar la participación de la sociedad en la interpretación de manuscritos antiguos adjudicando a cada individuo un importante papel en este proceso, permitiéndole aportar su visión y conocimiento personal sobre el mensaje del documento, enriqueciéndose a su vez con nuevo conocimiento y dando lugar así al ejercicio de la ciencia ciudadana.

Gracias al novedoso sistema de marcas en el documento, el transcriptor podrá enlazar el texto que va tipeando inmediatamente a la línea o párrafo en la página digitalizada, facilitando la búsqueda y ofreciendo la oportunidad a que otros usuarios opinen y ayuden a la transcripción del autor. Para elementos de contenido no textual (tales como firmas y sellos, lugares y referencias a otros textos o personas), Transcriptor permite la definición de marcas semánticas, que consisten en anotaciones que estructuran la información utilizando representaciones conceptuales de objetos del mundo real. La plataforma incorpora las características de foros de discusión y votos sobre las transcripciones realizadas, brindando un espacio a la comunidad para expresar sus opiniones sobre las distintas transcripciones y sobre los términos que se adecuan más a lo que el texto quiere transmitir, todo esto desarrollado con conceptos de gamificación que motivan y hacen más entretenida la colaboración de los usuarios con logros y desafíos propuestos por el sitio. Transcriptor organiza la información recabada de manera ordenada, en distintas colecciones y categorías, lo que permite a los usuarios encontrar proyectos de diversos tópicos, y acceder a ellos referenciando lugares, géneros, versiones e idiomas. Finalmente, Transcriptor estará integrado al proyecto Cientópolis, lo que permitirá a esta comunidad cargar y transcribir colaborativamente manuscritos digitalizados (imágenes, Archivos PDF, entre otros) para seguir explotando y motivando la ciencia ciudadana.

Referencias

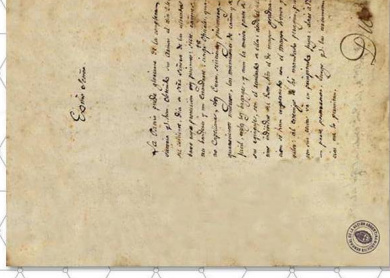
- [1] C. Huitfeldt and C. Sperberg-McQueen, "What is transcription?" *Literary and Linguistic Computing*, vol. 23, pp. 295–310, 09 2008.
- [2] "OCR, ICR y OMR. ¿qué son y para qué sirven?," Nov 2016.
- [3] M. J. Brown, "Experiments with early modern manuscripts and computer-aided transcription," Feb 2019.
- [4] M. Moyle, J. Tonra, and V. Wallace, "Manuscript transcription by crowdsourcing: Transcribe bentham" *Liber Quarterly*, vol. 20, p. 3,01 2010.

Transcriptor

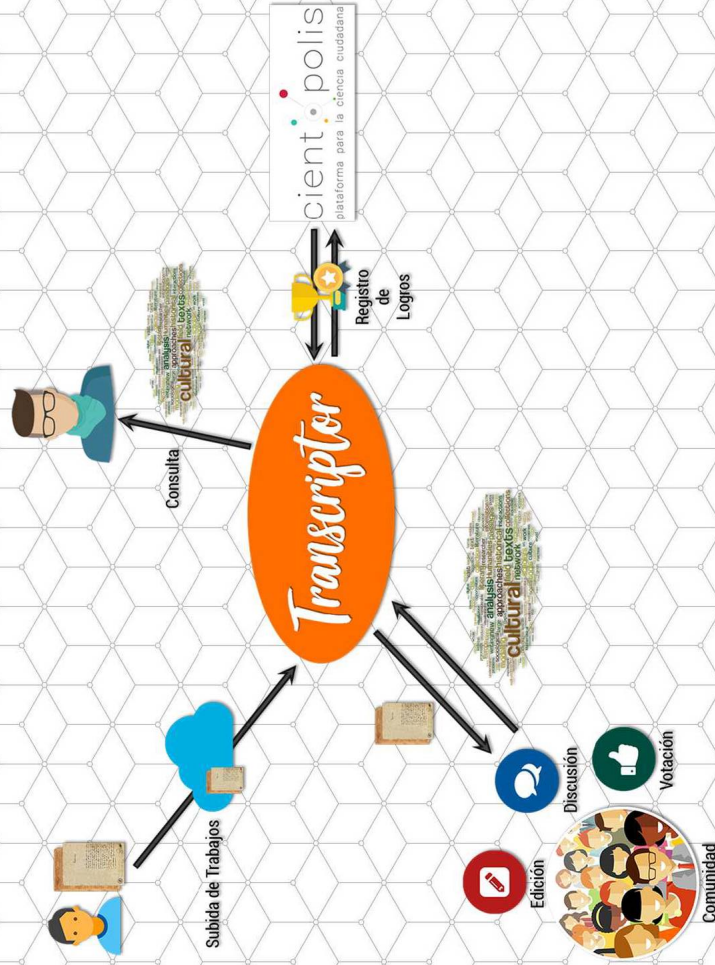
Sistema colaborativo para la transcripción de manuscritos históricos

Imaginate que tenes un montón de cartas o documentos manuscritos viejos. Que además, esas cartas las escribieron personas importantes en la historia, como por ejemplo, Manuel Belgrano, cuya carta podés ver en la imagen. Ahora imagináte que además queremos que el texto de esas cartas pueda encontrarse usando un buscador por Internet para que sea utilizado en las escuelas o en otro lugar.

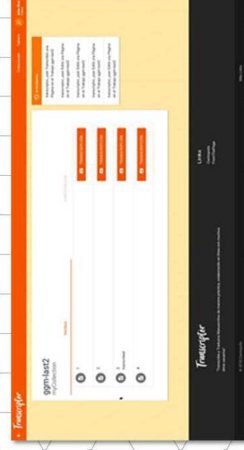
Transcriptor es un proyecto que busca crear una plataforma de ciencia ciudadana para realizar transcripciones de documentos que fueron digitalizados y que no han sido transcritos en forma automática. Así vas a poder copiar, pegar, buscar, consultar y editar contenidos propios o escritos hace décadas!



Flujos de trabajo



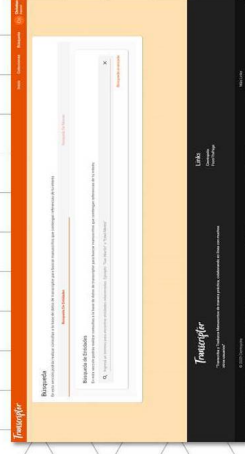
Funciones



Facil navegación a través de colecciones de manuscritos



Métodos de Transcripción enfocados en el Usuario



Recompensas y retos en la colaboración a la comunidad

Consulta semántica de la información transcrita

