



Ingeniería y competitividad

ISSN: 0123-3033

ISSN: 2027-8284

Facultad de Ingeniería, Universidad del Valle

Peña-Torres, Jefferson A.; Giraldo-Alegria, Santiago;  
Arango-Pastrana, Carlos Alberto; Bucheli, Víctor Andrés  
Un chatbot para asistir a las necesidades de información en tiempos de COVID-19  
Ingeniería y competitividad, vol. 24, núm. 1, e21111004, 2022, Enero-Junio  
Facultad de Ingeniería, Universidad del Valle

DOI: <https://doi.org/10.25100/iyc.24i1.11001>

Disponible en: <https://www.redalyc.org/articulo.oa?id=291371829011>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org  
UAEM

Sistema de Información Científica Redalyc  
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

# A chatbot to support information needs in times of COVID-

## 19

INGENIERÍA DE SISTEMAS

# Un chatbot para asistir a las necesidades de información en tiempos de COVID-19

Jefferson A. Peña-Torres<sup>1</sup> , Santiago Giraldo-Alegría<sup>1</sup> , Carlos Alberto Arango-Pastrana<sup>2</sup> , Víctor Andrés Bucheli<sup>1</sup>

<sup>1</sup>*Universidad del Valle, Escuela de Ingeniería de Sistemas y computación, Cali, Colombia*

<sup>2</sup>*Universidad del Valle, Departamento de Administración y Organizaciones, Cali, Colombia*

<sup>§</sup>*jefferson.amado.pena@correounalvalle.edu.co, giraldo.santiago@correounalvalle.edu.co,  
carlos.arango.pastrana@correounalvalle.edu.co, victor.bucheli@correounalvalle.edu.co*

**Recibido:** 11 de febrero de 2021 – **Aceptado:** 31 de mayo de 2021

## Abstract

This paper aims to describe the design, development, and testing of a question-answering chatbot to inform COVID-19 at Cali town. The chatbot is based on the model of natural language processing, and it is capable of holding a question-and-answer conversation about the pandemic. This document presents the sources of information to solve information needs in the Cali town's risk scenery from March to December 2020; Based on the sources of information, a corpus with 636 sentences was built. Three models were trained bases on the corpus. The models were trained incremental prototyping: A baseline model that responds to general questions, cases, active cases, and deaths at a geographic point of an area or region of interest (BC), the baseline model, zones and news, decrees or regulations generated by the Government during the risk situation (BCN) and the final model that responds the previous items and to frequently asked questions (BFAQ). A satisfaction survey of the prototype was developed to evaluate the chatbot, and the models were evaluated by metrics of PLN precision, coverage, and F-measure. The analysis and results showed that the final model (BFAQS) showed values upper 88% in the three measurements., besides, the BFACS held 1480 conversations, with an average conversation engagement of 4.12 minutes. Furthermore, the survey results show that 87% would use the chatbot again to obtain information about COVID-19.

**Keywords:** *chatbot, COVID-19, natural language processing, virtual assistant.*

Como citar:

Peña-Torres JA, Giraldo-Alegría S, Arango-Pastrana CA, Bucheli VA. Un chatbot para asistir a las necesidades de información en tiempos de COVID-19. INGENIERÍA Y COMPETITIVIDAD. 2022;24(1):e211e11004. <https://doi.org/10.25100/iyc.v24i1.11004>



Este trabajo está licenciado bajo una Licencia Internacional Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0

## Resumen

Este artículo describe el diseño, desarrollo y pruebas de un chatbot pregunta-respuesta acerca del escenario de riesgo covid-19 en la ciudad de Cali. El chatbot se implementa bajo un modelo de procesamiento de lenguaje natural (PLN) y es capaz de sostener un diálogo que brinda información en un escenario de pandemia. En el documento se presenta la identificación de fuentes de datos para solventar las necesidades de información para la ciudad de Cali en el escenario de pandemia, para el periodo marzo diciembre de 2020; con base en las fuentes de información se construye un corpus de información con 636 oraciones. A partir del corpus construido se construyen tres modelos usando prototipado incremental: modelo de línea base para casos generales (BC) el cual responde a preguntas generales y relacionadas con número de casos en puntos geográficos de una zona en la ciudad de Cali; modelo de línea base, zonas geográficas y normativa (BCN) que adicionalmente ofrece información de noticias, decretos o normativas generadas por el Gobierno durante el escenario de riesgo y un modelo de línea base, zonas geográficas, normativa y preguntas frecuentes (BFAQS) que responde adicionalmente a preguntas frecuentes. Para la evaluación se desarrolló una encuesta de satisfacción de la versión prototipo del chatbot, en el caso de la evaluación de los modelos se utilizaron las métricas usualmente utilizadas en procesamiento de lenguaje natural: precisión, cobertura y medida-F. El análisis y los resultados muestran que el modelo final (BFAQS) muestra valores superiores al 88% en las tres medidas, de igual forma el BFAQS sostuvo 1480 conversaciones, con un promedio de enganche o atracción de 4.12 minutos, igualmente, los resultados de la encuesta de la versión prototipo de la herramienta indicó que la 87% volvería a usar el chatbot para obtener información sobre la COVID-19.

**Palabras clave:** asistente virtual, chatbot, COVID-19, procesamiento de lenguaje natural.

## 1. Introducción

Los chatbots impulsados por inteligencia artificial (IA) o asistentes virtuales durante la pandemia del COVID-19 han ganado un lugar destacado, la verificación y seguimiento de síntomas, la gestión de reclamaciones, la respuesta a preguntas comunes y en general la asistencia médica han sido apoyadas por este tipo de herramientas<sup>(1-5)</sup>. Los chatbots están en auge porque pueden desarrollarse según las necesidades de los usuarios, algunos se comunican a través de redes sociales o con aplicaciones de mensajería instantánea (IM) como Twitter, Messenger, Telegram, Slack, entre otros. Este tipo de herramientas interactúan con los usuarios a través de una conversación fluida en lenguaje natural, donde con técnicas de procesamiento de lenguaje natural (PLN) extraen características propias del lenguaje, entienden los mensajes, y dan una respuesta con base las fuentes de información disponibles. Además, para comunicarse con un chatbot no es necesario la instalación de aplicaciones dedicadas<sup>(6)</sup>.

Actualmente, existen varias herramientas de software que permiten la creación de chatbots

como Google's Dialog Flow<sup>(7)</sup> and IBM's Watson assistant<sup>(8)</sup>, Azure bot<sup>(9)</sup>, Amazon Lex<sup>(10)</sup> de uso privativo u otras de código abierto como Rasa<sup>(11)</sup> y uso limitado como Flow XO<sup>(12)</sup> que utilizan diferentes enfoques y técnicas de PLN. Aunque la mayoría permiten el desarrollo de asistentes, comprender ¿cómo implementar un chatbot para asistir a las necesidades de información bajo un escenario de riesgo de pandemia? Es por esta razón que es necesario proponer soluciones para enfrentar situaciones de riesgo como el de la pandemia de la COVID-19, el estudio se realizó entre los meses de marzo y diciembre del 2020 y el trabajo de investigación se desarrolló en el marco de un proyecto financiado por la convocatoria Mincienciación del Ministerio de Ciencia Tecnología e Innovación

Por lo general, un chatbot responde preguntas generales como por ejemplo ¿Qué es el covid-19?, ¿qué medidas tomar en el caso de tener síntomas de covid-19?, sin embargo, en un escenario de riesgo de pandemia, los requerimientos de información son diferentes, en este sentido los requerimientos de información

son específicos, por ejemplo, las preguntas: ¿cuántos son los casos que hay en una comuna o barrio de la ciudad? ¿Sí es seguro o no visitar un centro comercial? ¿Cuál es la resolución vigente? ¿Si hay toque de queda o no? son preguntas específicas para las que es necesario un conjunto de muestras representativo que permita el entrenamiento de modelos bajo una arquitectura apropiada para dichos requerimientos. Por otro lado, los reportes de casos están en diferentes formatos, las noticias y los reportes contienen información sobre el mismo hecho, así el reto no es sólo en la construcción del modelo de PLN, sino también en la construcción del corpus que integra diferentes fuentes de información; finalmente, los datos cambian y las cifras varían en el tiempo, por lo que los retos de investigación requieren diseñar una arquitectura que soporte un chatbot pregunta-respuesta acerca del escenario de riesgo COVID-19 en la ciudad de Cali.

Así, integrar las diferentes fuentes de datos heterogéneas y con grandes volúmenes de información, es un reto del proyecto, de igual forma diseñar la arquitectura de PLN, para posteriormente modelar y entrenar los modelos de PLN; por último, evaluar los modelos y el prototipo desarrollado. En este artículo se analiza y extiende la funcionalidad de un marco de desarrollo Rasa <sup>(11)</sup> para crear un chatbot prototípico que responda a las necesidades de información en escenarios de pandemia. A partir de la información de casos, ubicación geoespacial de casos, normativas gubernamentales, noticias, preguntas frecuentes, entre otros, se aplican los modelos de PLN o inteligencia artificial conversacional, los modelos de aprendizaje automático tienen la capacidad de comunicarse en lenguaje natural y responder preguntas. Finalmente, con base en este modelo se construye una herramienta prototípica, la cual se evalúa al igual que el modelo propuesto. Los resultados muestran que

el modelo de PLN muestra valores superiores al 88% en las medidas de precisión, cobertura y f-measure, de igual forma el modelo sostuvo 1480 conversaciones, con un promedio de enganche o atracción de 4.12 minutos, igualmente, los resultados de la encuesta de la versión prototípica de la herramienta indicó que la 87% volvería a usar el chatbot para obtener información sobre la COVID-19. Los resultados son prometedores y muestran que el diseño de la arquitectura y el corpus construido permiten desarrollar un chatbot para escenarios de pandemia.

Este artículo consta de cinco secciones distribuidas así: En la sección 1 se presenta la introducción, en la sección 2 se presenta la motivación, en la sección 3 se describe el marco metodológico aplicado en la investigación. Luego en la sección 4 se presentan los resultados más relevantes del estudio, centrados en el desempeño de un chatbot prototípico para asistir diferentes necesidades de información sobre el COVID-19. Finalmente, en la sección 5 se presentan las conclusiones más importantes de la investigación.

## 2. Motivación

El mundo se enfrenta al riesgo del COVID-19 y muchos países, incluido Colombia, experimentaron la desinformación, las noticias falsas, los mitos, los remedios caseros, entre otros. De otra parte, la escasa información sobre el escenario de pandemia provocó angustia, estrés, y más importante tomar decisiones no informadas. Debido al incremento de las necesidades de información las agencias gubernamentales y los proveedores de atención médica desarrollaron boletines y piezas de información gráfica para difundir información relevante, pero debido a la cantidad de consultas y a otras preocupaciones urgentes causadas por la pandemia, estas fuentes no fueron suficientes. Además, algunas respuestas requieren un conocimiento específico sobre la pandemia.

Los asistentes virtuales o chatbots conversacionales están dentro de las herramientas de inteligencia artificial, con el procesamiento de lenguaje natural, con las que se puede desarrollar una solución tecnológica (13,14). Un bot puede brindar información que previamente ha sido creada por un experto o simplemente desde una base de conocimiento o corpus, también puede apoyar el diagnóstico médico al realizar preguntas de un cuestionario previamente construido e incluso puede mitigar, evitar o apoyar la diseminación de información ante una situación de riesgo, es decir, son capaces de suplir algunas necesidades de información sin necesidad de la interacción humano-humano (15). Este tipo de herramientas tiene muchos beneficios como: la disponibilidad, ya que puede ser accedido en cualquier momento y la concurrencia, ya que puede mantener varias conversaciones en simultáneo (5). Cada usuario obtiene la información asociada a su solicitud, lo que minimiza algunos problemas relacionados con la falta de información durante la pandemia del 2020.

Las necesidades de información nunca fueron más urgentes: cómo prevenir el contagio, cómo se propaga el virus, cuáles son los síntomas, cuántos casos hay, entre otras preguntas, pueden obtener respuesta a través de un chat que informa a la comunidad caleña, teniendo casos de desinformación que pueden propagarse incluso más rápido que el virus (16). Otros trabajos de chatbots de COVID-19 muestran que este tipo de herramientas en el mundo han sido una forma de construir soluciones tecnológicas e innovadoras para proporcionar información específica sobre el virus, responder preguntas relacionadas con la pandemia (13,17-22). Los chatbots se han desarrollado para lugares estratégicos como aeropuertos u hospitales, donde incluso asisten a profesionales de la medicina (3,5,23). Esto sugiere que las personas están cada vez más cómodas con este tipo de herramientas.

Este trabajo se enmarca en un proyecto financiado por la convocatoria Mincienciatión del Ministerio de Ciencia Tecnología e Innovación (Colombia). El proyecto se desarrolló entre los meses de marzo a diciembre y propuso diferentes herramientas tecnológicas para afrontar el escenario de pandemia. Este componente se enfocó en comprender ¿cómo implementar un chatbot para asistir a las necesidades de información bajo un escenario de riesgo de pandemia? De esta manera se desarrolló una tecnología capaz de ofrecer información específica y enfrentar así situaciones de riesgo como el de la pandemia de la COVID-19 en la ciudad de Cali para las fechas marzo a diciembre del 2020. En un escenario de riesgo de pandemia, los requerimientos de información son específicos, y seguramente para otros momentos de tiempo las preguntas estarán asociadas a las vacunas, casos reincidentes u otras especificidades del escenario de riesgo, así los retos de investigación requieren diseñar una arquitectura que soporte un chatbot pregunta-respuesta acerca del escenario de riesgo COVID -19 en la ciudad de Cali.

### 3. Metodología

La construcción de una herramienta chatbot en tiempos de la COVID-19, requiere diseñar un sistema de inteligencia artificial (IA) que mantenga una conversación en lenguaje natural (24,25). Lo que se busca con el sistema desarrollado es adaptarse para responder a las necesidades de información (18). Para ello, se hace una revisión de tecnologías existentes, se integran los conjuntos de datos, se construyen los modelos de PLN, se implementa una herramienta prototipo y se evalúan tanto los modelos como la herramienta. De esta forma, a continuación, se presenta una revisión de literatura la cual muestra las tecnologías en inteligencia artificial conversacional o procesamiento de lenguaje natural, la Tabla 1, muestra las características de las tecnologías,

ventajas y desventajas. Para el desarrollo de la herramienta prototipo se selecciona una de estas tecnologías como marco de desarrollo (*Framework*). Acorde a este marco de desarrollo se diseña la arquitectura que se presenta en líneas más adelante. Acorde con lo anterior, se implementó la herramienta conversacional prototipo chatbot, para suplir las necesidades de información acerca la COVID-19 en la ciudad de Cali, en el periodo marzo-diciembre 2020.

Los modelos de PLN o de inteligencia artificial conversacional, se construyen a partir de datos, es decir que la construcción de un modelo de PLN supone un proceso de entrenamiento donde el modelo obtiene los parámetros y patrones que mejor se ajustan a los datos. Con estos patrones aprende a identificar las intenciones del usuario y cómo debe responder ante la solicitud.

De esta forma en la subsección 3.2 se presenta la construcción de los conjuntos de datos para el entrenamiento de los modelos, en las subsección 3.3 se presenta la construcción de los modelos la cual se sigue un método de desarrollo de software llamado el prototipado incremental, el cual a partir de un modelo base permite mejorar y agregar nuevas características y habilidades de respuesta, así el primero modelo es capaz de responder preguntas simples, pero no es capaz de responder preguntas de la normatividades asociadas a la pandemia, pero el último modelo es capaz de responder una generalidad amplia de preguntas. Finalmente, se evalúan los modelos de acuerdo con métricas propias de PLN.

Para la implementación de la herramienta chatbot prototipo se utilizó el lenguaje de programación Python y JavaScript, para la implementación se siguió la arquitectura que se presenta en la Figura 1. La evaluación de la herramienta se llevó a cabo por medio de una encuesta de satisfacción a 32 personas, la encuesta se realizó de forma virtual, previamente el usuario utiliza el chatbot. Posteriormente responde las preguntas relacionadas con la calidad de la herramienta, el enganche de la

conversación con el chatbot, la información que provee sobre el escenario de riesgo de la pandemia y finalmente sobre la usabilidad de dicha herramienta prototipo

### 3.1. Tecnologías con Procesamiento de Lenguaje Natural o Inteligencia Artificial conversacional

Entre las tecnologías cognitivas se destacan aquellas que operan sobre el lenguaje y su entendimiento, con plataformas de procesamiento que incluso pueden simular una conversación humana, como lo son para cada caso: Google con Dialog Flow, Amazon con Lex, IBM con Watson, Rasa, Flow XO, entre otras. Cada tecnología implementa diferentes técnicas y enfoques, y esta es la base para el desarrollo de los sistemas conversacionales, a continuación, se presenta una revisión de las características y el alcance que las tecnologías existentes proveen.

En la Tabla 1 se presentan las tecnologías revisadas y se consideran siete criterios para su comparación, los criterios se describen a continuación:

- **Costo cero:** Para este tipo de herramientas, si es gratuita, la réplica de los resultados y el desarrollo de nuevos proyectos no estaría limitado por el pago de licenciamiento.
- **Instalación:** si el marco de desarrollo es instalable, la instancia de la herramienta puede ser desplegada o configurada de acuerdo con los requerimientos. Además, se pueden incluir herramientas de PLN y personalizar partes del software si se requiere.
- **Programación:** si permite adicionar funcionalidades, es programable no sólo parametrizable, sino que abre la posibilidad de desarrollar acciones personalizadas como el procesamiento de la entrada y conexión e interoperabilidad con otras herramientas, etc.

### Instancia de un bot conversacional

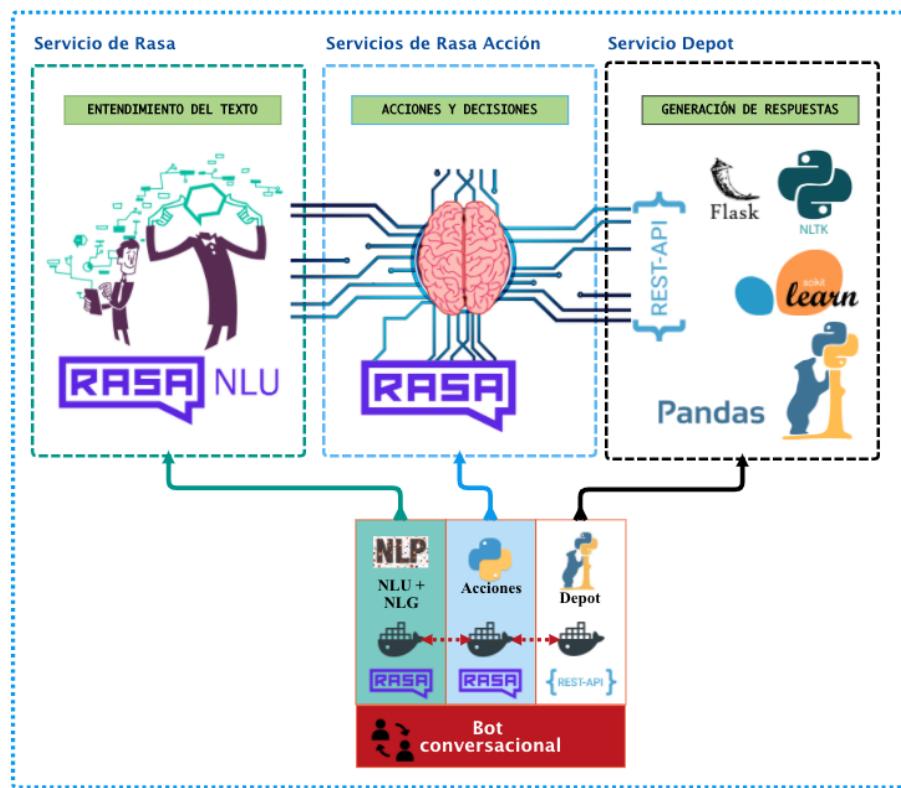


Figura 1. Arquitectura del chatbot prototípico. Fuente: Elaboración propia

Tabla 1. Comparación de tecnologías para desarrollo de Chatbots

Criterios	Rasa	Dialog Flow	Watson	Lex
Costo cero	Si	Si	Si	No
Instalación	Si	No	No	No
Programación	Si, Avanzado	Si, Básico	Si, Básico	Si, Básico
NER	Si	Si	Si	Si
GUI	No	Si	Si	Si
Integración con IM	Ninguno	Slack, Messenger, Telegram, otros	Slack, Messenger, Telegram, otros	Slack, Messenger, Telegram, otros
Cobro por demanda	Gratis	0.002 USD/Msg	0.0025 USD/Msg	0.00075 USD/Msg

Fuente: Elaboración propia

- **NER:** se implementa el Reconocimiento de Entidades Nombradas (NER, del inglés, Named Entity Recognition), esta técnica de PLN permite extraer de un texto automáticamente los nombres propios de lugares, organizaciones, fechas, entre otros, En la tabla se evalúa si la tecnología permite la extracción automática de entidades nombradas a partir de un entrenamiento con oraciones.
- **GUI:** posee una interfaz gráfica de usuario (GUI, del inglés, Graphical User Interface) no se requieren conocimientos avanzados de interfaces de comandos u otros conocimientos para la construcción de asistentes y diálogos.
- **Integración con IM:** se integra con sistemas de mensajería instantánea (IM) está soportado el despliegue a través de un IM como por ejemplo messenger o facebook.
- **Cobro por demanda:** hay costos asociados a la cantidad de recursos que se utilizan o por algunas funcionalidades especiales. La tecnología puede ser gratuita para algunos recursos gratuitos, pero para ciertas funcionalidades se pueden generar costos.

La Tabla 1 muestra que para los criterios costo cero, instalación, programación, GUI las herramientas son muy similares, a excepción de Rasa que permite una programación avanzada y si permite la instalación. Otro criterio que permite tomar una decisión de la tecnología a utilizar para la implementación es el cobro por demanda, pues esto implica que para el desarrollo de nuevos proyectos está limitado por el pago de acuerdo con el número de mensajes.

De esta forma, la comparación de las tecnologías muestra que Rasa tiene ventajas y se adapta más a nuestro contexto en términos de costos por

instalación y cobro por demanda. De igual manera permite la programación y la adición de nuevas funcionalidades, lo cual es clave para un proyecto de investigación o de desarrollo, así para este este trabajo se elige a Rasa <sup>(11)</sup>, la cual es una biblioteca para la comprensión del lenguaje natural que clasifica intenciones, extrae entidades nombradas y otras características del lenguaje natural directamente desde el texto. La tecnología se caracteriza por ser código abierto, gratuita y requiere de instalación local. Aunque, no incluye integración predefinida con sistemas de mensajería instantánea (IM), sistemas como Telegram o Whatsapp al permitir la programación avanzada se puede realizar una conexión de interoperabilidad con dichos sistemas (IM).

La elección de Rasa permite la instalación y la ausencia de interfaz de usuario para su configuración (GUI). Con la instalación se pueden configurar y desplegar microservicios con tareas específicas que puedan responder a múltiples situaciones de riesgo. Con la interfaz de línea de comandos (CLI, del inglés, Command Line Interface) se logra un control granular del sistema operativo y las dependencias de instalación. Lo anterior permite el uso eficiente de los recursos de la máquina y mayor control sobre el desarrollo y las funcionalidades de la herramienta prototipo chatbot.

### 3.2. Entrenamiento y preparación de los conjuntos de datos

Un chatbot puede considerarse como una estrategia para diseminar información correcta para las personas en un escenario de riesgo. Con el marco conversacional de Rasa <sup>(11)</sup> primero, se establece el dominio o contexto del bot conversacional y las intenciones conversacionales, necesidades de información u oraciones que definen el tema de la conversación, estas pueden ser preguntas tipo y

respuestas asociadas para resolver la necesidad de información. De acuerdo con estos elementos de diseño, se recopila un conjunto de datos que contiene una muestra de las intenciones conversacionales y otros textos que acompañan y guían la conversación.

**Dominio de la información:** Una herramienta chatbot debe recibir el lenguaje natural directamente desde el texto de entrada del usuario y entender la demanda de información. En este trabajo, el modelo es de propósito específico o de dominio es cerrado, en especial para responder al escenario de riesgo de pandemia COVID-19. En este sentido, los textos, preguntas y conversaciones tendrán varias intenciones relacionadas con cuatro grupos: las relacionadas con casos, las que involucran normas generadas durante la pandemia, preguntas frecuentes y generales.

Para que el chatbot comprenda el texto es necesario descubrir la intención conversacional que tiene el usuario al enviar el mensaje, esta intención se construye a partir de la información lingüística obtenida del mensaje: la estructura sintáctica, palabras claves y entidades nombradas. Un ejemplo de esta clasificación puede ser:

- **Entrada:** ¿Cuántos casos hay en el barrio Meléndez?
- **Intención:** Cantidad de casos en el barrio Meléndez (*healt\_cases\_neigborhood*)
- **Entidades:** Meléndez

**Conjunto de datos de entrenamiento:** Para desarrollar una herramienta con inteligencia artificial conversacional, se utiliza el marco de desarrollo Rasa<sup>(11)</sup>, es necesario construir un conjunto de datos para entrenar los modelos. Para este trabajo el conjunto de datos contiene diversas muestras de las intenciones del usuario,

o lo que es lo mismo preguntas y textos que un usuario enviaría para consultar información, los cuales el modelo de PLN toma como fuente de conocimiento para entrenarse. De esta manera, intenciones, textos muestra y otros de requerimiento de información particular, se estructuran en un corpus del que se utiliza un 80% para el entrenamiento del modelo y 20% para probarlo. El conjunto de datos contempla 32 muestras por cada intención de usuario, un subconjunto muestral de los textos que se presentan en la Tabla 2.

Los datos se construyeron a partir de la información del municipio de Cali, principalmente información de la vulnerabilidad del contagio en puntos específicos de la ciudad, cantidad de casos por barrio y comuna. También se extrajo datos departamentales y regionales de Esri Colombia<sup>(26-28)</sup> y Datos-Abiertos-Colombia<sup>(29)</sup>, los decretos y normativas fueron obtenidas directamente de los portales gubernamentales, todo de manera automática con técnicas de extracción de información web (Web scraping). Para las noticias se utilizó rutas de distribución de noticias (RSS, del inglés, Really Simple Syndication) directamente de periódicos colombianos y de la ciudad de Cali. Las preguntas y respuestas comunes sobre síntomas, prevención, contagio, vacunas se obtuvieron desde el portal de los Centros para el control y prevención de enfermedades (CDC)<sup>(30, 31)</sup> y coronavirus Colombia<sup>(32)</sup>.

Teniendo en cuenta que los modelos se construyen como prototipado incremental, los conjuntos de datos de entrenamiento siguen la misma lógica, de esta manera, el conjunto de datos es compilado de forma automática para cada una de las intenciones de comunicación asociadas al prototipo incremental: línea base para casos generales (BC); línea base, zonas geográficas y normativa (BCN); y línea base, zonas geográficas, normativa y preguntas frecuentes (BFAQS).

**Tabla 2.** Muestra las intenciones de comunicación por cada grupo

Generales	Relacionadas con casos	Normas generadas durante la pandemia	Preguntas frecuentes
<p><b>agent.sorry:</b> Intenciones de cortesía con el bot</p> <p><b>agent.greetings_bye:</b> Intención de despedirse del bot</p> <p><b>agent.greetings_good_morning:</b> Intenciones de saludar el bot</p> <p><b>agent.knowledge_about:</b> Intenciones de conocer la información del bot</p> <p><b>agent.confirmation_no:</b> Intenciones de negación para el bot</p> <p><b>out_of_scope:</b> Intenciones no clasificadas en otra intención</p>	<p><b>health_crisis_comune:</b> Intenciones relacionadas con la cantidad de casos en una comuna de Cali</p> <p><b>health_crisis_neigborhood:</b> Intenciones relacionadas con la cantidad de casos en algún barrio de Cali</p> <p><b>health_crisis_departamento:</b> Intenciones relacionadas con la cantidad de casos en el departamento del Valle</p> <p><b>health_crisis_country:</b> Intenciones relacionadas con la cantidad de casos en Colombia</p> <p><b>health_crisis_zone:</b> Intenciones relacionadas con la cantidad de casos en algún lugar de la ciudad de Cali</p> <p><b>agent.knowledge_date:</b> Intenciones de conocer la cantidad de contagios en una fecha</p>	<p><b>health_crisis_normative:</b> Intenciones relacionadas con alguna norma presentada por el gobierno.</p> <p><b>health_crisis_news:</b> Intenciones relacionadas con alguna noticia</p>	<p><b>health_crisis_faqs:</b> Intención relacionadas con alguna pregunta frecuente</p> <p><b>health_crisis_myths:</b> Intención relacionadas con algún mito acerca del virus</p> <p><b>health_crisis_precaution:</b> Intención relacionadas con la prevención del virus</p> <p><b>health_crisis_spread:</b> Intención relacionadas con la propagación del virus</p> <p><b>health_crisis_symptoms:</b> Intención relacionadas con los síntomas</p> <p><b>health_crisis_test:</b> Intención relacionadas con las pruebas y las vacunas.</p>

Fuente: Elaboración propia

De esta forma para cada modelo hay su corpus de entrenamiento correspondiente, al cual se agregan preguntas, textos y consultas o requerimientos tipo que contienen la estructura de sintáctica de las preguntas, o requerimientos de información que se esperan del usuario. El corpus de entrenamiento total cuenta con 636 oraciones aproximadamente y 20 consultas por cada intención. Cada modelo fue entrenado siguiendo el enfoque supervisado y el conjunto de muestras se dividía en 80% para entrenamiento y 20% de las muestras para pruebas.

El conjunto de entrenamiento contiene textos con las entidades nombradas (EN), que permiten guiar al Chatbot por un grupo de intenciones relacionadas con la consulta de casos en sitios de interés de la ciudad, como barrios, comunas e incluso fechas. De igual manera los textos para entrenamiento incluyeron palabras claves que son elementos fijos en la solicitud del usuario y guían al Chatbot por las intenciones relacionadas con preguntas frecuentes, mitos, síntomas, decretos, normas y otras diferentes a los casos, ejemplo de ellas son: lavado de manos, dolor en el pecho, diabetes, síntomas, decretos, entre otras. En la Tabla 3, se presentan ejemplos del conjunto de entrenamiento, entre las que se

**Tabla 3.** Textos para entrenamiento y pruebas

Clase	Subclase	Texto
	Comuna	<ul style="list-style-type: none"> <li>- Casos de covid-19 en la [Comuna 13](commune)</li> <li>- Cuántos son los casos en la [Comuna 3](commune)</li> <li>- Casos cerca a la [Comuna 5](commune)</li> <li>- Recuperados en la [Comuna 8](commune)</li> </ul>
Casos		<ul style="list-style-type: none"> <li>- El barrio [Calimio](neighborhood)</li> <li>- El barrio [La Flora](neighborhood)</li> <li>- Número de fallecimientos en el barrio [calimio](neighborhood)</li> <li>- Recuperados en el barrio [Prados De Oriente](neighborhood)</li> </ul>
	Decretos	<ul style="list-style-type: none"> <li>- Cuál es el decreto hay para el transporte de pasajeros</li> <li>- Circular existe para reducir la exposición</li> <li>- Reglamentación para el protocolo de bioseguridad</li> </ul>
Documentos de gobierno	Noticias	<ul style="list-style-type: none"> <li>- Qué hay nuevo sobre coronavirus</li> <li>- Noticias sobre covid o covid19 o covid-19</li> <li>- Últimas noticias del virus</li> <li>- Últimas noticias de la pandemia</li> </ul>
	Comunes	<ul style="list-style-type: none"> <li>- ¿Puedo donar sangre?</li> <li>- ¿Cómo puedo protegerme?</li> </ul>
	Prevención	<ul style="list-style-type: none"> <li>- ¿Debo enjuagarme la nariz con solución salina?</li> <li>- ¿Qué vacunas puedo recibir contra el coronavirus?</li> <li>- ¿Cómo evito el covid covid19 covid-19?</li> </ul>
Preguntas frecuentes	Propagación	<ul style="list-style-type: none"> <li>- Me puedo volver a contagiar</li> <li>- propagación del coronavirus covid o covid19 o covid-19</li> <li>- cómo se propaga el virus enfermedad covid covid19 covid-19</li> </ul>
	Síntomas	<ul style="list-style-type: none"> <li>- Tengo congestión o moqueo, náuseas o vómitos y diarrea</li> <li>- Tengo pérdida reciente del olfato y el gusto, dolor de garganta</li> <li>- Tengo síntomas</li> <li>- ¿Si siento malestar tengo covid?</li> </ul>
	Vacunación y pruebas	<ul style="list-style-type: none"> <li>- Existe algún medicamento contra el covid-19</li> <li>- ¿Debería hacerme una prueba de detección del COVID-19?</li> <li>- ¿Cuándo puedo realizarme la prueba de COVID-19?</li> </ul>

Fuente: Elaboración propia

destacan las entidades que hacen referencia a las comunas y barrios; y las palabras claves, noticias sobre el escenario de pandemia, entre otras.

### 3.3. Arquitectura de referencia para el prototipo

La arquitectura diseñada para la herramienta chatbot prototipo permite la estandarización de una solución para el despliegue de modelos de PLN. La aplicación está basada en microservicios con funciones específicas, autónomos que trabajan juntos y que están acoplados. La arquitectura está basada en microservicios de diseño modular, escalable, flexible y reutilizable <sup>(33)</sup>. Este prototipo puede ser desplegado en cualquier máquina para responder a un escenario de riesgo de pandemia, la cual es portable y es flexible al permitir agregar, modificar o quitar funcionalidades sin dañar la funcionalidad de la herramienta como un todo. La Figura 1 presenta las tecnologías y las relaciones entre los microservicios de la herramienta. La arquitectura se puede describir como el despliegue con tres microservicios:

- El servicio Rasa, este contiene los modelos de PLN y permite previo entrenamiento, entender el texto enviado por el usuario, clasificar automáticamente las intenciones de comunicación y acciones; finalmente enviar las acciones que deben realizarse para generar la respuesta.
- El Servicio de Acciones, diseñado para ejecutar un conjunto de acciones y programas de construcción de la respuesta, la cual se le entregará a cada solicitud del usuario.
- El Servicio Depot, diseñado puede consultar y almacenar la información estructurada o no estructurada desde las diferentes fuentes de datos.

La herramienta prototipo que se describe en este documento se basa en microservicios los cuales están desarrollados con el patrón arquitectural RESTFUL, con una comunicación basada en estructuras JSON. La implementación se hizo con Python v3.5, JavaScript v8.0 y Rasa v2.5. La arquitectura soporta los siguientes requerimientos de información: i) estadística de casos en barrios, comunas, el municipio de Cali y el departamento del Valle del Cauca, Colombia, ii) decretos, circulares, noticias y normativas para Colombia durante el escenario de riesgo de pandemia; y iii) preguntas frecuentes sobre covid-19, su propagación, la vacunación, el escenario de riesgo de pandemia entre otras que son relevantes para la población y su contexto.

El marco de desarrollo en el proyecto fue Rasa <sup>(11)</sup> que incorpora módulos de *Natural Language Understanding* (NLU), una rama de PLN, la cual permite el entendimiento del lenguaje natural. Bajo este marco de trabajo el chatbot aprende a clasificar automáticamente las intenciones, a partir de texto, identificando datos adicionales como entidades nombradas, palabras clave o cualquier otra información que permita dar respuesta a la necesidad de información del usuario. Así, Rasa es utilizado para entrenar los modelos de NLU o PLN, clasificar automáticamente las intenciones comunicacionales de las entradas del usuario, leer los archivos de configuración, las muestras y dar repuestas a las preguntas solicitadas.

La búsqueda de la respuesta se lleva a cabo en el módulo de acciones, las acciones se construyen con un agente basado en Rasa <sup>(11)</sup>, una vez el asistente es capaz de entender lo que el usuario le solicita, debe ejecutar una secuencia de acciones para obtener la respuesta. El mensaje de respuesta debe ser coherente y en lenguaje natural. Es por esta razón que para los casos de barrios, comunas o sitios de interés se crea una

lista de plantillas de respuesta, que especifican el número de casos activos, casos confirmados y fallecimientos en el sitio especificado en la solicitud. Los datos se actualizan a diario y se almacenan de manera estructurada. Para todos los otros casos las respuestas se construyen a partir de los textos recopilados en el conjunto de datos. Esto es todo el corpus de datos considerado en el conjunto de datos que no poseen una estructura y están escritas en lenguaje natural. Así, hallar una respuesta en los textos representa un reto de minería de texto. En este trabajo para la caracterización y extracción de textos se incluyeron técnicas de PLN, que permiten el preprocesamiento del texto, normalización de textos, eliminación de caracteres extraños, saltos y espacios en blanco, tokenizado y lematización de palabras<sup>(34, 35, 36, 37)</sup>. Todas estas tareas se llevan a cabo en Rasa<sup>(11)</sup> y permite la construcción de una respuesta a partir de un texto en lenguaje natural.

Para hallar una respuesta, El Servicio Depot, diseñado puede consultar y almacenar la información estructurada o no estructurada desde las diferentes fuentes de datos, la versión prototipo utiliza las técnicas de Matriz de términos del documento (DTM, del inglés, *Document Term Matrix*) y la similaridad coseno sobre el conjunto de datos permite identificar automáticamente la relación entre el requerimiento de información y la respuesta asociada, para más información sobre el método revisar<sup>(36)</sup>. Lo que significa que la respuesta posee una ponderación numérica en relación con la consulta del usuario, el chatbot posee diversas formas de contestar a una pregunta, pero la más cercana a la consulta será entregada como respuesta al usuario.

### 3.4. Prototipado incremental

Las arquitecturas basadas en microservicios aportan la modularidad necesaria para permitir la reutilización y evolución de prototipos<sup>(33, 38, 39)</sup>.

Tres prototipos fueron puestos a prueba para demostrar la usabilidad de este tipo de soluciones y evidenciar problemas potenciales. La herramienta prototipo que incorpora el modelo línea base (BC) se configuró para suplir las necesidades de información estadística de casos en barrios, comunas, municipios y departamentos, como prueba de concepto se estableció para el contexto de la ciudad de Cali, Colombia. Con un mínimo de 20 oraciones para el entrenamiento y pruebas que incluían Entidades Nombradas (EN) para reconocer los nombres de los barrios y otros lugares específicos de la ciudad, después del entrenamiento este modelo se incorporó al servicio de entendimiento del lenguaje y los datos de los barrios, comuna y de más puntos de interés se almacenaron en el servicio Depot. El diseño incremental utilizado es altamente reutilizable y personalizable para evitar la reconstrucción de los prototipos desde cero. Una vez terminado un modelo, se evalúa y se toma decisiones de mejora, para la evaluación de cada modelo se puso a disposición de los usuarios a través de en un sitio web y de la plataforma Telegram.

Con el prototipado incremental los usuarios finales tienen una participación activa durante la construcción y despliegue de herramientas. Es por esta razón que una vez se avala el comportamiento de una de las versiones de la herramienta prototipo se realizan ajustes, cambios y se incorporan nuevas características para una nueva versión prototipo. Para la herramienta prototipo que responde a noticias, normativas, se mantienen las intenciones con entidades nombradas y se agregan al entrenamiento aquellas que le permiten a Rasa identificar que el usuario requiere nueva información relacionada con los documentos gubernamentales generados durante la pandemia.

Para poder suplir las necesidades de información de una población en su contexto, se deben

considerar diferentes tipos de preguntas acerca del escenario de riesgo de la pandemia, la prevención, el contagio entre otras que son de vital importancia. Para lograrlo se entrenó e incorporó el modelo BFAQ a la herramienta prototipo, que en este artículo presenta un mejor desempeño en comparación con el modelo base, con una respuesta aceptable. Esta versión incluye mensajes de saludo, despedidas con el objetivo de guiar y simular una conversación humana, presenta algunas preguntas iniciales para ayudar al usuario que no sabe cómo iniciar la conversación. El prototipo se conectó a la plataforma de mensajería instantánea Telegram con el nombre de @SigeloBotCali y está actualmente disponible para su utilización.

### 3.5. Evaluación de los prototipos

Para conocer la capacidad de generalización de un modelo, es decir, la capacidad de hacer bien el trabajo de clasificación ante una consulta nueva es importante su evaluación. Para realizar esta evaluación es necesario contar con dos conjuntos de muestras: entrenamiento y pruebas.

El conjunto de entrenamiento se utiliza para que el modelo aprenda los parámetros automáticamente y minimice el error de clasificación. Luego con el conjunto de prueba se calcula el error que comete el modelo sobre las muestras de ese conjunto. Esta división permite conocer de manera fiable la capacidad de generalización. Es común para este tipo de herramientas utilizar las medidas de precisión, cobertura (recall) y valor-F (F-Score). Con la precisión se mide la cantidad de textos clasificados en la intención a la que realmente pertenecen, con la cobertura se mide la cantidad de textos que se han clasificado en la intención correcta y el valor-F es la media armónica entre precisión y cobertura. Cuanto mayor sea el resultado del valor-F, mejor se comporta el modelo<sup>(41)</sup>.

Aunque las medidas son comúnmente usadas para evaluar el desempeño, para este tipo de herramientas que incorporan chatbots es necesario evaluar el contenido, la satisfacción del usuario y otros aspectos funcionales<sup>(42)</sup>. Es por esta razón que en este trabajo se evalúa la herramienta prototipo con una encuesta de seis preguntas: ¿Con qué frecuencia utiliza asistentes virtuales para obtener información? ¿Cómo ha sido la experiencia con el asistente virtual? ¿Utilizaría nuevamente esta herramienta para obtener información bajo situaciones de riesgo (sanitarias, naturales, entre otras)? ¿Qué fue lo más te llamó la atención? ¿En qué grupo enmarcaría las preguntas que le realizó al bot? y ¿Cómo calificaría la respuesta que entregó la herramienta?

Con estas preguntas se espera que los encuestados den a conocer su opinión sobre el prototipo y entender sus necesidades de información.

## 4. Resultados

Los resultados muestran que los modelos de NLU o PLN responden de forma acertada a los requerimientos de información para el escenario de riesgo de pandemia en la ciudad de Cali. Para la evaluación de los modelos propuestos y de las herramientas prototipo se siguieron dos estrategias, los modelos se evaluaron a través de medidas estándar en PLN, las medias utilizadas son precisión, cobertura y f-measurement, mientras que para la evaluación de la herramienta prototipo chatbot, la cual integra los modelos propuestos, se utilizó una encuesta de satisfacción. En esta sección se presenta la evaluación del rendimiento de los modelos incorporados en la herramienta, así como también, un estudio que muestra cómo interactúan la herramienta y los usuarios, de igual forma se observó con el estudio el grado de satisfacción de los usuarios.

#### 4.1. Efectividad de los modelos de PLN

Entre los elementos más importantes de la investigación es la comprensión de los componentes necesarios de NLU y PLN para responder a los requerimientos e información en un escenario de pandemia. Los cuales se encargan de clasificar las intenciones de comunicación del usuario directamente desde el texto. Además, permite encontrar la respuesta más cercana a un requerimiento de información. La compresión del lenguaje natural es considerada como un problema de clasificación múltiple. Para evaluar el desempeño de los modelos en la tarea de clasificación se cuenta con las métricas de precisión, cobertura (recall) y Valor-F (F-Score, por su nombre en inglés). La precisión permite medir el grado de efectividad del modelo, la cobertura o exhaustividad la cual mide la cantidad de ejemplos que el modelo es capaz de identificar correctamente frente al total de ejemplos correctos y el valor-F que es una medida representativa de un bajo número falsos positivos y falsos negativos, lo que significa que el modelo realiza la tarea de clasificación con pocos errores. En la Tabla 4 se presenta el rendimiento de cada uno de los modelos considerados en la investigación, en esta se resalta el rendimiento del modelo BFAQ que en comparación con los otros modelos de referencia logra una mejor valor-F. Es por esta razón que la herramienta prototipo utiliza el modelo BFAQ para responder a las necesidades de información de la ciudad de Cali.

**Tabla 4.** Evaluación de modelos

Modelo	Cobertura (Recall)	Precision (Precision)	Valor-F (F-Score)
BC	0.861	0.857	0.854
BCN	0.873	0.880	0.860
BFAQ	0.886	0.890	0.883

Fuente: Elaboración propia

Las métricas permiten comparar los modelos y la efectividad del método de desarrollo prototipado incremental. Aunque es común utilizar estas métricas, la matriz de confusión del clasificador de intenciones ayuda a reconocer cuáles intenciones tienen mayor confianza.

En la Figura 2, se presenta una matriz de correlación de las intenciones de comunicación, las filas representan las intenciones y en las columnas representan la clasificación realizada por el modelo automáticamente. La matriz de confusión del modelo BFAQ contiene colores oscuros, o con un mayor valor en la diagonal principal y con conteo bajo en otras zonas de la matriz. Esto significa que el modelo predice bien cuando las preguntas del usuario están relacionadas con covid-19, la mitigación, la precaución, la propagación, las medidas de prevención, entre otras preguntas comunes. Para preguntas relacionadas con el número de casos el modelo presenta tiende a fallar más, ya que la sintaxis de este tipo de preguntas es similar, preguntar por el número de casos en barrios, departamento o municipios poseen la misma estructura sintáctica y así el modelo presenta más errores en la matriz. La herramienta posee una combinación de tecnologías de NLU y PLN, para lograr la integración de datos de diferentes fuentes, datos públicos y el conocimiento disponible en las páginas web, noticias y documentos gubernamentales. Además, incorpora en su diseño un marco conversacional autónomo que responde de forma acertada a preguntas bajo situaciones de riesgo.

La construcción de los conjuntos de datos, la implementación de las tecnologías y el desarrollo de la tecnología prototipo se hizo en tres meses, lo que muestra que frente a un escenario de riesgo de pandemia la arquitectura acá presentada y las tecnologías acá descritas permitiría crear un chatbot en menor tiempo. La cual puede ser una herramienta de soporte a la toma de decisiones de los ciudadanos, para que

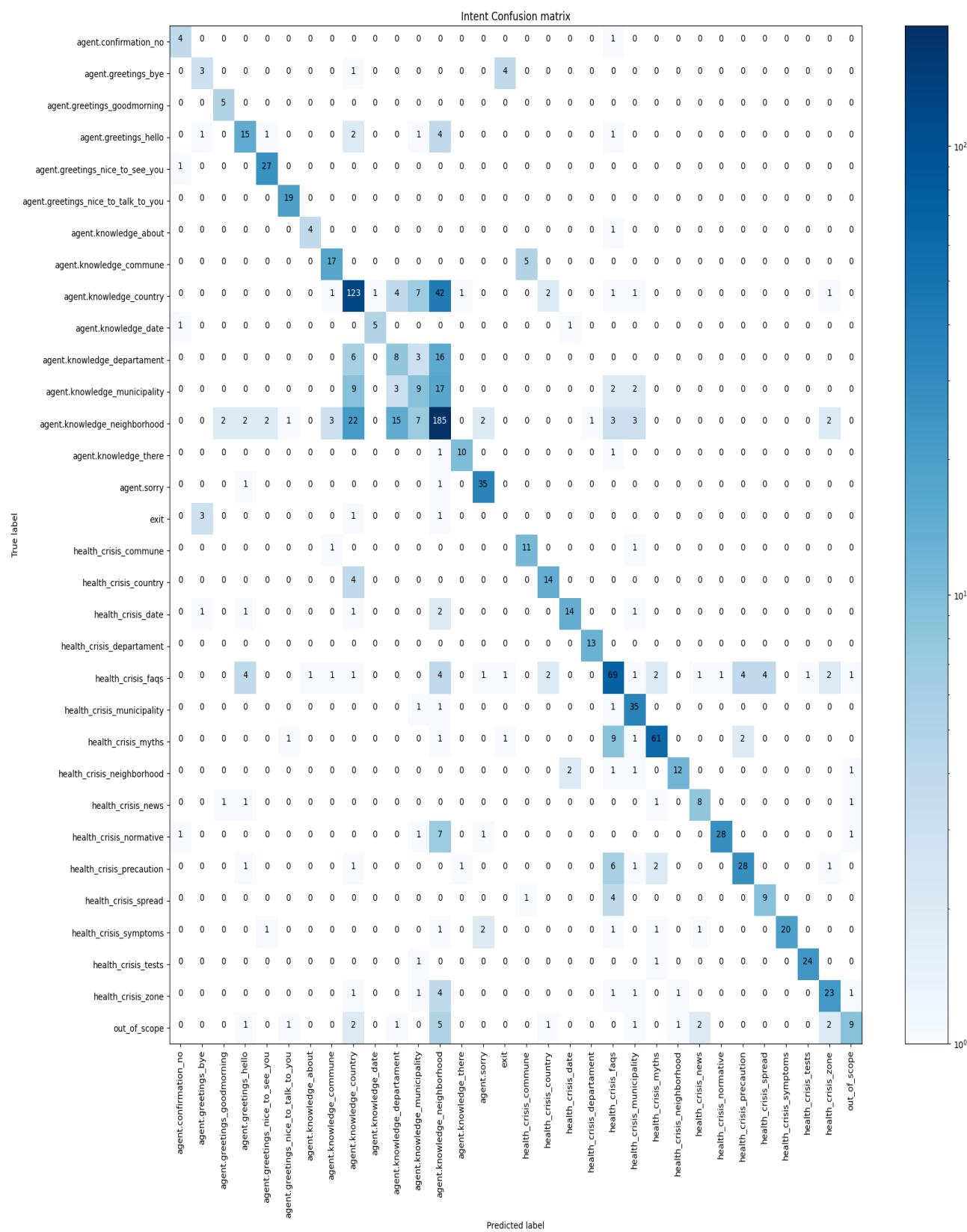


Figura 2. Matriz de confusión para el modelo BFAQ. Fuente: Elaboración propia

ellos se mantengan actualizados, informados y tomen decisiones informadas en cuanto a movilidad, normatividad y demás aspectos relacionados en un escenario de pandemia. Finalmente, dado que la herramienta se conecta con sistemas de información de IM, el flujo de información con los usuarios es transparente, pues ellos pueden comunicarse como con un contacto en facebook, messenger u otro, ver Figura 3

#### 4.2. Encuesta de satisfacción

Como prototipo funcional se evalúa la herramienta chatbot prototipo con una encuesta de 10 preguntas politómicas. Esta evaluación se realizó con 32 participantes, profesionales con diferentes necesidades de información sobre la COVID-19 en Cali, con el fin de conocer su nivel de satisfacción. En Figura 3, se presenta el saludo inicial de la herramienta, la respuesta que ofrece al preguntar por el número de casos en la ciudad y una muestra de las preguntas que más le han realizado y que pueden ser de interés, esta muestra es de la versión que utiliza el sistema de mensajería Telegram

Los resultados de la encuesta evidencian que la frecuencia con la que se usan este tipo de herramientas para la población de encuestados la cual es: el 50% indicó que nunca o casi nunca las han utilizado, el porcentaje restante indicó que usa bots o asistentes para obtener información. A pesar de la poca inmersión de este tipo de herramientas en la población, en la Figura 4 se presenta como fue la experiencia que indicaron los encuestados. Los resultados muestran que el 87.5% de los encuestados indicaron como buena o muy buena, esto sugiere que este tipo de herramientas posee un gran potencial para resolver necesidades de información.

De igual manera en la Figura 5, se preguntó si utilizaría nuevamente esta herramienta para obtener información bajo situaciones de riesgo (sanitarias, naturales, entre otras). El 87.5% de

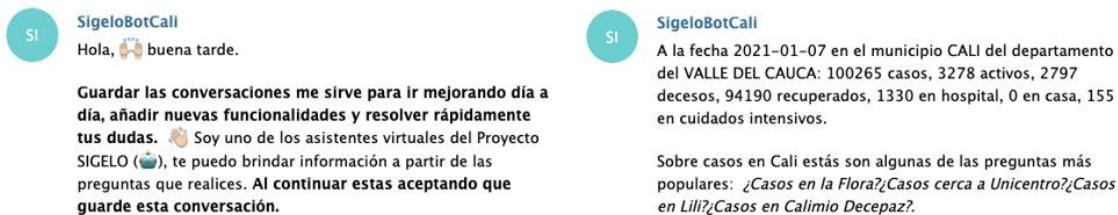
los encuestados indicó que usaría la herramienta. Mientras que el 62.5% de los encuestados indicaron que sí o probablemente sí utilizarían esta herramienta de estas características para obtener información en situaciones de riesgo.

Por otra parte, con el objetivo de evidenciar la pertinencia de la herramienta de acuerdo con la Figura 6, el 75% de los encuestados indicó que el tiempo de respuesta es bueno o muy bueno, el 57% de los encuestados expresó que las respuestas del chatbot eran claras y que en un 70% las respuestas eran útiles o realmente suplían la necesidad de información. Mientras que el 37.5% indicó que la herramienta era fácil de usar y el 31.25% indicó que la herramienta era rápida en responder las preguntas.

Para analizar la pertinencia de los datos y las necesidades de información, de acuerdo a la Figura 7. Los encuestados indicaron que sus necesidades de información, o las preguntas que realizaron se enmarcaron en cantidad de casos y lugares de riesgo en la ciudad de Cali, en menor medida temas de COVID y normativas generadas durante la pandemia. Las necesidades de información acerca de noticias obtuvieron el menor porcentaje, quizás porque ya existen portales o fuentes de noticias que suplen estas necesidades de mejor manera.

#### 5. Conclusiones

Este trabajo presenta el desarrollo de una herramienta prototipo basada en RASA <sup>(11)</sup> para asistir las necesidades de información relacionadas con la pandemia del COVID-19. El diseño y construcción de la herramienta se basó en un modelo que logró de forma efectiva responder preguntas relacionadas con el escenario de riesgo de pandemia. Este trabajo permitió explorar las herramientas existentes para la implementación de un chatbot, seleccionar la más indicada, construir los conjuntos de datos para el caso del escenario de



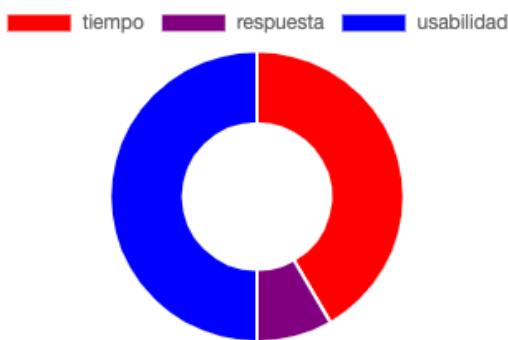
**Figura 3.** Ejemplo de interacción en Telegram. Fuente: Elaboración propia



**Figura 4.** Experiencia con el prototipo, cómo ha sido la experiencia con el asistente virtual del proyecto SIGELO. Fuente: Elaboración propia



**Figura 5.** Uso de la herramienta en escenario de riesgo de pandemia, utilizaría nuevamente esta herramienta para obtener información bajo situaciones de riesgo (sanitarias, naturales, entre otras). Fuente: Elaboración propia



**Figura 6.** Evaluación de la usabilidad, ¿Qué fue lo más te llamó la atención? Fuente: Elaboración propia



**Figura 7.** Interés de los encuestados, De acuerdo con su interés se puede agrupar las preguntas. Fuente: Elaboración propia

riesgo de pandemia de la Ciudad de Cali en el periodo marzo diciembre de 2020. Entrenar y evaluar los modelos de PLN e implementar con base en ellos una herramienta prototípico, la cual también fue evaluada mediante una encuesta de satisfacción. Los requerimientos de información que el chatbot es capaz de entregar es el número de casos a nivel de barrio y comuna en la ciudad de Cali, las noticias, los decretos y toda normativa desarrollada por el Gobierno Nacional para afrontar la pandemia Covid-19.

Además de responder a preguntas comunes acerca de la propagación, la vacunación, el rastreo de contagios, síntomas, trato de las mascotas, niños, limpieza y desinfección, entre otras. Se sigue una metodología de desarrollo de software llamado prototipado incremental, se desarrollaron tres modelos que se entrenaron progresivamente con datos recopilados manualmente de acuerdo con las necesidades de información documental para Cali y la satisfacción de estas necesidades. Los resultados muestran que el modelo final (BFAQS) muestra valores superiores al 88% en las tres medidas, de igual forma el BFAQS sostuvo 1480 conversaciones, con un promedio de enganche o atracción de 4.12 minutos, igualmente, los resultados de la encuesta de la versión prototípico de la herramienta indicó que la 62% volvería a usar el chatbot para obtener información sobre la COVID-19.

Como estudio futuro además de realizar una exploración más amplia de las necesidades de información, según su fuente, su tipología, la relevancia que esta tenga para la población objetivo. Así como intenciones de comunicación que cambian de acuerdo con el momento actual de la pandemia, por ejemplo, vacunación, nuevas olas de contagio, etc. Se puede incluir una exploración con multimedia, la voz y las imágenes también son formas naturales que usan las personas para comunicarse.

Finalmente se concluye que la herramienta desarrollada, permite a las personas encontrar información pertinente, relevante y actualizada que satisfaga las necesidades de información que se pueden presentar en una situación de riesgo. Las preguntas o dudas son resueltas a través de una conversación en lenguaje natural a través de un dispositivo móvil o un portal digital. Con este tipo de desarrollos se impulsa la automatización de flujos y procesos de comunicación en la atención médica y sanitaria que en este momento son de utilidad para apoyar a los ciudadanos a tomar decisiones informadas.

## 6. Agradecimientos y declaración de financiación

Este artículo fue desarrollado gracias a la financiación otorgada por el Ministerio de Ciencias, Tecnología e Innovación de Colombia

– MINCIENCIAS, en el proyecto “Sistema de gestión logística local en escenarios de riesgo de pandemia COVID-19 – SIGELO” con código 1106101577569, proyecto financiado en el marco de la convocatoria MINCIENCIATON, que tenía por objetivo “Fomentar el desarrollo de soluciones para afrontar problemáticas ocasionadas por la pandemia de COVID-19 y otras infecciones respiratorias agudas (IRA) de gran impacto en salud pública, mediante la selección y financiación de proyectos que promuevan la obtención de resultados científicos y tecnológicos en torno al diagnóstico, tratamiento, mitigación y monitoreo de las enfermedades correspondientes”. Los autores también agradecemos a la Universidad del Valle por el soporte técnico, académico y financiero durante la formulación y ejecución del proyecto.

## 7. Referencias

- (1) Bickmore TW, Pfeifer LM, Jack BW. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the SIGCHI conference on human factors in computing systems. 2009. p. 1265-74. <https://doi.org/10.1145/1518701.1518891>
- (2) Jerald J. The VR book: Human-centered design for virtual reality. Morgan & Claypool; 2015. <https://doi.org/10.1145/2792790>
- (3) Amato F, Marrone S, Moscato V, Piantadosi G, Picariello A, Sansone C. Chatbots meet ehealth: Automatizing healthcare. CEUR Workshop on Artificial Intelligence with Application in Health. 2017;1982(6):40-49.
- (4) Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 2018;25(9):1248-58.
- (5) Herriman M, Meer E, Rosin R, Lee V, Washington V, Volpp KG. Asked and answered: Building a chatbot to address covid-19-related concerns. NEJM Catal Innov Care Deliv. In press 2020. <https://doi.org/10.1093/jamia/ocv072>.
- (6) Shevat A. Designing bots: Creating conversational experiences. O'Reilly Media, Inc.; 2017.
- (7) Dialogflow [Internet]. [citado 22 de diciembre de 2020]. Disponible en: <https://dialogflow.cloud.google.com/#/getStarted>.
- (8) Watson Assistant - Overview [Internet]. 2021 [citado 10 de febrero de 2021]. Disponible en: <https://www.ibm.com/cloud/watson-assistant>
- (9) Azure Bot Service|Microsoft Azure [Internet]. [citado 10 de febrero de 2021]. Disponible en: <https://azure.microsoft.com/es-es/services/bot-services/>
- (10) Chatbot | Deep learning | Amazon Lex [Internet]. [citado 10 de febrero de 2021]. Disponible en: <https://aws.amazon.com/es/lex/>
- (11) Bocklisch T, Faulkner J, Pawlowski N, Nichol A. Rasa: Open-source language understanding and dialogue management. ArXiv Prepr ArXiv171205181. 2017. <https://arxiv.org/abs/1712.05181>.
- (12) Flow XO – Easy to use chatbot platform [Internet]. Flow XO. [citado 10 de febrero de 2021]. Disponible en: <https://flowxo.com/>
- (13) Martin A, Nateqi J, Gruarin S, Munsch N, Abdarrahmane I, Zobel M, et al. An artificial intelligence-based first-line defence against COVID-19: digitally

- screening citizens for risks via a chatbot. Sci Rep. 2020;10(1):1-7. <https://doi.org/10.1038/s41598-020-75912-x>
- (14) Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. Npj Digit Med. 2020;3(1):1-4. <https://doi.org/10.1038/s41746-020-0280-0>
- (15) Sears A, Jacko JA. Human-computer interaction fundamentals. CRC press; 2009.
- (16) Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. 2018;359(6380):1146-51. <https://doi.org/10.1126/science.aap9559>
- (17) Ni L, Lu C, Liu N, Liu J. Mandy: Towards a smart primary care chatbot application. In: International symposium on knowledge and systems sciences. Springer. 2017. p. 38-52.
- (18) Srivastava S, Kalra S, Prabhakar T. Contextual Reactive Pattern on Chatbot building Platforms. In Proceedings of the European Conference on Pattern Languages of Programs (EuroPLoP '20). Association for Computing Machinery. 2020. p.1-8. <https://doi.org/10.1145/3424771.3424815>
- (19) Rodsawang C, Thongkliang P, Intawong T, Sonong A, Thitiwatthana Y, Chottanapund S. Designing a competent chatbot to counter the COVID-19 pandemic and empower risk communication in an emergency response system. OSIR Journal. 2020 Jun 30;13(2):71-77.
- (20) Jiao A. An Intelligent Chatbot System Based on Entity Extraction Using RASA NLU and Neural Network. JPhCS. 2020;1487(1):012014. <https://doi.org/10.1088/1742-6596/1487/1/012014>
- (21) Ouerhani N, Maalel A, Ghézela HB. SPeCECA: a smart pervasive chatbot for emergency case assistance based on cloud computing. Clust Comput. 2020;23(4):2471-82. <https://doi.org/10.1007/s10586-019-03020-1>
- (22) Judson TJ, Odisho AY, Young JJ, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. J Am Med Inform Assoc. 2020;27(9):1450-5. <https://doi.org/10.1093/jamia/ocaa130>
- (23) Yan M, Castro P, Cheng P, Ishakian V. Building a chatbot with serverless computing. In: Proceedings of the 1st International Workshop on Mashups of Things and APIs. 2016. p. 1-4. <https://doi.org/10.1145/3007203.3007217>
- (24) Chittò P, Baez M, Daniel F, Benatallah B. Automatic generation of chatbots for conversational web browsing. In: International Conference on Conceptual Modeling. Springer. 2020. p. 239-49. [http://dx.doi.org/10.1007/978-3-030-62522-1\\_17](http://dx.doi.org/10.1007/978-3-030-62522-1_17)
- (25) Baez M, Daniel F, Casati F, Benatallah B. Chatbot integration in few patterns. IEEE Internet Comput. 2020. <https://doi.org/10.1109/MIC.2020.3024605>
- (26) Esri Colombia [Internet]. Esri Colombia. [citado 10 de febrero de 2021]. Disponible en: <https://esri.co/covid-19/>

- (27) Dangermond J. Geodesign and GIS—designing our futures. *Proc Digit Landsc Archit.* 2010;502-14.
- (28) Dangermond J, De Vito C, Pesaresi C. Using GIS in the Time of the COVID-19 Crisis, casting a glance at the future. A joint discussion. *J-Read-J Res Didact Geogr.* 2020;195-205. Available from: <http://www.j-reading.org/index.php/geography/article/view/257>.
- (29) Datos Abiertos Colombia | Datos Abiertos Colombia [Internet]. la plataforma de datos abiertos del gobierno colombiano. [citado 10 de febrero de 2021]. Disponible en: <https://www.datos.gov.co/>
- (30) CDC. Coronavirus Disease 2019 (COVID-19) [Internet]. Centers for Disease Control and Prevention. 2020 [citado 6 de mayo de 2020]. Disponible en: <https://www.cdc.gov/coronavirus/2019-ncov/php/principles-contact-tracing.html>
- (31) CDC. espanol.cdc.gov [Internet]. Centers for Disease Control and Prevention. 2020 [citado 10 de febrero de 2021]. Disponible en: <https://espanol.cdc.gov/coronavirus/2019-ncov/faq.html>
- (32) Coronaviruscolombia [Internet]. [citado 10 de febrero de 2021]. Disponible en: <https://coronaviruscolombia.gov.co/Covid19/>
- (33) Dragoni N, Lanese I, Larsen ST, Mazzara M., Mustafin R., Safina L. (2018) Microservices: How To Make Your Application Scale. In: Petrenko A., Voronkov A. (eds) Perspectives of System Informatics. PSI 2017. Lecture Notes in Computer Science, vol 10742. Springer, Cham. [https://doi.org/10.1007/978-3-319-74313-4\\_8](https://doi.org/10.1007/978-3-319-74313-4_8)
- (34) Uysal AK, Gunal S. The impact of preprocessing on text classification. *Inf Process Manag.* 2014;50(1):104-12.<http://dx.doi.org/10.1016/j.ipm.2013.08.006>
- (35) Bird S, Loper E. NLTK: The Natural Language Toolkit. In: Proceedings of the ACL Interactive Poster and Demonstration Sessions [Internet]. Barcelona, Spain: Association for Computational Linguistics; 2004. p. 214-7. Disponible en: <https://www.aclweb.org/anthology/P04-3031>
- (36) Srinivasa-Desikan B. Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras. Packt Publishing Ltd; 2018.
- (37) Plisson J, Lavrac N, Mladenic D, others. A rule-based approach to word lemmatization. In: Proceedings of IS04. 2004. p. 83-6.
- (38) Thomas N. An e-business chatbot using AIML and LSA. In: 2016 International Conference on Advances in Computing, Communications, and Informatics (ICACCI). IEEE; 2016. p. 2740-2.
- (39) Abdul-Kader SA, Woods J. Question answer system for online feedable newborn Chatbot. In: 2017 Intelligent Systems Conference (IntelliSys). IEEE; 2017. p. 863-9.
- (40) Petrasch R. Model-based engineering for microservice architectures using Enterprise Integration Patterns for inter-service communication. In: 2017 14th International Joint Conference on Computer Science and Software

- Engineering (JCSSE). IEEE; 2017. p. 1-4.
- (41) Tharwat A. Classification assessment methods. *Applied Computing and Informatics*. 2021;17(1):168-192. <https://doi.org/10.1016/j.aci.2018.08.003>
- (42) Maroengsit W, Piyakulpinyo T, Phonyiam K, Pongnumkul S, Chaovalit P, Theeramunkong T. A survey on evaluation methods for chatbots. In: *Proceedings of the 2019 7th International Conference on Information and Education Technology*. 2019. p. 111-9.