

Inteligencia Artificial aplicada a la estadística pública. Chatbot para facilitar el acceso de los ciudadanos a los datos estadísticos

**Jesús Alberto González Yanes - Instituto Canario de Estadística (ISTAC)
Arte Consultores Tecnológicos S.L.**

Resumen:

El Instituto Canario de Estadística (ISTAC) y Arte Consultores Tecnológicos han creado 'ISTAC-Bot', un asistente automatizado que ayuda a los usuarios a localizar datos estadísticos. Está disponible en la aplicación de mensajería instantánea Telegram y es capaz de entender un centenar de indicadores y palabras clave. 'ISTAC-Bot' es uno de los primeros proyectos de inteligencia artificial del Gobierno de Canarias y el primer chatbot estadístico del mundo [1][2].

Para el desarrollo de este chatbot se ha utilizado la herramienta RASA stack, compuesta por un sistema de procesamiento de lenguaje natural (NLU ó *Natural Language Understanding*) acoplado a un sistema de manejo de diálogo. En este artículo se detallan las características de estos dos sistemas y se exponen los resultados obtenidos mostrando diálogos reales entre un usuario que consulta los datos estadísticos y el chatbot.

Palabras clave: IA, NLP, Machine Learning, chatbot, bot, difusión estadística.

Introducción

El Instituto Canario de Estadística (ISTAC) y Arte Consultores Tecnológicos (ARTE) han estado trabajando en mejorar la comunicación entre el público generalista y el ISTAC. Como parte de esa iniciativa se propuso un proyecto piloto que permitiera que los usuarios accedieran de forma mucho más amigable a los datos publicados por el ISTAC. Bajo esta premisa, se crea 'ISTAC-Bot', un asistente automatizado que ayuda a los usuarios a localizar datos estadísticos.

ISTAC-bot pretende ofrecer al usuario una experiencia más interactiva y eficaz, respondiendo a las preguntas, interactuando y mostrando los indicadores estadísticos requeridos por el usuario. La ventaja de este sistema con respecto al motor de búsqueda tradicional es que el usuario tiene una interacción más humana y realista. Además, el chatbot permite al usuario navegar por las bases de datos del ISTAC de una manera más eficaz. Así, una vez que el usuario ha obtenido el dato que estaba buscando, el chatbot le sugiere nuevo contenido, mostrándole indicadores que guardan relación con el indicador por el que preguntó en el primer momento, ampliando de este modo su horizonte de búsqueda y complementando la información.

ISTAC-Bot está disponible en la aplicación de mensajería instantánea Telegram y es capaz de entender un centenar de indicadores y palabras clave [1][2]. Para que los usuarios puedan interactuar con el chatbot, basta con instalar la aplicación Telegram en su móvil y a continuación buscar a 'ISTAC-Bot' y añadirlo como amigo, con esto podrá iniciar una conversación con el asistente virtual y empezar a pedir los datos que necesita.

En su primera fase de desarrollo, 'ISTAC-Bot' facilita datos de determinados indicadores estadísticos para un periodo de tiempo determinado, una isla o municipio canario en concreto y para un determinado sexo. Así, por ejemplo, el asistente virtual es capaz de ofrecer al usuario información sobre tasas de paro o IPC para un año elegido, por islas o comarcas [1][2]. El usuario también puede preguntar, por ejemplo, por el Paro Registrado, y distinguir incluso entre los datos para paro registrado en función del sexo (hombres o mujeres).

Para crear este chatbot se han empleado técnicas de procesamiento de lenguaje natural que se explicarán en este artículo. Además de la metodología, también se expondrán los resultados y las conclusiones.

Metodología

Estructura del Chatbot

Los chatbots son softwares de inteligencia artificial diseñados para interpretar el lenguaje natural y facilitar la información que el usuario solicita de forma automática a través de una conversación simulada [1][2].

Para el desarrollo de ISTAC-Bot se ha hecho uso de la herramienta *Open Source RASA Stack* [3]. RASA consiste en una API de alto nivel para el desarrollo de chatbots en lenguaje Python, que utiliza librerías de procesamiento de lenguaje natural (NLP ó *Natural Language Processing*) y de *Machine Learning* (ML), tales como:

- SpaCy [4] y Mitie [5]: Librerías de procesamiento de lenguaje natural que permiten digitalizar y analizar texto en varias lenguas.
- Scikit-learn [6] y TensorFlow [7]: Librerías que implementan múltiples herramientas de ML y grafos de redes neuronales.

RASA es un buen compromiso entre manejabilidad, libertad y personalización. El chatbot desarrollado con RASA consiste en dos sistemas: una parte de procesamiento de texto, RASA NLU, y un sistema de control de diálogo, RASA core. En la Figura 1 se presenta un esquema de los dos sistemas.

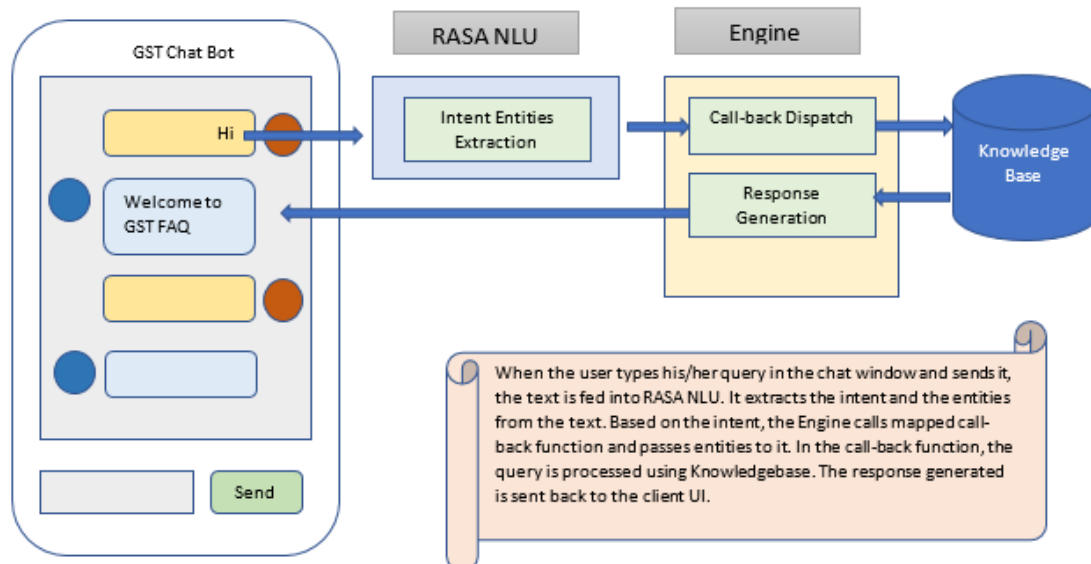


Figura 1: Arquitectura del chatbot. RASA NLU analiza el texto enviado por el usuario y extrae la intención y las entidades. Luego, el sistema de manejo de diálogo decide qué acción tomar, para generar una respuesta que es enviada al usuario. Imagen tomada de [8].

- RASA NLU

Este primer sistema se encarga de analizar el texto ingresado por el usuario con el fin de extraer la **intención** de la frase y las **entidades** o variables importantes para el diálogo. Por ejemplo para la frase: "¡Hola!", la intención es Saludar.

- Gestión de diálogo

En función de la intención y del historial de la conversación, este segundo sistema se encarga de decidir la siguiente acción que el chatbot debe efectuar. En el ejemplo anterior, el bot debe responder al saludo del usuario y eventualmente preguntarle: "¿En qué te puedo ayudar?".

RASA NLU

En esta sección se detallan las características del primer sistema del chatbot desarrollado para el ISTAC: el procesamiento de lenguaje natural. Se emplea la configuración que utilizan las librerías Spacy y Scikit-learn para el análisis de texto y el clasificador de intenciones respectivamente. Las intenciones que se han utilizado para entrenar este clasificador son:

- Saludo: el usuario saluda. Por ejemplo: "¡Buenos días!"
- Pedir información: el usuario pide alguna información contenida en la base de datos del ISTAC. Por ejemplo: "Quiero saber la tasa de desempleo de Canarias."
- Despedida: el usuario se despide. Por ejemplo: "¡Muchas gracias, adiós!"

Se crea un dataset de entrenamiento con múltiples ejemplos de frases en cada una de las clases, y se entrena el clasificador para que pueda detectar la intención del texto ingresado por el usuario. Asimismo, se incluyen en el dataset ejemplos de las entidades que se quieren retener de las frases del usuario.

- Entidad de lugar. Por ejemplo, el nombre de un municipio o isla de Canarias, así como códigos postales.
- Entidad de tiempo. Por ejemplo, un año o mes.
- Entidad de indicador. Correspondiente a los indicadores estadísticos del ISTAC, tales como la tasa de desempleo o la población.

Así, cuando el usuario pide el valor de un indicador, haciendo referencia a un lugar determinado, en un período determinado, el sistema guarda estas tres entidades en memoria para poder acceder a la información correspondiente de la base de datos del ISTAC.

Sistema de diálogo de RASA

El sistema de diálogo del chatbot decide qué acción efectuar en función de lo que el primer sistema detecta. Para esto fue necesario crear un dataset con la estructura típica de las conversaciones posibles entre el usuario y el chatbot. Así, cuando el usuario saluda, el chatbot saluda nuevamente y propone su ayuda. Si el usuario pide una información con las entidades mínimas necesarias, el chatbot responde.

Si en la petición del usuario llegase a faltar una o varias entidades, el chatbot las pide. Además, una vez que la petición está completa, el chatbot propone una nueva búsqueda variando las entidades.

Si el usuario llegase a equivocarse en la ortografía de una de las entidades, el chatbot es capaz de proponer las entidades más parecidas presentes en la base de datos. Se ha añadido también un sistema de sugerencias de posibles indicadores interesantes para el usuario, en función de su búsqueda actual. De esta forma, una vez concluida una búsqueda, el chatbot propone al usuario los indicadores relacionados con el tema de la última búsqueda efectuada.

Resultados

Tal y como se ha mencionado, RASA NLU analiza el texto enviado por el usuario y extrae información sobre las intenciones del usuario y sobre los parámetros que nos proporciona. Posteriormente, el sistema de gestión de diálogos decide qué acción tomar para proceder a generar la respuesta que se enviará al usuario. En este punto, en el caso de que el chatbot considere que tiene toda la información necesaria, acudirá al sistema de indicadores del ISTAC (eDatos) para obtener los datos que debe mostrar al usuario en la respuesta final (Figura 2).

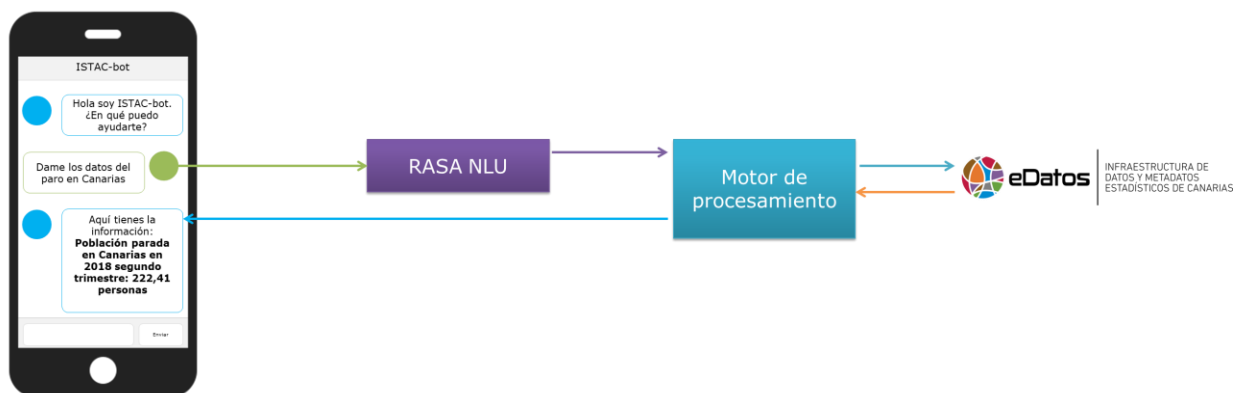


Figura 2: Diagrama de funcionamiento del chatbot para la obtención de datos estadísticos desde el sistema de indicadores del ISTAC (eDatos).

Para ver claramente el funcionamiento del chatbot, en la Figura 3 se muestra un ejemplo real de un usuario que interactúa con ISTAC-Bot. En este escenario, un usuario saluda a ISTAC-Bot y a continuación pregunta por los datos de “Venta de cemento para el año 2018”. El chatbot responde inmediatamente con el dato buscado por el usuario y, además, le da información acerca de la variación de ese dato con respecto al año anterior. En esa misma respuesta, el chatbot sugiere una nueva consulta variando las entidades y, por último, le muestra un listado de otros indicadores que pueden ser de su interés y que guardan relación con el indicador que ha consultado.

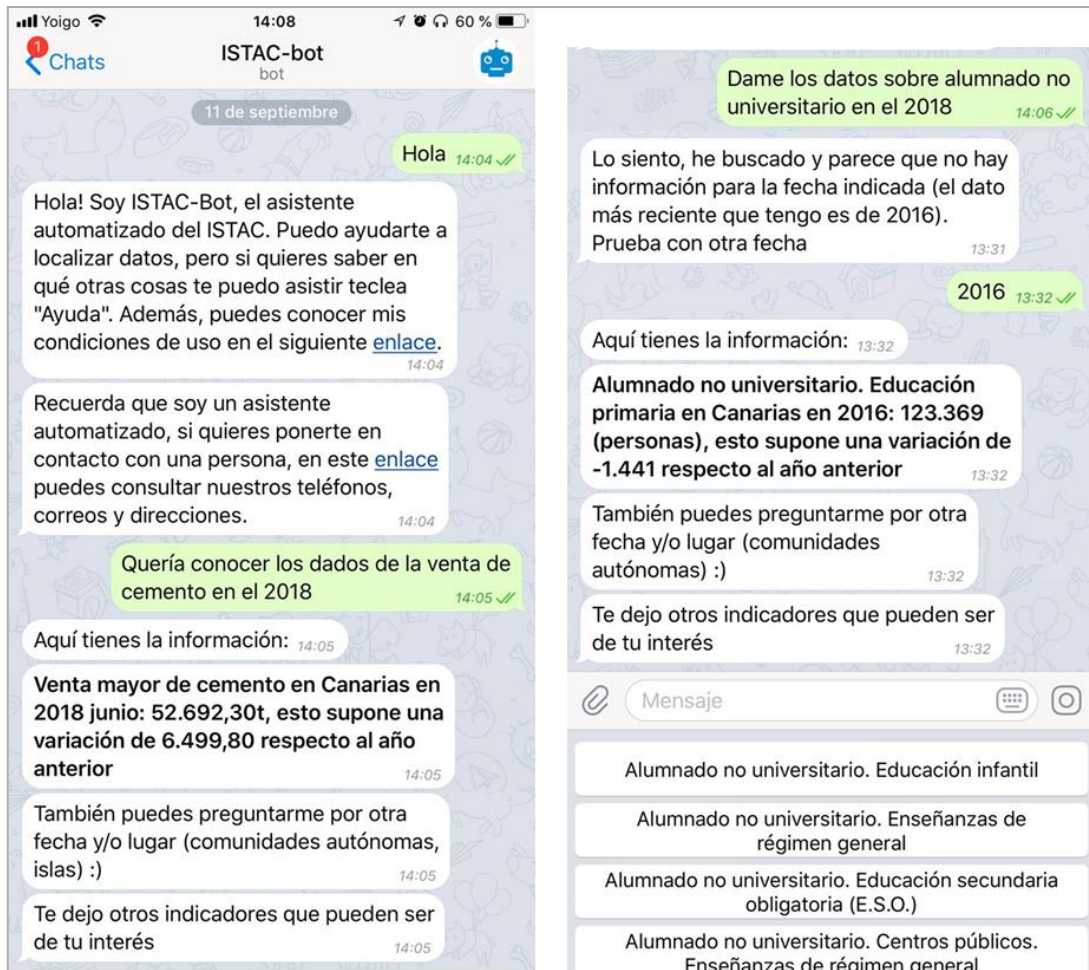


Figura 3: Ejemplo de interacción de un usuario con ISTAC-Bot.

Con este ejemplo queda demostrado que ISTAC-Bot es capaz de entender la petición del usuario, consultar el dato al sistema de indicadores del ISTAC (eDatos), dar la información y, además, sugerir contenido.

Conclusiones

Se ha presentado un chatbot desarrollado en el marco de la mejora de la comunicación entre el público generalista y el ISTAC: ISTAC-Bot. Éste es capaz de detectar la intención del texto introducido por el usuario gracias a su sistema de procesamiento de lenguaje natural. También extrae los datos importantes de la petición del usuario para poder buscar la información correspondiente en los sistemas del ISTAC. Además, el chatbot posee un sistema de sugerencias por si el usuario no da con la ortografía correcta, así como para mostrar indicadores relacionados con la búsqueda que se ha llevado a cabo.

Con este chatbot el ISTAC logra cumplir con los objetivos marcados, respondiendo, interactuando y mostrando los datos de los principales indicadores estadísticos publicados por el ISTAC. Con esta herramienta, el ISTAC conseguirá aumentar el número de usuarios que acuden al ISTAC en busca de datos estadísticos; facilitar el acceso a los datos del ISTAC mediante las herramientas que los usuarios utilizan en su día a día; mejorar la interfaz de consulta y las interacciones entre los usuarios y el ISTAC, y hacer mucho más amigable el acceso de los usuarios a los datos publicados por el ISTAC.

Este proyecto se encuentra en su primera fase de desarrollo, en fases posteriores se espera trasladar a ISTAC-Bot a otros sistemas de mensajería, como Whatsapp, o la propia web del ISTAC. Además, ISTAC-Bot aprende constantemente de las consultas de los usuarios, por lo que será cada vez más preciso y con el tiempo podrá dar mucha más información, incluyendo no solo el conjunto de indicadores sino también todos los conceptos que se relacionan con los datos estadísticos publicados.

Bibliografía

- [1] Gobierno de Canarias, «El Istac desarrolla un asistente virtual para localizar datos estadísticos mediante mensajería instantánea,» 28 09 2018. [En línea]. Disponible en: <http://www.gobiernodecanarias.org/noticias/h/100554/istac-desarrolla-asistente-virtual-localizar-datos-estadisticos-mediante-mensajeria-instantanea>.
- [2] El Día, «El Istac desarrolla un asistente virtual para localizar datos estadísticos vía Telegram,» 28 09 2018. [En línea]. Disponible en: <https://eldia.es/canarias/2018-09-25/6-Istac-desarrolla-asistente-virtual-localizar-datos-estadisticos-Telegram.htm>
- [3] Rasa, «Rasa Stack,» 28 09 2018. [En línea]. Disponible en: <https://rasa.com/products/rasa-stack/>.
- [4] SpaCy, «Industrial-Strength Natural Language Processing,» 28 09 2018. [En línea]. Disponible en: <https://spacy.io/>.
- [5] «MITIE: library and tools for information extraction,» 28 09 2018. [En línea]. Disponible en: <https://github.com/mit-nlp/MITIE>.
- [6] «Scikit-Learn,» 28 09 2018. [En línea]. Disponible en: <http://scikit-learn.org/stable/>.
- [7] TensorFlow, «An open source machine learning library for research and production,» 28 09 2018. [En

línea]. Disponible en: <https://www.tensorflow.org/>.

[8] Analytics Vidhya, «Bot-Arch,» 28 09 2018. [En línea]. Disponible en: <https://www.analyticsvidhya.com/blog/2018/01/faq-chatbots-the-future-of-information-searching/bot-arch/>.

Datos de contacto

Jesús Alberto González Yanes

igonyanp@gobiernodecanarias.org

Instituto Canario de Estadística (ISTAC)

922922800

Arte Consultores Tecnológicos S.L.

info@arte-consultores.com

922201617