Revista Internacional de Investigación e Innovación Tecnológica

Página principal: www.riiit.com.mx

Análisis Cualitativo de documentos de texto en español
Qualitative Analysis of Spanish text document

Hernández-Jiménez, G.[1], Cuevas-Rasgado, A.D.[1]*, Hernández-Espinosa, R.[2], Niño-Membrillo, Y.E.[1]

[1] Ingeniería en Computación, Universidad Autónoma del Estado de México, Centro universitario Texcoco, Av. Jardín Zumpango S/N El Tejocote, Texcoco, Estado de México, México C. P. 56259.

[2] Ciencias Sociales, Universidad Autónoma del Estado de México, Centro universitario Texcoco, Av. Jardín Zumpango S/N El Tejocote, Texcoco, Estado de México, México C. P. 56259

jgerardo.jmz@gmail.com; adcuevasr@uaemex.mx*; rhernandeze@uaemex.mx; yeninom@uaemex.mx

Innovación tecnológica: Método CodCual que emplea Procesamiento de Lenguaje Natural para mejorar y automatizar el análisis de extensos documentos de texto.

Área de aplicación industrial: En las ciencias sociales, estudios de mercado, análisis de opinión pública, educación, periodismo, facilitando la codificación e interpretación automatizada de textos cualitativos con procesamiento de lenguaje natural.

Recibido: 22 diciembre 2024
Aceptado: 07 septiembre 2025

Abstract

This paper approaches the problem of analyzing long texts, such as interviews, within the field of qualitative research. In social science studies, qualitative analysis tools are fundamental for organizing and extracting relevant information from textual corpora. These data are essential for research in diverse areas such as economic studies, market research, electoral opinions, and important social issues such as gender-based violence and the COVID-19 pandemic. The are some complex challenges in the analysis of opinions and testimonies, for example: the difficulty in distinguishing between opinions and facts, the structural complexity of interviews; and the time required by the researcher to manually read and analyze each interview. To address these challenges, the Qualitative Coding (CodCual in spanish) method has been developed, which incorporates Natural Language Processing (NLP) techniques, such as automatic sentence segmentation, using the NLTK (Natural Language Toolkit) library in the Python programming language. This segmentation was performed using the nltk.sent_tokenize() method, which divides the text into sentence units. This technique facilitates the identification and marking of relevant fragments in spanish and textual interviews. Although complex NLP models are not used, this approach speeds up manual coding, better organizes textual information, and reduces the

operational burden of the analysis. The effectiveness of CodCual was evaluated through testing with 145 students of social sciences, computer engineering and Informatic areas, in a test for two hours. Functionality with 75.2 percent + 16.6 percent and Usability with 57.9 percent + 28.3 percent are properties of quality of software evaluated demonstrating that the method can be effectively used by people with varying levels and areas of experience in qualitative analysis. The results highlight that CodCual significantly improves the efficiency of coding and information retrieval, allowing researchers and students to focus on in-depth analysis rather than exhaustive reading of texts. This makes CodCual a valuable tool for increasing productivity and accuracy in qualitative research.

Key words: Coding, Code, Interviews, Method, Qualitative analysis.

Resumen

El artículo aborda la problemática del análisis de textos extensos, como entrevistas, dentro del ámbito de la investigación cualitativa. En los estudios de ciencias sociales, las herramientas de análisis cualitativo son fundamentales para organizar y extraer información relevante de corpus textuales. Estos datos son esenciales para investigaciones en áreas diversas como estudios económicos, investigación de mercados, opiniones electorales y temas sociales importantes como la violencia de género y la pandemia de COVID-19. El análisis de opiniones y testimonios presenta desafíos complejos, tales como: 1) la dificultad para distinguir entre opiniones y hechos, lo cual puede afectar la objetividad de los resultados; 2) la complejidad estructural de las entrevistas, que pueden variar significativamente en formato y contenido; y 3) el tiempo considerable que debe invertir el investigador para leer y analizar manualmente cada entrevista, lo cual limita la eficiencia y la productividad. Para enfrentar estos desafíos, se ha desarrollado el método de Codificación Cualitativa (CodCual), que incorpora técnicas de Procesamiento de Lenguaje Natural (PLN), como la segmentación automática de oraciones, utilizando la biblioteca NLTK (Natural Language Toolkit) en lenguaje de programación Python. Esta segmentación permite dividir el texto en unidades oracionales, facilitando la identificación y el marcado de fragmentos relevantes en entrevistas en español. Si bien no se emplean modelos complejos de PLN, este enfoque permite acelerar la codificación manual, organizar de mejor manera la información textual para reducir la carga operativa del análisis. La efectividad de CodCual fue evaluada mediante pruebas con 145 usuarios, demostrando que el método puede ser utilizado eficazmente por personas con diferentes niveles de experiencia en análisis cualitativo. Los resultados de las pruebas de funcionalidad con 75.2%+16.6% y usabilidad con 57.9%+28.3% como propiedades de la calidad de los productos de software destacan que CodCual mejora significativamente la eficiencia en la codificación y recuperación de información, permitiendo a investigadores y estudiantes concentrarse en el análisis profundo en lugar de en la lectura exhaustiva de textos. Esto convierte a CodCual en una herramienta valiosa, gratuita y de software libre para incrementar la productividad y precisión en la investigación cualitativa.

Palabras clave: Análisis cualitativo, Código, Codificación, Entrevistas, Método.

1. Introducción

Las entrevistas semiestructuradas son un instrumento de recopilación de datos cualitativos ya que facilitan la obtención de datos a través del contacto cercano entre el investigador y los participantes, este enfoque resulta particularmente útil para reunir perspectivas valiosas que contribuyan al análisis y la identificación de descubrimientos relevantes en estudios de investigación cualitativa (Babativa Salamanca et al., 2024). Es por ello por lo que la entrevista "se trata de una técnica que se caracteriza por ser una conversación más o menos dirigida (según el tipo de entrevista) entre el investigador (remitente) y el sujeto de estudio (receptor) con un objetivo siempre bien determinado y centrado en la resolución de problemas, los objetivos y preguntas de investigación de trabajos" (Lopezosa, 2020). Por lo tanto, entre los desafíos cruciales que enfrentan investigadores y profesionales es la identificación precisa de párrafos relevantes en un vasto cuerpo de texto, especialmente en documentos con contenido de opiniones (entrevistas), el proceso actual suele ser manual o semiautomático. Para abordar este problema, se remite a la necesidad imperante de optimizar la eficiencia en la codificación mediante el análisis cualitativo, como se describe en (Rädiker & Kuckarts, 2020, pág., 195) y la recuperación de información (RI) que en literatura clásica como en (Pérez-Carballo & Strzalkowski, 2000) lo define como "una tarea de documentos relevantes desde un gran archivo en respuesta a una pregunta formulada y ordenarlos de acuerdo con su relevancia", mientras que (Baeza-Yates & Ribeiro-Neto, 1999), se centra más en la representación, el almacenamiento, la organización de la información y para el clásico (Salton, et. al., 1975) es una serie de documentos que satisfacen las necesidades de información expresadas por el usuario a partir de preguntas y respuestas. Todos ellos son ideas que reflejan la necesidad de identificar temas importantes en documentos relevantes, a través de un tema o concepto de interés como es la codificación.

El uso de la codificación automatizada permite a los usuarios identificar rápidamente los tópicos relevantes sin la necesidad de leer la transcripción completa y como menciona (Cházaro-Arellano, 2024) permite ubicar los datos de análisis. Un método utilizado para analizar las entrevistas es la codificación cualitativa, que permite la organización sistemática de la información recolectada, lo que implica identificar y registrar una o más citas de texto asociándolos con un nombre de código de esa idea o concepto, como lo menciona (Gibbs, 2012). La manera en cómo se comunican los seres humanos es importante para las áreas del conocimiento, en este sentido "el lenguaje natural es un tipo de lenguaje originado de forma espontánea. Es un tipo de discurso escrito o hablado distinto a los lenguajes formales (p.ej., lógica formal, lógica matemática) o los lenguajes de programación. Para analizar automáticamente el lenguaje humano se emplean técnicas de computación, en este ámbito conocidas como Procesamiento de Lenguaje Natural (PLN)" (Ibáñez de Opacua Lomoschitz, 2019). El uso del PLN es fundamental para permitir que las computadoras comprendan, interpreten y generen lenguaje humano de una manera valiosa, es así como el PLN es la fuerza motriz de la Inteligencia Artificial (IA) en muchas aplicaciones en el mundo actual. Por lo tanto, el objetivo de este artículo es presentar el método *CodCual* aplicado a entrevistas de textos en español. Se ha utilizado el ciclo de vida de desarrollo de software llamado: Scrum[1] definido como un marco ágil de trabajo en la gestión de proyectos de software en la utilización de un ciclo de vida adaptativo (Nader & Turley,

[1] ScrumGuides.org

2019), que ha permitido trabajar ciclos cortos llamados Sprint, cada uno resultando en una versión mejorada del software. El impacto del método *CodCual* es significativo en términos de eficiencia y productividad, ya que reduce notablemente el tiempo necesario para realizar investigaciones a través de la búsqueda cualitativa, así mismo permite a los usuarios centrarse más en el análisis y la interpretación de los datos, en lugar de invertir tiempo considerable en la lectura total de las transcripciones. Esta herramienta será útil beneficiando principalmente a los estudiantes y docentes del área de ciencias sociales de la Universidad Autónoma del Estado de México (UAEMex), así como también a personas de otros ámbitos como investigadores y periodistas, quienes pueden emplearla en actividades de investigación y toma de decisiones basadas en una información precisa.

El apartado “Marco Teórico” se centra en el codificado de palabras, un campo que une lingüística computacional, PLN e IA. Mencionando que desde los años 50 la investigación en materia ha avanzado gracias a grandes conjuntos de datos y algoritmos de aprendizaje automático, mejorando la captura semántica, sintaxis y otras propiedades lingüísticas, lo que ha facilitado el análisis eficiente de datos textuales y de voz.

El apartado “Materiales y Métodos” se describe el desarrollo de *CodCual* utilizando Visual Studio Code (VS Code), un entorno de desarrollo con soporte para depuración y edición de código, empleando Python como lenguaje de programación. Así mismo para la creación de *CodCual* se usó la investigación de tipo cualitativa, recopilando información de entrevistas y análisis de softwares de codificación.

El apartado "Trabajos Relacionados" estudia los programas de análisis cualitativo que permiten a los investigadores clasificar, filtrar datos sin procesar y crear informes. También se ofrece una visión general de investigaciones previas con estas herramientas, destacando su utilidad en diversas áreas. La revisión de estos trabajos ayuda a comprender las posibilidades y el alcance de las herramientas cualitativas.

El "Diseño" se enfoca en la funcionalidad central del método *CodCual*, mediante un diagrama de flujo que describe el diseño y la implementación del método de codificación cualitativa, destacando su capacidad para analizar y codificar textos basados en palabras clave (códigos).

En el área de “Métodos” se describe el funcionamiento de *CodCual* mediante diagramas de flujo que representan el proceso que se lleva a cabo en cada función realizada en específico por el codificador.

La sección "Pruebas" describe el uso de *CodCual* donde se abordan varios aspectos en la práctica como la aplicación de colores de subrayado a textos seleccionados, la visualización de las citas codificadas, y su recuperación manteniendo la coherencia del texto original. Además, se destaca la capacidad del método para gestionar datos cualitativos.

En la sección de “Resultados” se presenta los datos y descripciones de las respuestas obtenidas por los 145 usuarios que probaron el método *CodCual* en forma práctica, analizando las experiencias específicas de cada grupo sobre la funcionalidad del método. De tal forma que todos los usuarios lograron utilizar la aplicación de manera efectiva, a pesar de sus diferentes niveles de familiaridad con herramientas tecnológicas.

La sección "Discusión" del artículo analiza la eficacia de *CodCual* en la identificación y recuperación de citas de texto relevantes. Se describe el procedimiento de las pruebas.

Además, se discuten los resultados obtenidos a partir de las pruebas realizadas. También se abordan las métricas de funcionalidad y usabilidad, destacando la facilidad de uso y la eficiencia del método de codificación, así como la satisfacción en general que presentaron los usuarios con la herramienta y su capacidad para gestionar y organizar datos cualitativos.

2. Marco Teórico

En la metodología cualitativa en ciencias sociales la codificación es el procedimiento básico o fundamental en el proceso del análisis de los datos. Estos datos, invariablemente se presentan en forma de texto, ya sea que se hayan producido mediante entrevistas o incluso de observaciones directas. Por lo tanto "una entrevista es una metodología cualitativa que nos ayuda a recabar datos muy detallados y profundos, principalmente porque se adaptan al entrevistado desde aspecto muy variados como su contexto, su conocimiento sobre el tema, sus creencias, etc." (Lopezosa, 2020). Para poder analizar e interpretar los datos en información es primordial dentro de las investigaciones cualitativas, el análisis que implica la reducción de datos, el análisis descriptivo e interpretación, lo cual se logra mediante la categorización, codificación, registro y tabulación (Rueda Sánchez et al., 2023). Este proceso facilita la organización de los datos, permitiendo identificar patrones significativos que conducen a hallazgos en la investigación.

Para (Gibbs, 2012) la codificación "consiste así en identificar uno o más pasajes de texto que ejemplifiquen alguna idea temática y relacionarlos con un código, que es una referencia abreviada a la idea temática". Codificar es "la manera de indexar o categorizar el texto para establecer un marco de ideas temáticas sobre él" (Gibbs, 2012). Estas definiciones siguen siendo fundamentales, ya que permiten comprender cómo la codificación transforma los datos en significados organizados que posibilitan la construcción teórica.

Por lo tanto, la codificación le permite al investigador social identificar categorías abstractas que servirán como base para construir conceptos iniciales. A partir de estos conceptos, es posible formular hipótesis sobre las relaciones entre categorías, lo cual facilita el desarrollo progresivo de un modelo teórico interpretativo aplicado a un contexto social específico. Sin embargo, la tarea manual del codificado presenta limitaciones evidentes en términos de velocidad y precisión por lo que es esencial implementar técnicas que ayuden a satisfacer la necesidad del uso requerido de herramientas y métodos para obtener información. "Las primeras aplicaciones de PLN eran sistemas codificadas a mano y basadas en reglas que podían realizar ciertas tareas de PLN, pero no podían adaptarse fácilmente a un flujo aparentemente interminable de excepciones o a los crecientes volúmenes de datos de texto y voz" (IBM, 2024). El codificado de palabras es un campo de investigación que se encuentra en la intersección de la lingüística computacional (LC), el PLN, y la IA, los métodos como el codificado pueden capturar la semántica de las palabras (el significado), la sintaxis (la estructura gramatical), y otras propiedades lingüísticas.

El proceso de codificado tiene sus antecedentes en la investigación en LC y PLN que se ha desarrollado desde la década de 1950, esta área específicamente en el análisis y la generación de lenguaje se ha ido desarrollando significativamente en las últimas décadas debido al aumento de grandes conjuntos de datos y a los avances en algoritmos de aprendizaje automático. Es por ello por lo que, desde la década de 1980, se han buscado formas de aprovechar los desarrollos de la informática para facilitar

esta labor. Así mismo dentro de los años 80, la informática comienza a vislumbrarse como un auxiliar para la investigación cualitativa, esencialmente desde las primeras experiencias que tienen lugar en Inglaterra, Alemania, y Estados Unidos.

En la actualidad, el codificado e identificación de entidades se ha utilizado en diversas aplicaciones, como la extracción de información, la traducción automática y la generación de resúmenes de texto. El avance tecnológico ha facilitado la llegada de tecnologías como ChatGPT[2] de recopilación de datos OpenAI[3] o Gemini de Google[4], las cuales han tenido un impacto significativo en la investigación al proporcionar herramientas que abren nuevas vías de aplicación para una amplia gama de metodologías cuantitativas y cualitativas respaldadas por IA (Goyanes y Lopezosa, 2024).

Es así como los softwares informáticos se han convertido en valiosos aliados para la investigación cualitativa. Permiten ajustar y optimizar los procesos de sistematización e interpretación de textos. Además, contribuyen a la validez de los hallazgos al dejar rastros transparentes del proceso de indagación realizado por el analista.

2.1 Hipótesis de investigación

A partir de los antecedentes revisados sobre la codificación cualitativa y el uso de PLN en las ciencias sociales, se plantea la siguiente hipótesis de investigación: Si se aplica el método CodCual con segmentación automática de oraciones, los usuarios identificarán y codificarán fragmentos relevantes de manera más rápida y organizada que con la codificación manual. En esta hipótesis integrada por su variable causal categórica: método CodCual, como un conjunto de software, y las variables numéricas de efecto: tiempo, funcionalidad y eficiencia, parte de la necesidad de optimizar los procesos de análisis cualitativo, reduciendo el tiempo de operación del usuario y aumentando la eficiencia en su productividad.

2.2 Marco conceptual

La codificación cualitativa ha sido el eje del análisis de datos en ciencias sociales, ya que constituye el procedimiento fundamental para organizar e interpretar información textual (Gibbs, 2012; Rueda Sánchez, Armas & Sigala-Paparella, 2023). En este contexto, el Procesamiento de Lenguaje Natural (PLN) constituye un puente entre la metodología cualitativa y la informática, al automatizar tareas como la segmentación de oraciones y la extracción de información relevante (Ibáñez de Opacua Lomoschitz, 2019; IBM, 2024). Las herramientas CAQDAS (NVivo, MAXQDA, ATLAS.ti) han consolidado esta relación, integrando funciones que permiten a los investigadores gestionar, codificar y visualizar datos cualitativos (Lopezosa & Codina, 2023; Mativi, Niveiros & Carneiro Junior, 2020). Sin embargo, sus costos y la curva de aprendizaje se convierten en barreras para estudiantes y docentes (Marjaei, Yazdi & Chandrashekara, 2019; Consoli, 2021). Frente a ello, CodCual se plantea como una alternativa de software libre y gratuito, que integra técnicas de PLN para ofrecer funciones esenciales de codificación sin las limitaciones de acceso económico o técnico que se plantea en (Goyanes & Lopezosa, 2024).

El desarrollo de CodCual se basó en Python y la biblioteca NLTK (Natural Language Toolkit), herramientas utilizadas en el ámbito académico por ser accesibles y contar con abundante documentación. Python, como lenguaje de programación, destaca por su

[2] https://openai.com/es-ES/
[3] https://openai.com/es-ES/
[4] https://www.google.com/

sintaxis sencilla y clara, lo que lo convierte en una opción idónea en programación. Por su parte, NLTK es una biblioteca reconocida en el campo del PLN, ya que permite realizar tareas como la tokenización[5], segmentación y análisis de texto. Su uso en CodCual demuestra que es posible acercar técnicas de PLN a estudiantes de ciencias sociales y humanidades, ofreciendo un recurso educativo que no solo mejora la eficiencia en el análisis cualitativo, sino que también promueve la alfabetización digital en contextos universitarios.

3. Materiales y Métodos

La plataforma utilizada para el desarrollo del método *CodCual* fue VS Code (*versión 1.91.1*), que proporciona soporte para la depuración, resaltado de sintaxis, edición de código, fragmentos y refactorización de código de programación. El lenguaje de programación empleado fue Python, fungiendo como un lenguaje de programación de alto nivel, interpretado y de propósito general.

Dentro de las bibliotecas más importantes utilizadas en el desarrollo de la aplicación se hizo uso de: *filedialog*[6], *colorchooser*[7], *textwrap*[8], *nltk*[9], *docx*[10], *fitz*[11] y *os*[12]. Estas bibliotecas en conjunto permiten crear una aplicación robusta que interactúa con el sistema operativo, maneja archivos de diversos formatos y permite la manipulación de texto, colores y documentos de manera sencilla.

Por lo tanto, para el desarrollo del proyecto se realizó una investigación de tipo cualitativa, que tradicionalmente se ha utilizado en las ciencias empíricas. La investigación cualitativa se aplicó con base a la siguiente Figura 1.

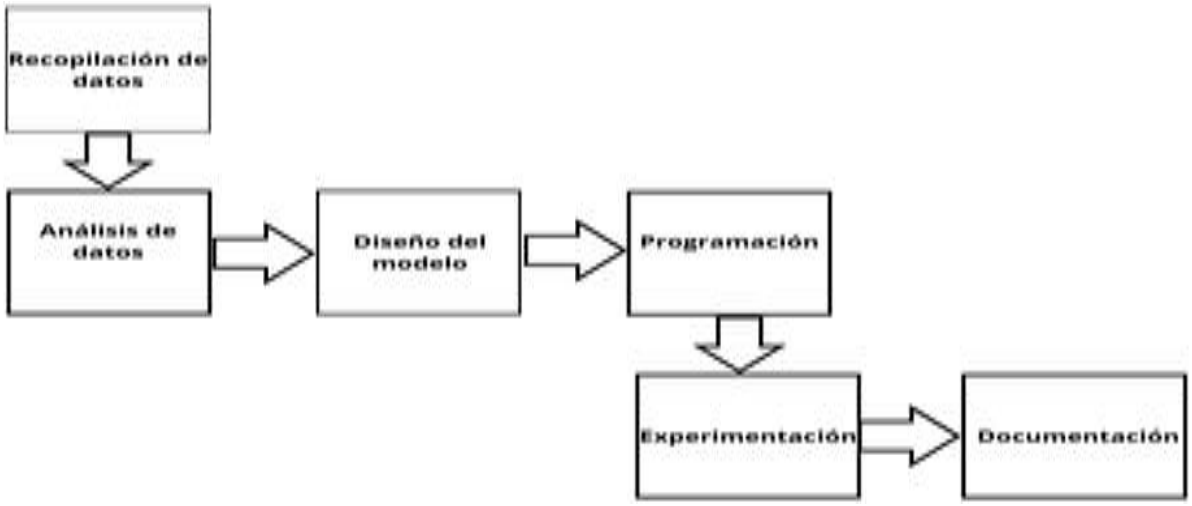


Figura 1. Diagrama de bloques de la secuencia de etapas de la investigación.

[5] Sustitución de datos sensibles por símbolos de identificación únicos que conservan toda la información esencial de los datos, https://www.sydle.com/es/blog/tokenizacion-65e8cd225d245147cf6e412e

[6] *Es la librería de Python que se encarga de facilitar la interacción con archivos y sistema de archivos.*

[7] *Librería que permite a los usuarios seleccionar colores a través de la interfaz gráfica.*

[8] *Librería que ayuda a ajustar a ajustar y organizar el texto para que se vea bien en la pantalla, evitando que las líneas sean demasiado largas y difíciles de leer.*

[9] *Es la librería que facilita el análisis automático del lenguaje escrito para extraer información relevante en investigaciones sociales.*

[10] *Librería que permite manipular archivos de Word, útil para automatizar la generación de informes, editar documentos existentes o extraer información.*

[11] *Es la librería que puede usarse para analizar documentos en formato PDF, como artículos, encuestas o reportes, de forma automatizada.*

[12] *Librería que permite organizar archivos de datos, mover o renombrar documentos, y acceder a diversos directorios.*

3.1 Recopilación de datos

Se recopilaron datos cualitativos sobre la codificación en contenido de opiniones, como:

- *Análisis de documentos de entrevistas.* El análisis se ha utilizado para identificar los desafíos que los usuarios encuentran al codificar o buscar información importante.
- *Análisis de softwares que realizan codificación.* Se ha observado la funcionalidad y la usabilidad en estos tipos de documentos y el uso de las herramientas de preprocesamiento previo al codificado y búsqueda semántica. En este sentido se ha analizado los sistemas *NVivo, MAXQDA y ATLAS. Ti*. Sin embargo, el buen desempeño de estos depende de su eficacia y de la calidad de los datos de entrada. Si estos datos no están bien organizados, estructurados o codificados de manera adecuada, el sistema puede generar resultados imprecisos o incompletos.

3.2 Análisis de datos

Se recopilaron los datos, posteriormente se analizaron para identificar patrones y tendencias. Se usaron técnicas cualitativas, como:

- *Análisis de contenido,* para identificar temas y patrones en los datos. Mediante el método *CodCual* fueron codificadas y categorizadas las respuestas de una entrevista abierta, lo que ha permitido identificar temas recurrentes como el entorno de vida de las personas.
- *Análisis de discurso,* para comprender el significado de los datos en el contexto. *CodCual* se ha aplicado para analizar transcripciones de entrevistas, lo que ha ayudado a entender cómo las personas perciben el contexto de sus propias experiencias y expectativas de vida.

3.3 Diseño del Modelo

A través de reuniones con un experto en el manejo de codificación en documentos de entrevistas se fueron diseñando las interfaces gráficas, el proceso de identificación de palabras, frases y oraciones, el diseño de los diagramas de flujo de datos. La identificación de conceptos y marcado de párrafos. Cada idea y recomendación se fue transformando en pasos para cumplir el objetivo general: Diseñar un método de codificación cualitativa de textos de documentos en español funcional, gratuito y accesible, la Figura 2 muestra los pasos del modelo.

3.4 Programación

Cada diagrama de flujo de datos se transformó en código de programación, bajo la norma de la programación orientada a datos, con métodos que a su vez usan otros métodos de PLN. Cada fragmento de código fue mostrado al experto para comprobar el cumplimiento del objetivo. Se hicieron adaptaciones y cambios constantemente usando el ciclo de vida Scrum hasta asegurarse que se cubrieran todos los requisitos planteados por el experto.

3.5 Experimentación

La etapa de experimentar la funcionalidad del modelo contempló la elección de forma aleatoria de una población de 145 estudiantes de la Licenciatura en Administración, Informática e Ingeniería en Computación con distintos niveles de comprensión de textos según el semestre cursando, que bajo la inspección de sus profesores analizaron 6 archivos con entrevistas textuales y con preguntas abiertas a personas sobre cómo se desarrollaron sus actividades cotidianas durante la pandemia de COVID-19. Esta etapa se desarrolló en diferentes horarios, pero de forma grupal en ambiente dedicado como los laboratorios de cómputo. Durante 2 horas en un determinado día se desarrolló la experimentación con el grupo de

Administración, luego 2 horas en distintos días con los otros 2 grupos Informática e Ingeniería.

El muestreo fue no probabilístico por conveniencia, al seleccionarse grupos previamente organizados y disponibles en los horarios establecidos. Esta estrategia permitió garantizar la participación en condiciones homogéneas y controladas.

3.6 Documentación

Derivado de la experimentación se obtuvieron resultados representados en gráficas y porcentajes para solo 2 de las características según la Ingeniería de Software, tales como: la Funcionalidad, la Usabilidad. Otras características como la Eficiencia, Confiabilidad y Mantenibilidad no se encuentran documentados en este artículo.

El desarrollo del método CodCual se diseñó bajo un enfoque de código abierto, de modo que cualquier investigador pueda emplearlo, o mejorarlo. Tanto el código fuente como una versión ejecutable se encuentran disponibles en línea para su descarga[13].

4. Trabajos Relacionados

El desarrollo de las tecnologías de la información y la comunicación han propiciado el desarrollo de programas cuya evolución ha sido descrita en algunos trabajos previos (Cisneros Puebla, 2003; 2011; Cisneros Puebla y Davidson 2012); se trata de programas capaces de codificar electrónica y semiautomáticamente datos ayudando a que los estudios del investigador ganen en rigor y eficacia (Makkar y Powell, 2024).

Los programas de *Computer-Assisted Qualitative Data Analysis Software*, por sus siglas en inglés (CAQDAS) o *Software de análisis de datos cualitativos asistido por computador*, facilitan a los usuarios ejecutar muchas operaciones analíticas tales como clasificar y filtrar datos, así como hallar y crear relaciones entre ellos, asignar y definir temas y categorías para la semántica de los datos, la muestra de los resultados del análisis de datos y al final presentar los informes.

“Cabe señalar que las aplicaciones CAQDAS cuentan con dos elementos clave sobre los que gira todo el proceso: los códigos, y los informes que se producen como resultado de aplicar esta codificación a las entrevistas” (Lopezosa y Codina, 2023).

A través de la Tabla 1, se exploran las capacidades y funcionalidades de los programas *NVivo*, *MAXQDA* y *ATLAS.ti*, en el ámbito del análisis cualitativo de datos se ha descrito su funcionamiento frente al método desarrollado de *CodCual*.

Tabla 1. Comparativa de funciones entre herramientas de análisis cualitativo.

Característica	NVivo	MAXQDA	ATLAS.ti	CodCual
Importar archivos de texto (docx, PDF, txt)	✓	✓	✓	✓
Importar audio y video	✓	✓	✓	✗
Codificación manual por selección de texto	✔	✓	✓	✓

[13] *https://doi.org/10.5281/zenodo.17013099*

Asignación de códigos por colores	✓	✓	✓	✓
Contador de citas codificadas	✗	✗	✗	✓
Segmentación automática de oraciones (PLN)	✗	✗	✗	✓
Visualización de códigos y citas codificadas	✓	✓	✓	✓
Subrayado visual y alteración del texto codificado	✗	✓	✓	✓
Consulta o búsqueda de texto codificado	✓	✓	✓	✓
Visualización mediante mapas o redes	✓	✓	✓	✗
Creación de memos o notas analíticas	✓	✓	✓	✗
Anexar un código dentro otro (Anidación)	✓	✓	✓	✓
Guardado y exportación de citas codificadas	✓	✓	✓	✓

Por lo tanto, cabe señalar que la ventaja clave del método CodCual es que es un método de código abierto y gratuito, lo que lo hace accesible frente a NVivo, MAXQDA y ATLAS.ti, que tienen costos de licencia significativos. Aunque estas últimas ofrecen funciones avanzadas, el método desarrollado se enfoca en lo esencial e importante: la codificación y recuperación de citas o fragmentos en documentos de textos codificados. En conjunto, CodCual ofrece una alternativa ligera, de fácil uso y centrada en la codificación y recuperación eficiente de citas textuales, lo cual lo convierte en una opción viable para investigadores que requieren una herramienta práctica, es más simple y directa, ideal para usuarios que no necesitan la complejidad o la curva de aprendizaje asociada a las herramientas de las opciones comerciales.

En la Tabla 2, se muestran los trabajos relacionados de algunos de los estudios más destacados sobre esta investigación de herramientas, ordenados cronológicamente que ayudan a comprender las posibilidades investigativas y el alcance de las herramientas de análisis cualitativo. La lista contempla tanto trabajos comparativos y de análisis de funcionalidades como de estudios de caso de uso.

Tabla 2. Investigaciones y características de las herramientas de análisis cualitativo**.**

Trabajo relacionado	Características principales
Marjaei (2019)	Se realizó una revisión de MAXQDA centrada principalmente en el desarrollo de estudios de LIS (*Library and Information Sciences*). En este estudio se confirmó que los investigadores de LIS necesitan respaldar y mejorar sus estudios preparando los datos cualitativos con MAXQDA, codificando con su sistema de codificación y organizando los datos para el análisis, utilizando programas para el análisis cualitativo asistido por computadora.
Mativi, Niveiros & Carneiro Junior (2020)	Realizaron una revisión sistemática de artículos publicados en SciELO y SPELL sobre el uso de herramientas CAQDAS en estudios empíricos en administración. Identificaron que NVivo, Atlas.ti y MAXQDA fueron los softwares más empleados, valorando su aporte a la codificación, organización y triangulación de datos. El trabajo ofrece una mirada crítica y actualizada sobre el uso de CAQDAS en el ámbito investigativo.
Consoli (2021)	Utilizó MAXQDA como herramienta para llevar a cabo investigaciones narrativas y para examinar su epistemología. El uso de este software ayudó a facilitar el análisis narrativo al tratar con un gran conjunto de datos de diferentes fuentes, y también permitió legitimar la metodología narrativa como un enfoque investigativo riguroso.
Lopezosa, Codina & Freixa (2022)	Desarrollaron una propuesta metodológica para el uso de ATLAS.ti en entrevistas semiestructuradas. Ofrece una guía paso a paso de codificación y categorización que busca reforzar la rigurosidad del análisis cualitativo.
Lopezosa & Codina (2023)	Analizaron la integración de ChatGPT con programas CAQDAS como ATLAS.ti, NVivo y MAXQDA. Su estudio describe cómo la IA puede apoyar tareas como la pre-codificación, generación de categorías y organización narrativa, sin sustituir el juicio analítico del investigador. Esta propuesta metodológica aporta una vía práctica para optimizar el análisis cualitativo incorporando IA sin comprometer el rigor interpretativo.
Sun & Peng (2023)	Estudio metodológico en el ámbito de la investigación en lenguas extranjeras, en el que se aplicó un enfoque de análisis mixto combinando herramientas como SPSS (*Statistical Package for the Social Sciences*) para el tratamiento de datos cuantitativos y Atlas.ti para el análisis cualitativo. A través de esta integración, los autores reflexionan sobre las fortalezas y limitaciones de ambos enfoques, subrayando cómo Atlas.ti permitió una codificación de las percepciones de los estudiantes, mientras que SPSS facilitó la sistematización de datos estadísticos.
Goyanes & Lopezosa (2024)	Presentan una revisión sistemática sobre el uso de la IA, especialmente ChatGPT, en investigaciones cualitativas y cuantitativas en ciencias sociales. Subrayan su integración con herramientas CAQDAS como NVivo, MAXQDA y ATLAS.ti para apoyar tareas de codificación y análisis temático, destacando beneficios, riesgos y desafíos éticos de su aplicación en el análisis cualitativo.
Lopezosa, Goyanes & Codina (2024)	Crearon una guía práctica sobre el uso de IA en el análisis de entrevistas cualitativas. Propone un protocolo paso a paso para el uso de herramientas como ChatGPT, Whisper y Gemini junto con herramientas CAQDAS.
Makkar & Powell (2024)	Exploran cómo los CAQDAS pueden hacer frente a los retos que surgen desde un procedimiento manual tradicional, esbozando métodos de codificación cualitativa y el papel de variantes específicas de los CAQDAS en este proceso. Destacan ventajas y desventajas del uso de CAQDAS en el análisis cualitativo para investigadores principiantes.
Rodríguez Garay & Slimovich (2025)	Revisión sistemática sobre programas CAQDAS (como MAXQDA, NVivo) aplicados al análisis de documentos pre-digitales. Incluye un estudio de caso que compara el desempeño de softwares en la codificación de textos escaneados mediante OCR (*Optical Character Recognition*).
Durán, R. J., Sandoval & Alarcón, L. (2025)	Llevaron a cabo un estudio de caso que utiliza análisis cualitativo asistido por software (NVivo y matrices en Excel) para explorar percepciones juveniles. Aunque no usa directamente CAQDAS, aplica codificación abierta, análisis temático y teoría fundamentada con procedimientos sistemáticos equivalentes.

A diferencia de los trabajos revisados en esta sección, CodCual no solo se limita a analizar o utilizar herramientas comerciales de análisis cualitativo, sino que propone el

desarrollo e implementación de una herramienta propia, de código abierto, diseñada específicamente para la codificación de entrevistas en español. Mientras los estudios anteriores se enfocan en el uso de programas comerciales establecidos o en discusiones metodológicas sobre su aplicación, CodCual incorpora una funcionalidad técnica concreta basada en PLN: la segmentación automática de oraciones mediante NLTK en Python, lo que permite trabajar con unidades mínimas de análisis desde el momento de cargar el texto.

Además, introduce un contador dinámico de codificaciones y un sistema de subrayado con asignación visual de códigos, facilitando la gestión de fragmentos codificados en tiempo real. CodCual, además de ser funcional y usable, es una propuesta técnica ligera, accesible y adaptable, orientada a investigadores que requieren funciones esenciales sin depender de licencias comerciales ni curvas de aprendizaje elevadas.

En este contexto, la principal aportación de CodCual es ofrecer una herramienta gratuita, de código abierto y en español para el análisis cualitativo. Su diseño sencillo y accesible permite prescindir de licencias costosas y facilita la codificación sin requerir conocimientos técnicos avanzados. De este modo, contribuye a democratizar el acceso al análisis cualitativo y a resolver el desafío recurrente de analizar manualmente documentos en español, especialmente en investigaciones en el área de las ciencias sociales.

5. Diseño del Modelo

La Figura 2, representa el diagrama general del modelo y está diseñado para procesar documentos de contenido de opiniones, así como textos narrativos para analizar y extraer lo más importante del texto, identificando a los párrafos u oraciones con palabras clave, de esta forma se va codificando en el proceso el archivo de texto importado, con el objetivo de realizar un análisis detallado.

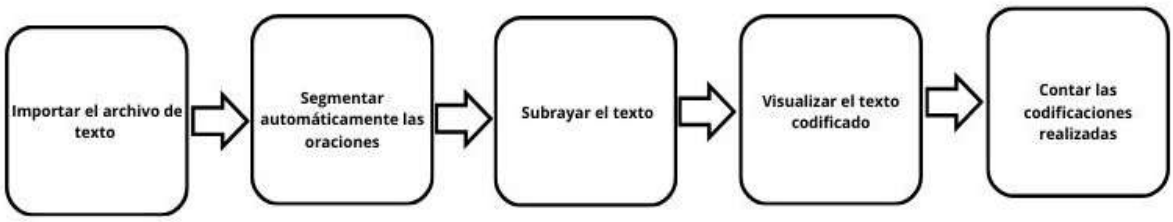


Figura 2. Diagrama general del modelo en donde se muestran los pasos principales.

El algoritmo de *CodCual*, que se presenta en la Figura 3, el proceso comienza con la ***importación del documento***, a medida que el usuario ***selecciona el párrafo o fragmento*** y lo ***codifica***, el sistema aplica el código correspondiente asignado por el usuario a la sección pertinente del texto codificando así el fragmento.

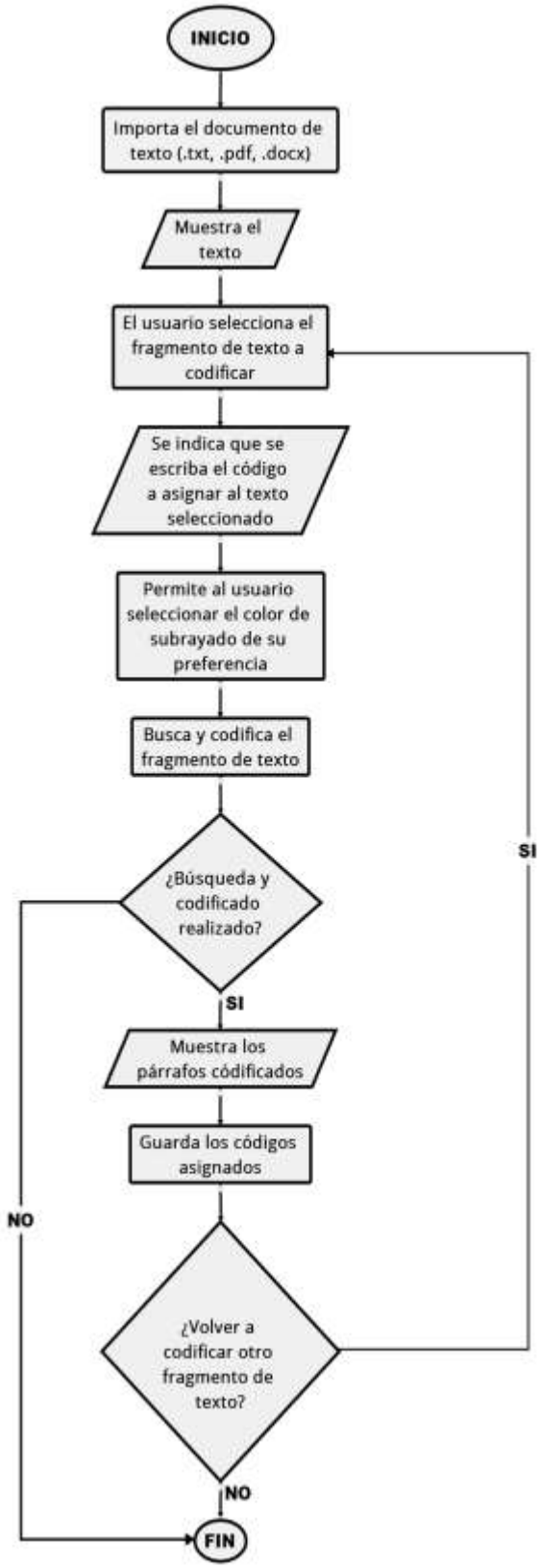


Figura 3. Diagrama de flujo de la funcionalidad bajo el método principal del codificado cualitativo.

6. Métodos

6.1 Segmentación automática de oraciones (tarea de PLN)

En el proceso de codificación, se implementó una técnica de Procesamiento de Lenguaje Natural (PLN) basada en la segmentación automática de oraciones, como se ilustra en la Figura 4. Esta funcionalidad se activa al momento de cargar un archivo de texto, ya sea en formato .docx o .pdf, y permite dividir el contenido completo en oraciones individuales.

Para ello, se empleó el tokenizador de oraciones de la biblioteca NLTK, el cual identifica automáticamente los límites entre enunciados. Como resultado, se genera una lista de oraciones que se muestra de forma organizada en la interfaz del sistema, facilitando su análisis.

Esta segmentación previa agiliza la revisión del texto, permitiendo al usuario seleccionar y codificar manualmente fragmentos específicos con mayor claridad y precisión, reduciendo la carga cognitiva asociada al análisis de textos extensos.



Figura 4. Diagrama del proceso de segmentación y codificación de oraciones.

Funciones complementarias al proceso de codificación

Además de la segmentación automática como tarea principal de PLN, el método CodCual incorpora funcionalidades adicionales diseñadas para mejorar la interacción del usuario con el texto y optimizar la experiencia de codificación. A continuación, se describen tres de estas funciones clave:

6.2 Subrayado de texto

Al codificar un fragmento de texto, se utilizó el algoritmo *subrayado de texto* de *CodCual* cuya función se ilustra en la Figura 5, este algoritmo es capaz de detectar el rango seleccionado de texto de inicio a fin.

Una vez identificado este rango, se aplica un color de subrayado específico además de modificar el color de la letra para resaltar el texto de manera efectiva. Este proceso permite una mejor visualización y diferenciación del texto subrayado, facilitando así la lectura y el análisis del contenido destacado.

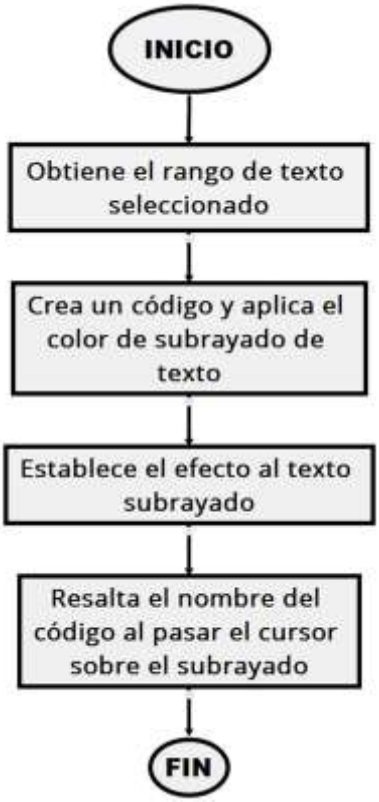


Figura 5. Diagrama acerca de la función de subrayado del fragmento de texto.

6.3 Visualización del texto codificado

El proceso se lleva a cabo cuando se ejecuta la funcionalidad representada en la Figura 6, en la que se ilustra el método que se encarga de ordenar que tanto el código y su cita resultante se presenten de forma clara y visible en la interfaz.

Este método es crucial para asegurar que los usuarios puedan navegar y utilizar la aplicación de forma eficiente. Al mostrar el texto y la cita de manera organizada, se facilita la comprensión y el acceso a la información, mejorando así la experiencia del usuario.



Figura 6. Diagrama de la función para visualizar el código y las citas generadas.

6.4 Contador de codificaciones realizadas

En el diagrama mostrado en la Figura 7, se describe una funcionalidad esencial para el sistema la cual consiste en llevar un registro preciso de la cantidad de fragmentos del texto que han sido codificados haciendo uso de un mismo código asignado o de uno diferente, y a medida que se aplican los códigos mediante el método *CodCual*, este contador se actualiza dinámicamente, reflejando en tiempo real el número de citas almacenadas que hay en cada código en específico. El propósito principal de este contador es asegurar un seguimiento eficiente y preciso del proceso de codificación, facilitando la gestión y el análisis de los datos codificados.

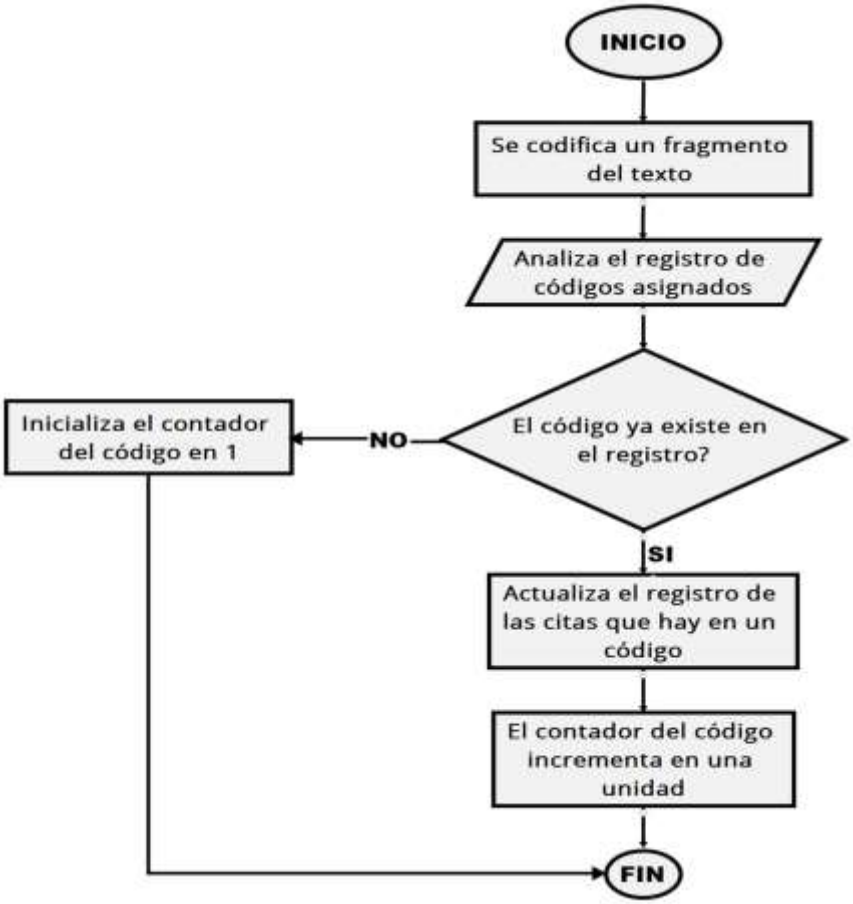


Figura 7. Diagrama del contador de códigos que incrementa en una unidad según el caso.

Con el fin de ilustrar de manera más clara la lógica interna del sistema, en la Figura 8 se presenta un fragmento de programación en Python responsable de la funcionalidad de codificado y subrayado de fragmentos textuales, parte del proceso de codificación cualitativa en CodCual. Este código permite al usuario asignar etiquetas personalizadas a los fragmentos seleccionados, aplicarles un color distintivo y desplegar un tooltip (mensaje emergente) con información contextual al pasar el cursor sobre el texto subrayado.

src > Interfaz > Interfaz_CodCual > EtiquetadoApp > cambiar_cursor_segun_posicion

```
87  class EtiquetadoApp:
368
369     def etiquetar_fragmento(self):
370         # Solicita al usuario el código de la etiqueta
371         etiqueta = simpledialog.askstring("Codificar", "Escribe un Código:")
372         if etiqueta:
373             # Elige color para el subrayado
374             color_subrayado = self.elegir_color_subrayado()
375             self.etiqueta_actual = etiqueta
376
377             # Inserta la etiqueta en el panel de etiquetas
378             self.texto_etiquetado.insert(
379                 tk.END, f'>>>({etiqueta})<<<\n', "negrita")
380             self.texto_etiquetado.tag_configure(
381                 "negrita", font=("Arial", 12, "bold"))
382
383             # Aplica el subrayado al fragmento seleccionado
384             tag_name = self.aplicar_subrayado(color_subrayado)
385
386             # Crea un tooltip asociado al fragmento
387             tooltip = Tooltip(self.texto_original, etiqueta)
388             self.texto_original.tag_bind(tag_name, "<Enter>",
389                 lambda event, t=tooltip, tn=tag_name: t.show_tooltip(event, tn))
390             self.texto_original.tag_bind(tag_name, "<Leave>", tooltip.hide_tooltip)
391             self.texto_original.tag_bind(tag_name, "<Motion>", tooltip.update_position)
392
393             # Guarda y actualiza la lista de fragmentos codificados
394             self.guardar_subrayados()
395             self.actualizar_lista_etiquetado()
396
```

Figura 8. Fragmento de programa en Python que implementa el proceso de codificado y subrayado de fragmentos textuales del método CodCual.

7. Pruebas

En las pruebas se explica cómo se ha empleado *CodCual* para codificar y organizar las citas del código permitiendo la visualización y recuperación de estas. Este proceso de codificación ha ayudado a clasificar y agrupar mediante citas relevantes, haciendo más accesible la búsqueda de datos clave dentro del sistema para su respectivo análisis.

7.1 Fragmento de texto subrayado

Al avanzar la investigación, la codificación inicial se ha hecho más nutrida y matizada, justamente este proceso de enriquecimiento progresivo es útil para distinguir la información clave dentro del texto. Es así

como se ha codificado un fragmento de texto al seleccionar un rango en específico, el cual se puede diferenciar del resto del texto debido al color aplicado con el subrayado y código, mostrando el nombre del código a través de un rectángulo de color amarillo cuando se pasa el cursor sobre el fragmento de texto.

7.2 Visualización de las Citas de Código

Al codificar, las citas resultantes se colocan como se van creando a través del método *CodCual*, en la sección: *Citas del Código* de la interfaz para su respectiva visualización y almacenamiento. De esta forma cuando se pulsa el botón generado del código asignado, que se encuentra en la sección de la *Lista de Codificaciones Realizadas* se pueden consultar las citas del código. Las citas consultadas del código en cuestión, se visualizan una seguida de la otra en el orden de la codificación aplicada y separadas por un salto de línea que marca el término de una cita, sin embargo, al pertenecer al mismo código aparecen agrupados con un solo código. Cabe destacar que las citas pueden provenir de diferentes partes del documento y que si se asignan a un mismo código se recuperan las citas correspondientes. Este enfoque garantiza que, a pesar de la codificación la esencia del texto original se mantiene intacta, permitiendo una interpretación precisa y efectiva. La aplicación de técnicas de (PLN) ha podido mejorar significativamente el proceso de recuperación de citas con su código correspondiente.

7.3. Contador de citas registradas por código asignado

En la práctica *CodCual* contabilizó cuántas citas acordes a ese código se han codificado, siendo útil en la parte estadística. De tal forma que por cada código asignado se registró el número de citas que contiene dicho código. Una prueba de ello es el código "*ENTREVISTAS*" en el área, el cual tiene almacenadas a dos citas con el mismo código aplicado visualizándose en el texto.

Finalmente, dentro de las pruebas *CodCual* facilitó la selección y codificación de distintos fragmentos de texto con los códigos correspondientes. Se aplicaron diversos códigos a oraciones de texto, utilizando la interfaz de la aplicación, como se muestra en su totalidad la Figura 9.

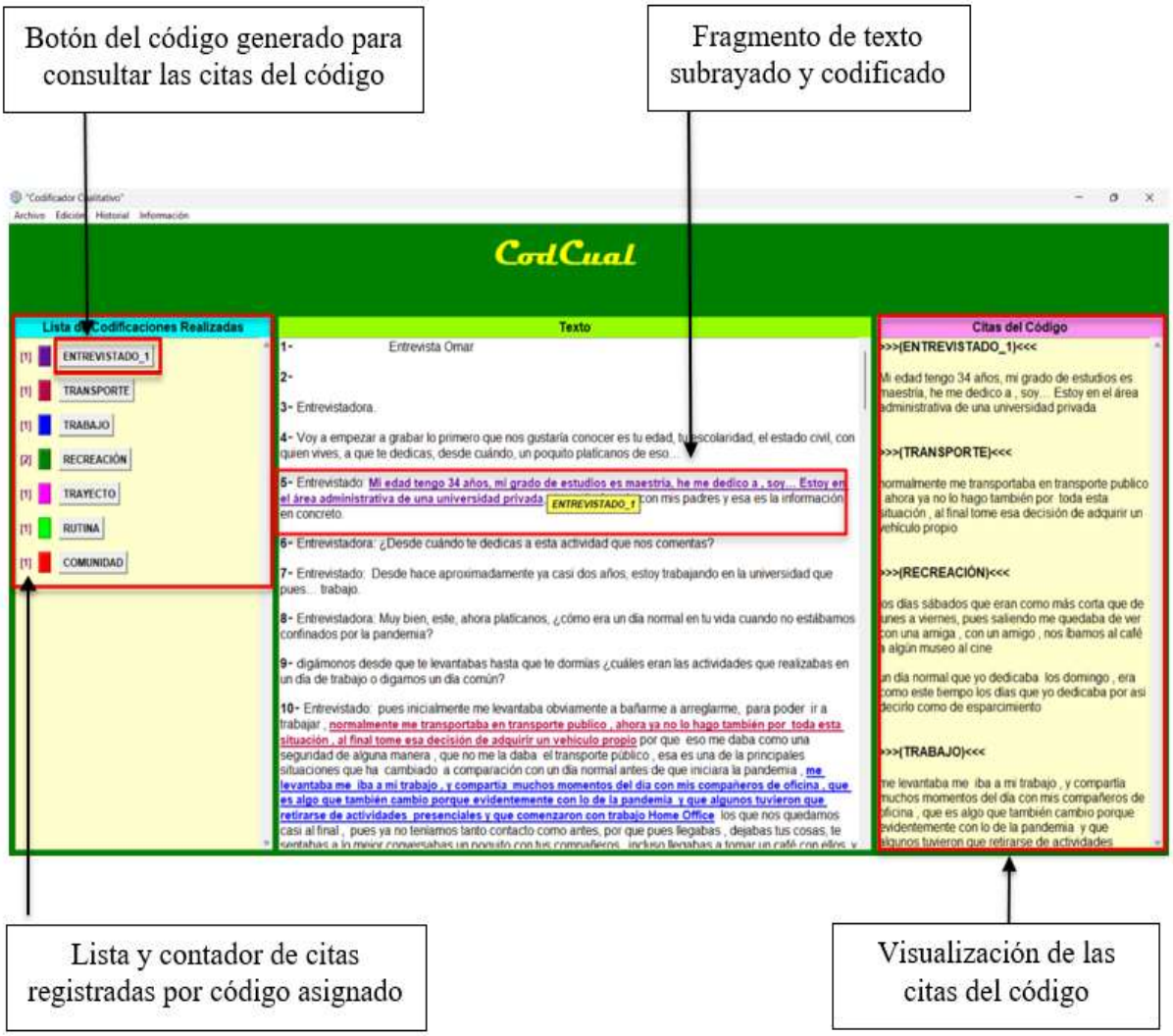


Figura 9. Interfaz en ejecución dentro de las pruebas, con subrayados de texto, códigos y citas**.**

8. Resultados

8.1 Ejemplo práctico de aplicación del método CodCual

Con el objetivo de ilustrar de manera clara el funcionamiento del método CodCual y su coherencia con los objetivos planteados, se presenta a continuación un caso práctico de ejemplo utilizando un fragmento de entrevista aplicada durante la etapa de pruebas.

8.1.1 Carga del documento

El usuario abre la aplicación CodCual y selecciona un archivo de texto de formato .docx que contiene la transcripción completa de una entrevista. La Figura 10 muestra el contenido de la entrevista en el rectángulo del centro.



Figura 10. Carga del documento de entrevista.

8.1.2 Selección y codificación

Al codificar una oración el sistema aplica automáticamente la segmentación de oraciones mediante la biblioteca NLTK. En este caso el usuario identifica un fragmento relevante, lo selecciona y le asigna el código *"GENTRIFICACIÓN"*, aplicando el subrayado en color morado, como se muestra en la Figura 11. Posteriormente, el usuario codifica otra oración con el código *"CONSECUENCIAS"*, utilizando esta vez el color rojo.

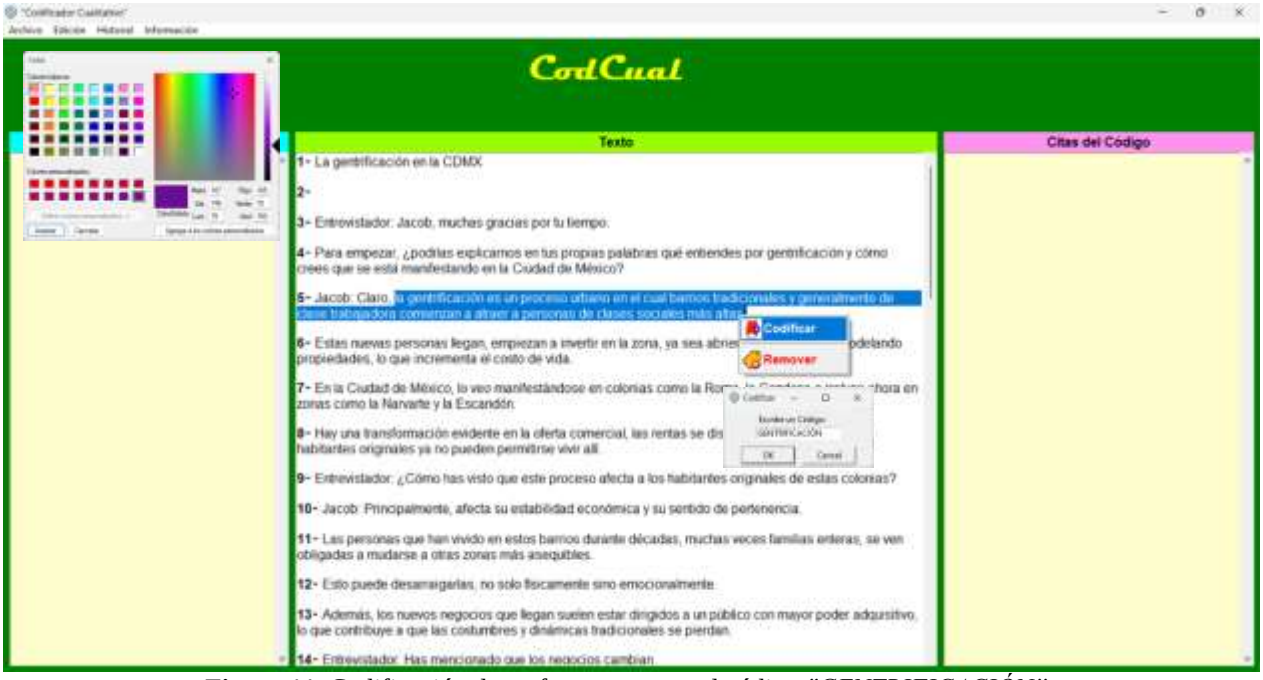


Figura 11. Codificación de un fragmento con el código *"GENTRIFICACIÓN"*.

8.1.3 Visualización y contador de codificaciones realizadas

El código recién creado se agrega automáticamente a la lista de codificaciones realizadas. Al hacer clic en el botón de una cita específica, el sistema despliega todas las citas del documento que comparten ese mismo código. El contador dinámico indica, por ejemplo, que el código *"GENTRIFICACIÓN"* cuenta con 3 citas registradas provenientes de distintas partes del texto, la Figura 12 lo presenta en el rectángulo de lado derecho con el título *Citas del Código*.

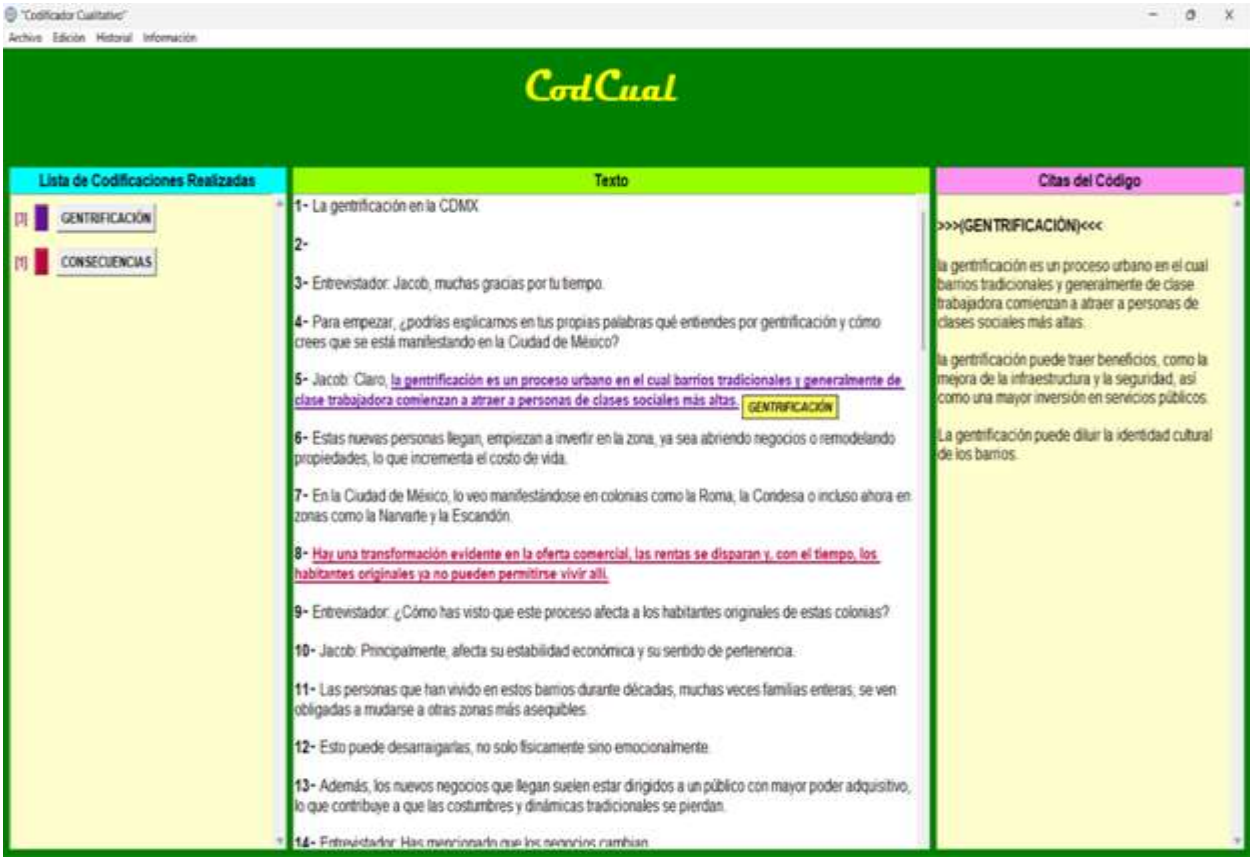


Figura 12. Visualización en la interfaz de las citas generadas y el contador de citas correspondiente.

8.1.4 Salida y exportado de citas codificadas.

El CodCual permite guardar y exportar las citas codificadas correspondientes a uno o varios códigos específicos. Este listado puede generarse en un archivo para su análisis posterior, la Figura 13 presenta el contenido de este archivo de texto.

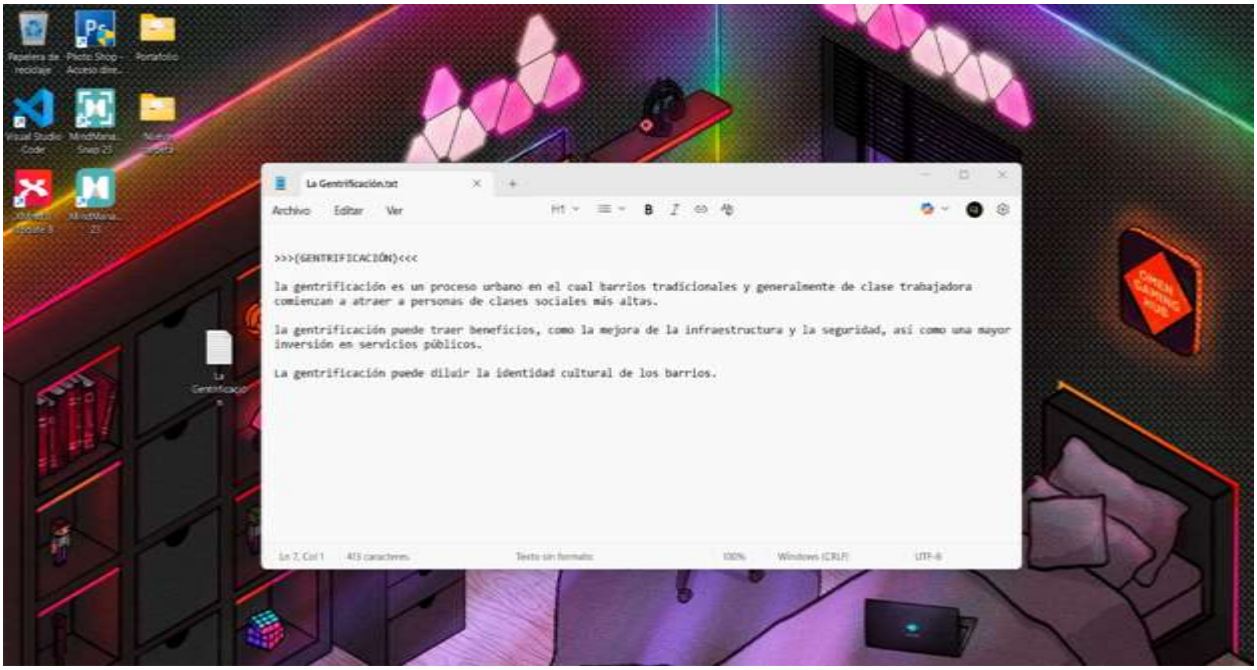


Figura 13. Almacenamiento de citas codificadas para su análisis posterior.

En este ejemplo, el método CodCual permitió identificar y agrupar de forma rápida y organizada fragmentos relacionados con la adaptación de rutinas durante la pandemia, reduciendo el tiempo de búsqueda y mejorando la claridad en la interpretación de los datos.

8.2 Pruebas con usuarios.

La Funcionalidad según la ISO/IEC 25000[14] es el grado en que un producto satisface los objetivos y tareas de los usuarios. La Usabilidad de un producto de software es la capacidad que tiene para ser entendido, aprendido, utilizado y resultar atractivo para el usuario. Estas dos características de calidad del software fueron evaluadas mediante una encuesta por parte de 145 estudiantes de la UAEMex- Este grupo fue dividido en seis grupos pertenecientes a tres diferentes disciplinas, por lo que dos grupos por carrera universitaria utilizaron la aplicación en un entorno educativo. La Lic. en Administración, Lic. en Informática y en Ing. en Computación, del CU Texcoco.

En general, *CodCual* fue recibido de buena forma por los usuarios de las tres carreras, destacándose por su facilidad de uso y su capacidad para realizar el proceso de codificación cualitativa.

En cuanto al desempeño de la herramienta en cada uno de los grupos evaluados, se observaron los siguientes resultados en base a la prueba realizada en el laboratorio:

- Usuarios de Administración: Los grupos de esta carrera reportaron una curva de aprendizaje moderada, con base a su primera experiencia en el uso de la aplicación desarrollada. A pesar de ello, apoyados con el manual de usuario lograron utilizar *CodCual* con éxito para generar citas de código, resaltando la utilidad en las

[14] *ISO/IEC 25000 es la norma ISO del Comité de Ingeniería Electrotécnica que se encarga también de regular el aspecto de software. Esta norma organiza y mejora las guías que se enfocan en dos procesos principales: definir claramente tan bueno es un software y por el otro, evaluar si el software realmente cumple con las expectativas de calidad.*

investigaciones relacionadas con estudios sociales.

- Usuarios de Computación: Los estudiantes de Ing. en Computación realizaron un análisis técnico de *CodCual*, identificando puntos fuertes y áreas a trabajar desde una perspectiva de ingeniería de software. Si bien destacaron la eficiencia del algoritmo de codificación, también sugirieron posibles optimizaciones a futuro.
- Usuarios de Informática: Los estudiantes de Informática demostraron un buen nivel de competencia con la herramienta, evaluando positivamente su capacidad para gestionar volúmenes de datos y para organizarlos. Además, destacaron la funcionalidad de las herramientas de análisis cualitativo integradas en la aplicación a partir del desarrollo técnico.

9. Discusión

Se ha determinado cuan bien el método *CodCual* es capaz de codificar, identificar y recuperar (mostrar) de manera precisa como eficiente las citas de texto relevantes. El procedimiento realizado fue el siguiente:

1) Descarga e instalación de la aplicación *CodCual*.
2) Uso de la aplicación con 6 documentos de entrevista semiestructurada.
3) Durante el proceso de uso, se utilizó el manual de usuario.
4) Después de una hora, los usuarios participantes respondieron las preguntas para evaluar la herramienta.

9.1 Método de Comprobación del Resultado de la Codificación

Se basó en la capacidad para identificar y recuperar citas de texto codificadas en buena forma. Mediante este enfoque se permitió saber la eficacia de *CodCual* en términos de su capacidad para identificar correctamente las citas con su código asignado.

1. ***Tipo de textos analizados***: Se codificaron fragmentos de texto provenientes de entrevistas semiestructuradas sobre temas sociales extraídos de investigaciones cualitativas. Estas citas de texto resultantes codificadas incluían respuestas a preguntas abiertas, comentarios y observaciones de los participantes.
2. ***Proceso de validación***: Este proceso incluyó dos etapas principales. La primera consistió en la codificación inicial por los usuarios, quienes utilizaron el manual de usuario como guía durante las pruebas. La segunda etapa implicó la evaluación de la aplicación de *CodCual* mediante un cuestionario de 17 preguntas sobre la funcionalidad y usabilidad de ésta. Los estudiantes realizaron la prueba y ellos validaron los resultados, esto significa que los usuarios corroboraron que la herramienta haya realizado correctamente el proceso de codificación.
3. ***Número de pruebas***: Se realizaron un total de 145 pruebas de forma síncrona entre los seis grupos de usuarios descritas anteriormente. Cada grupo fue instruido para realizar la codificación del documento de entrevista, siguiendo las indicaciones de la prueba. Las pruebas se llevaron a cabo en un entorno educativo en un laboratorio de cómputo, lo que permitió saber la experiencia en tiempo real de los usuarios en función con el uso de la aplicación.

9.2 Métricas de funcionalidad y usabilidad.

Como parte de la validación del desarrollo, se realizaron pruebas de funcionalidad y usabilidad con el objetivo de evaluar si *CodCual* cumple adecuadamente con su propósito: facilitar el análisis cualitativo de entrevistas en español de forma accesible, eficiente y sin requerimientos técnicos avanzados. Estas pruebas permiten comprobar si la herramienta no solo opera con corrección, sino que en realidad aporta valor práctico a los investigadores en ciencias sociales.

Las métricas de funcionalidad evaluaron la capacidad del sistema para ejecutar sus principales funciones sin fallos y garantizar la estabilidad operativa, como lo muestra la Figura 14, donde la pregunta: ***¿Los módulos principales del sistema funcionan?,*** obteniendo un porcentaje de 75.2% como funcionalidad total y 16.6% como mayoría, luego una de las preguntas sobre usabilidad es ***¿La terminología usada en las descripciones del software es adecuada para el usuario? Evitando el uso de términos técnicos de computación*** de la Figura 15 obteniendo 57.9% como adecuada y 28.3% como mayoría. Estas métricas tipo preguntas del cuestionario están diseñadas acorde al modelo Meca (Cuevas, 2003). Este modelo está constituido por una estructura arbórea donde las características o propiedades del producto de software que se encuentran en el primer nivel, luego las sub características como la Correctitud y Completitud que se encuentran en el segundo nivel y finalmente los atributos y métricas en el tercer nivel definidos a través de preguntas claras y precisas que el usuario sin conocimientos técnicos de computación puede entender. Este modelo para medir la calidad del software detalla sus antecedentes en (Cortés, 2003). En este artículo solo se presentan 2 preguntas, pero el cuestionario cuenta con un total de 17 preguntas, al final se tomó como resultado la cantidad de preguntas contestadas con el mayor porcentaje y el siguiente mayor (en color verde + el color amarillo o rojo). La encuesta se encuentra en esta dirección: *https://perma.cc/DFE6-QCUR.*

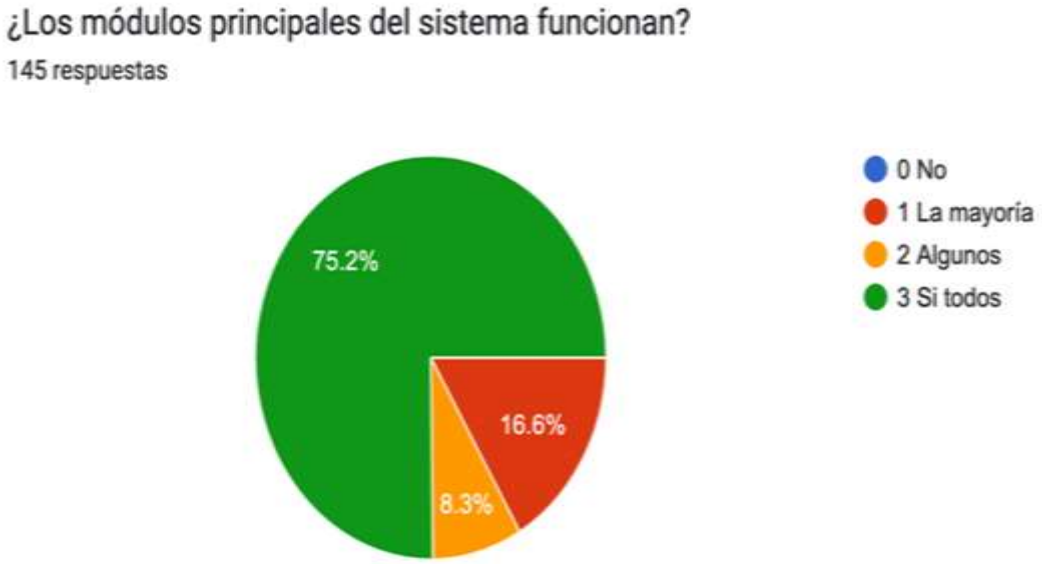


Figura 14. Gráfica de los resultados sobre funcionalidad.

Las métricas de usabilidad se enfocaron en la facilidad con la que los usuarios podían aprender y operar el software, así como en la claridad de la interfaz y la experiencia en general, como lo muestra la Figura 15.



Figura 15. Gráfica de los resultados sobre usabilidad.

Los resultados de ambas métricas permitieron valorar la utilidad real de la herramienta en escenarios típicos de codificación cualitativa, confirmando su efectividad, accesibilidad y pertinencia para el análisis textual, obteniendo finalmente buenos resultados.

9.3 Áreas de aplicabilidad

A partir de la funcionalidad y usabilidad comprobadas, el software muestra capacidad para aplicabilidad en diferentes áreas. En primer lugar, en el área de estudios sociales, ya que, como lo han venido haciendo los softwares comerciales, en sus principales ventajas se incluye una mejor administración o manejo de grandes cantidades de datos cualitativos, es decir texto, en términos de tiempo y recursos, aspecto que se requiere para proceder al trabajo interpretativo y de elaboración conceptual basada en los datos (Gibbs, 2012).

Dentro del área de estudios sociales en colaboración interdisciplinaria con el área de la salud, se vislumbra también su aplicabilidad, pues específicamente se ha desarrollado ampliamente un campo de estudios cualitativos en el ámbito de la enfermería, en torno a los procesos de atención, experiencias de la enfermedad, entre otros (Varela & Sutton, 2021, Naranjo-Hernández & González-Bernal (2021).

Destaca la aplicabilidad en el ámbito educativo de las ciencias sociales, dado que, al ser una herramienta de acceso libre, facilitará la disponibilidad a estudiantes y profesores de niveles medios y/o superiores que no disponen de licencias comerciales de este tipo de software, o bien son limitadas.

Por otro lado, en el ámbito industrial, las aplicaciones potenciales son en el área del marketing. Desde las últimas décadas del siglo pasado, este campo profesional ha recurrido a la investigación cualitativa para fortalecer sus estrategias publicitarias y mercantiles, a partir de explorar los aspectos subjetivos significativos de los consumidores. En este sentido, el software es

una herramienta disponible más para dicho campo (Moreno, 2005).

Recalcando su utilidad en el ámbito académico y de las ciencias sociales, el método CodCual presenta un área de oportunidad en las aplicaciones industriales y productivas. Por ejemplo, en la gestión de calidad, la herramienta puede analizar comentarios de clientes, encuestas abiertas, semiestructuradas o reportes de incidencias, permitiendo identificar patrones de consumo, necesidades recurrentes o áreas de mejora en los procesos. De igual forma, puede aplicarse al análisis de bitácoras de mantenimiento o documentación técnica, facilitando la detección de problemas frecuentes y la toma de decisiones más rápida. Por lo tanto, esto demuestra que, aunque fue concebido inicialmente para la investigación cualitativa en ciencias sociales, CodCual puede transferirse y adaptarse a entornos industriales, aportando valor a la productividad y eficiencia de organizaciones que requieren sistematizar información textual para mejorar sus procesos.

Aunque otras herramientas como NVivo, MAXQDA y ATLAS.ti son potentes, mayormente usadas y cuentan con las características mostradas en la Tabla 1, que pudiesen ser limitantes de CodCual, sin embargo, debido a su apretura cualquier sector académico y social se verá beneficiado con su potencial.

10. Conclusiones y trabajos futuros

En el presente artículo se cumplió con el objetivo al presentar el método de codificación cualitativa, *CodCual*, aplicado a textos en español que identifica palabras y citas con su código en las entrevistas, perfilándose este método como una herramienta que contribuye al análisis de datos cualitativos, lo que aporta al área de las ciencias sociales pero que se puede expandir a otras áreas como por ejemplo educación y salud. Así, los investigadores podrán analizar y organizar sus datos de manera más eficiente. Sin embargo, su verdadero valor radica en cómo los investigadores utilicen *CodCual* para mejorar la calidad y la profundidad de sus análisis. Además, esta herramienta aborda de manera efectiva los desafíos tradicionales del análisis manual, tales como la gran cantidad de tiempo invertido y la necesidad para discernir ideas concretas dentro del texto.

CodCual se evalúo con la participaron 145 estudiantes con distintos niveles de comprensión de textos. La evaluación consistió en la codificación de oraciones de seis entrevistas textuales sobre actividades cotidianas durante la pandemia de COVID-19. Para la codificación, se utilizó un sistema de resaltado por colores lo que facilitó la visualización y recuperación de citas con códigos idénticos. Los resultados revelaron que CodCual es una herramienta eficaz para usuarios con diversos niveles de experiencia en el análisis cualitativo. Además, se observaron patrones recurrentes en las entrevistas, como la adaptación y el compromiso en condiciones adversas, la expresión de ansiedad o estrés en contextos sociales complejos, la transformación de las rutinas cotidianas derivada del confinamiento por la pandemia, así como la influencia sociocultural del turismo extranjero en comunidades rurales. Estos hallazgos evidencian el potencial del método *CodCual* al descubrir patrones y tendencias que podrían pasar desapercibidos con métodos tradicionales.

Con respecto a la calidad del software, CodCual obtuvo puntuaciones de 75.2% en funcionalidad y 57.9% en usabilidad. Aunque se identifican áreas de mejora en la experiencia del usuario, los resultados evidencian el potencial de CodCual como una herramienta de apoyo no solo para investigadores, sino también para estudiantes

y docentes que buscan analizar, sistematizar y organizar el contenido de documentos mediante la codificación de texto.

Como trabajos futuros, se propone incorporar IA para aprender de las codificaciones previas y sugerir mejoras automáticas en la codificación futura, permitiendo una mejor experiencia para el usuario y una mayor eficacia de la aplicación. Así mismo implementar un sistema de generación automática de informes detallados que permita la exportación de resultados en diferentes formatos y la posibilidad de compartir los hallazgos de manera más dinámica. De esta manera también se propone adaptar a *CodCual* para que pueda analizar no solo texto, sino también otros tipos de datos cualitativos como audio, video o imágenes incluyendo el uso de técnicas PLN para una mejor comprensión del contexto y la semántica de los textos analizados.

11. Referencias

[1] Babativa Salamanca, H. A., Rubiano Cárdenas, P. A., Velásquez Pérez, T., Gaona Diaz, N. F., González Ortega, J. A., & Vega Gámez, M. I. (2024). La entrevista semiestructurada: una herramienta pertinente en la percepción de valores sociales para la vida. *Revista Lasallista de Investigación*, 21(1), 92–107. https://doi.org/10.22507/rli.v21n1a5 *(Consultado el 6 de septiembre de 2025)*

[2] Lopezosa, C. (2020). Entrevistas semiestructuradas con NVivo: pasos para un análisis cualitativo eficaz. En C. Lopezosa, J. Díaz-Noci, & L. Codina (Eds.), *Methodos. Anuario de Métodos de Investigación en Comunicación Social*, 1, 88–97. https://doi.org/10.31009/methodos.2020.i01.08 *(Consultado el 6 de septiembre de 2025)*

[3] Rädiker, S., & Kuckartz, U. (2020). *Análisis de datos cualitativos con MAXQDA: Texto, audio, video* (Trad. al español). MAXQDA Press. https://doi.org/10.36192/978-3-948768003 *(Consultado el 6 de septiembre de 2025)*

[4] Pérez-Carballo, J., & Strzalkowski, V. (2000). Natural language information retrieval: Progress report. *Information Processing & Management, 36*(1), 155–178. https://doi.org/10.1016/S0306-4573(99)00049-7 *(Consultado el 6 de septiembre de 2025)*

[5] Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. ACM Press. https://web.cs.ucla.edu/~miodrag/cs259-security/baeza-yates99modern.pdf *(Consultado el 6 de septiembre de 2025)*

[6] Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for information retrieval. *Communications of the ACM, 18*(11), 613–620. https://dl.acm.org/doi/10.1145/361219.361220 *(Consultado el 6 de septiembre de 2025)*

[7] Cházaro-Arellano, E. (2024). Análisis de datos en las investigaciones cualitativas: El reto frente al investigador. *Revista Arbitrada Interdisciplinaria Koinonía*, 9(17), 168-171. https://doi.org/10.35381/r.k.v8i17.3163 *(Consultado el 6 de septiembre de 2025)*

[8] Gibbs, G. (2012). *El análisis de datos en investigación cualitativa* Ediciones Morata ISBN:9788471126856, 8471126850.

[9] Ibáñez de Opacua Lomoschitz, M. (2019). *Named Entity Recognition y Topic Modeling: Metodología y aplicaciones al procesamiento de texto* [Tesis de grado en Ingeniería, Universidad Carlos III de Madrid]. e-Archivo UC3M. https://e-archivo.uc3m.es/handle/10016/30068 *(Consultado el 6 de septiembre de 2025)*

[10] IBM. (2024, 11 de agosto). *¿Qué es el procesamiento del lenguaje natural (PLN)?* IBM. https://www.ibm.com/es-es/topics/natural-language-processing *(Consultado el 6 de septiembre de 2025)*

[11] Nader, K., & Turley, F. (2019). *Los fundamentos de Agile Scrum*. Van Haren Publishing, ISBN:9789401805407, 9401805407

[12] Goyanes, M., & Lopezosa, C. (2024). ChatGPT en Ciencias Sociales: Revisión de la literatura sobre el uso de inteligencia artificial (IA) de OpenAI en investigación cualitativa y cuantitativa. *Anuario ThinkEPI*, 18, e18e04. https://doi.org/10.3145/thinkepi.2024.e18a04 *(Consultado el 6 de septiembre de 2025)*

[13] Rueda Sánchez, M. P., Armas, W. J., & Sigala-Paparella, S.-P. (2023). Análisis cualitativo por categorías a priori: reducción de datos para estudios gerenciales. *Ciencia y Sociedad*, 48(2), 83–96. https://doi.org/10.22206/cys.2023.v48i2.pp83-96 *(Consultado el 6 de septiembre de 2025)*

[14] Lopezosa, C., & Codina, L. (2023). *ChatGPT y software CAQDAS para el análisis cualitativo de entrevistas: Pasos para combinar la inteligencia artificial de OpenAI con ATLAS.ti, NVivo y MAXQDA*. Serie Editorial DigiDoc. DigiDoc Reports. Universitat Pompeu Fabra. http://hdl.handle.net/10230/55477 *(Consultado el 6 de septiembre de 2025)*

[15] Marjaei, S., Yazdi, F. A., & Chandrashekara, M. (2019). MAXQDA and its application to LIS research. *Library Philosophy and Practice*. https://perma.cc/TTV3-5SB4 *(Consultado el 6 de septiembre de 2025)*

[16] Consoli, S. (2021). Descubriendo la cara oculta del análisis narrativo: Una perspectiva reflexiva a través de MAXQDA. *System, 102*, 102611. https://doi.org/10.1016/j.system.2021.102611 *(Consultado el 6 de septiembre de 2025)*

[17] Cisneros Puebla, C. A. (2003). Análisis cualitativo asistido por computadora. *Sociologias*, 5(9), 288-313. https://doi.org/10.1590/s1517-45222003000100010 *(Consultado el 6 de septiembre de 2025)*

[18] Cisneros Puebla, C. A. (2011). La computación cualitativa en la investigación en ciencias sociales. *La Sociología en sus escenarios*, (23), 1-27. https://perma.cc/52AC-DNTJ *(Consultado el 6 de septiembre de 2025)*

[19] Cisneros Puebla, C. A. & Davidson, J. (2012). Qualitative computing and qualitative research: Addressing the challenges of technology and globalization. *Historical Social Research/Historische Sozialforschung*, 237-248. https://perma.cc/H4FM-W6D2 *(Consultado el 6 de septiembre de 2025)*

[20] Makkar, M., & Powell, A. (2024). Qualitative data software for coding. In *Handbook of Qualitative Research Methods in Marketing* (pp. 448-459). Edward Elgar Publishing. https://doi.org/10.4337/9781035302727 *(Consultado el 6 de septiembre de 2025)*

[21] Mativi, C. S., Niveiros, S. I., & Carneiro Junior, J. B. A. (2020). Softwares de análisis de datos cualitativos (QDA): Un análisis de artículos publicados en las bases de datos SPELL y SciELO. *Estudios y Pesquisas em Administração*, 4(2), 151–177. https://doi.org/10.30781/repad.v4i2.10318 *(Consultado el 6 de septiembre de 2025)*

[22] Lopezosa, C., Codina, L., & Freixa, P. (2022). ATLAS.ti para entrevistas semiestructuradas: guía de uso para un

análisis cualitativo eficaz. *Hipertext.net*, 25, 21–33. https://doi.org/10.31009/hipertext.net.2022.i25.03 *(Consultado el 6 de septiembre de 2025)*

[23] Sun, Y., & Peng, S. (2023). Análisis cuantitativo y cualitativo en la investigación de lenguas extranjeras: Reflexiones y perspectivas a partir del uso de SPSS y Atlas.ti. *TEISEL. Revista de Investigación en Didáctica de las Lenguas Extranjeras*, 2, e-a_inv01, 1–24. https://doi.org/10.1344/teisel.v2.41969 *(Consultado el 6 de septiembre de 2025)*

[24] Lopezosa, C., Goyanes, M., & Codina, L. (2024). *Acelerando la investigación cualitativa con inteligencia artificial: Una guía práctica para el diseño, desarrollo y ejecución de investigación con entrevistas*. Universitat de Barcelona. https://hdl.handle.net/2445/211948 *(Consultado el 6 de septiembre de 2025)*

[25] Rodríguez Garay, C. E., & Slimovich, A. (2025). Revisión sistemática de programas informáticos de análisis cualitativo para la sistematización de documentos pre-digitales. *Revista Latinoamericana de Metodología de las Ciencias Sociales (Relmecs)*, *15*(1), e154. https://doi.org/10.24215/18537863e154 *(Consultado el 6 de septiembre de 2025)*

[26] Durán, R. J., Sandoval, V. M., Hernández, S. A., Veliz, A. E., & Alarcón, L. (2025). Análisis cualitativo del consumo informativo y las actitudes políticas de jóvenes en relación con la democracia y la ciudadanía en el sur de Bolivia. *Contratexto*, 43, 57–80. https://doi.org/10.26439/contratexto2025.n43.7437 *(Consultado el 6 de septiembre de 2025)*

[27] Cuevas Rasgado, A. D. (2003). *La evaluación de los productos de software asistida por el sistema MECA* [Tesis de maestría, Centro de Investigación en Computación, Instituto Politécnico Nacional]. Repositorio Digital IPN. https://repositoriodigital.ipn.mx/handle/123456789/6894 *(Consultado el 6 de septiembre de 2025)*

[28] Cortés, F. (2003). *Actualización y automatización de un modelo cualimétrico para la evaluación de calidad del software* [Tesis de maestría, Instituto Politécnico Nacional]. Centro de Investigación en Computación, IPN.

[29] Varela, T. V., & Sutton, L. H. (2021). La codificación y categorización en la teoría fundamentada, un método para el análisis de los datos cualitativos. *Investigación en educación médica, 10*(40), 97-104. https://doi.org/10.22201/fm.20075057e.2021.40.21367 *(Consultado el 6 de septiembre de 2025)*

[30] Naranjo-Hernández, Y., & González-Bernal, R. (2021). Investigación cualitativa, un instrumento para el desarrollo de la ciencia de Enfermería. *Revista Archivo Médico de Camagüey, 25*(3). https://perma.cc/UM35-ETJW *(Consultado el 6 de septiembre de 2025)*

[31] Moreno, M. (2005). La investigación cualitativa en marketing: el camino hacia la percepción social del mercado. *Revista Colombiana de Marketing 4*(6), pp. 68-85. https://perma.cc/LGR9-WRTB *(Consultado el 6 de septiembre de 2025)*