

# Smile: Asistente virtual inteligente con detección de emociones

## *Smile: Intelligent virtual assistant with emotion detection*

Flor Miriam Plaza del Arco<sup>1</sup>, Salud María Jiménez-Zafra<sup>1</sup>,  
Daniel García-Baena<sup>2</sup>, Miguel Ángel García-Cumbreras<sup>1</sup>

<sup>1</sup>Departamento de Informática, SINAI, CEATIC, Universidad de Jaén, Jaén, España  
{fmplaza, sjzafra, magc}@ujaen.es

<sup>2</sup>I.E.S. San Juan de la Cruz, Úbeda (Jaén), España.  
daniel.gbaena@gmail.com

**Resumen:** Smile es un demostrador de un asistente virtual inteligente con detección de emociones. Se trata de una utilidad (*skill*), en concreto para Amazon Alexa, que haciendo uso de las Tecnologías del Lenguaje Humano incorpora un detector de emociones. Se ha diseñado un sistema de diálogo que interactúa con el usuario en función de la emoción detectada. Para el sistema de detección de emociones se ha utilizado un sistema que combina la aplicación de dos lexicones, teniendo como premisa que el sistema debe resolver la emoción en tiempo real para ser usable como *skill*.

**Palabras clave:** Detección de emociones, Asistente virtual inteligente.

**Abstract:** Smile is a demonstrator of an intelligent virtual assistant with emotion detection. It is a skill, specifically for Amazon Alexa, that makes use of Human Language Technologies and incorporates an emotion detector. A dialog system has been designed that interacts with the user depending on the detected emotion. For the emotion detection system we have used a system that combines the application of two lexicons, taking as a premise that the system must resolve the emotion in real time to be usable as a skill.

**Keywords:** Emotion detection, Intelligent virtual assistant.

## 1 Introducción y motivación

Un Asistente Virtual Inteligente (AVI) es un agente basado en software que tiene habilidades de aprendizaje permanente y cuyo objetivo consiste en ayudar a los usuarios de sistemas computacionales a la hora de realizar multitud de tareas, haciendo que la interacción entre la persona y la máquina sea mínima (Hayes, 2017). Gartner predice que el 25 % de los domicilios tendrán uno o más dispositivos con AVIs en 2021<sup>1</sup>.

Por otro lado, un Sistema Conversacional es un programa que permite al usuario simular una conversación con él mediante el lenguaje natural (voz o texto). Estos sistemas se caracterizan por ser menos inteligentes que los AVI ya que no siempre entienden el mensaje enviado por el usuario. Se basan en una serie de respuestas preparadas previamente.

El uso de los asistentes virtuales integrados en dispositivos comerciales (Google Home o Nest, Amazon Alexa o HomePod de Apple) ha aumentado de forma considerable en el uso doméstico. Son accesibles funcional y económicamente, lo que permite el diseño y desarrollo de aplicaciones dedicadas, por ejemplo para la mejora de la autonomía de las personas mayores. En este sentido proponemos un sistema que detecta la emoción de cada persona, pudiendo modificar la conversación o acción en función de la misma.

En este trabajo presentamos Smile, un demostrador de un asistente virtual inteligente que es capaz de gestionar la conversación con el usuario en función de la emoción detectada. Para el asistente inteligente hemos hecho uso de Alexa como núcleo del sistema. Para el sistema de detección de emociones hemos hecho uso de un sistema propio que utiliza un enfoque basado en lexicones, teniendo en

<sup>1</sup><https://gtnr.it/3xtqzCD>

cuenta que se trata de una aplicación donde el tiempo de respuesta es un requisito fundamental. La Figura 1 muestra un esquema general del sistema desarrollado.

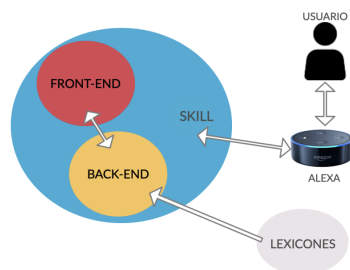


Figura 1: *Smile* - Esquema general.

## 2 Detección de emociones

Las emociones son rasgos básicos de los humanos y ya han sido estudiadas por investigadores de diversos campos como la psicología, la sociología y la informática. El reconocimiento de emociones se enmarca dentro del área de la Computación Afectiva, la cual abarca diversas disciplinas como el Análisis de Sentimientos (Yadollahi, Shahraki, y Zaiane, 2017), que se encarga del estudio y la evaluación del estado anímico de los humanos ante diferentes situaciones.

En el campo de la psicología, existen diversas teorías de clasificación de emociones, las cuales se dividen en dos vertientes principales: modelos de emociones discretas y modelos de emociones dimensionales (Canales y Martínez-Barco, 2014).

*Modelos de emociones discretas.* Esta teoría defiende que existe un número limitado de emociones que una persona puede experimentar. Dos de los principales modelos de emociones discretas son el de Ekman (Ekman, 1999), que propuso un modelo de clasificación de emociones basada en seis categorías (ira, asco, miedo, alegría, tristeza y sorpresa), y el de Plutchik (Plutchik, 1980), que incluyó dos más (confianza y anticipación). La mayoría de los estudios que se han realizado para la detección de emociones aplicando técnicas de PLN siguen este tipo de modelo (Mohammad y Turney, 2013; Plaza-del Arco et al., 2018; Plaza-del-Arco et al., 2020).

*Modelos de emociones dimensionales.* A diferencia del modelo anterior, las emociones son representadas en un espacio dimensional. Los modelos más conocidos son el de Russell (Russell, 1980), con una clasificación de las

emociones dentro de un espacio circular bidimensional, con una dimensión para la valencia y otra para la excitación, y el modelo de Mehrabian (Mehrabian, 1996), que tiene en cuenta tres dimensiones independientes para describir el estado de ánimo de las personas (placer, excitación y dominio).

### 2.1 Estudio y análisis de lexicones

Se denomina lexicón a un diccionario o listado de términos representativos de una lengua, región, etc. Los lexicones afectivos están formados por una lista de palabras asociadas a una categoría de emoción (tristeza, alegría, miedo, etc). Para el español encontramos actualmente escasos lexicones marcados con emociones:

1. **Linguistic Inquiry and Word Count (LIWC) (Tausczik y Pennebaker, 2010).** Disponible para inglés y español, está basado en un conjunto de diccionarios, entre los que se encuentra un diccionario de palabras sociales, signos de puntuación, palabras afectivas, etc. Proporciona un lexicón de términos etiquetados con una serie de factores, entre los que se encuentran las emociones relacionadas con enojo y tristeza. Fue desarrollado por investigadores interesados en psicología social, de salud, etc., por lo que las categorías del lenguaje fueron creadas para capturar estados psicológicos y sociales de las personas.
2. **NRC Emotion Lexicon o Emolex (Mohammad y Turney, 2013).** Es un lexicón en inglés formado por 10.000 términos, que también está disponible para español. Cada entrada incluye un término, una emoción y una medida de la asociación entre el término y la emoción. Sigue el modelo de emociones de Plutchik.
3. **SEL (Rangel, Sidorov, y Guerra, 2014).** Es un diccionario formado por 2.036 palabras en español, etiquetadas con las seis emociones básicas de Ekman. Se basa en WordNet-Affect <sup>2</sup>, un recurso construido para el inglés y etiquetado por categorías emocionales.

<sup>2</sup><https://wndomains.fbk.eu/wnaffect.html>

### 3 Descripción del sistema

El sistema está formado por el desarrollo de un *skill* de Alexa para la parte frontend y un módulo Python para el backend.

Veamos a continuación un ejemplo de diálogo de la aplicación desarrollada:

*Usuario*: “Alexa, abre Smile.”

*Alexa*: “Buenos días, ¿cómo llevas el día?”

*Usuario*: “Pues muy bien, porque acabo de aprobar el carnet de conducir.”

*Alexa (alegría detectada)*: “No tengo mannos pero ¡choca esos cinco!. ¿Quiéres contarme algo más?”

*Usuario*: “Sí claro. He preparado una comida riquísima.”

*Alexa (alegría detectada)*: “Estupendo. Si pudiera probarla estaría encantada. ¿Me cuentas algo más?”

*Usuario*: “No, eso es todo.”

*Alexa*: “Vale, hasta luego.”

#### 3.1 FrontEnd

El front-end de una *skill* de Amazon Alexa está formado por diversos componentes. El **Invocation Name** es el nombre con el que se invocará la *skill* por parte del usuario; los **Intents** son las diferentes acciones que proporciona la *skill* para que el usuario las utilice. En nuestro caso, se ha creado un único intent, el detector de emociones; Las **Utterances** son oraciones definidas para cada intent y que sirven para lanzarlos. En el caso de nuestra *skill*, se ha definido un único utterance que será el slot del mensaje que proporciona el usuario; por último, los **Slots** son registros dentro de *utterances* con los que se pretende obtener información del usuario. En nuestra *skill*, existe un único slot, que será el mensaje que proporcione el usuario para analizar y reconocer la emoción, y que será obligatorio.

Se ha definido un conjunto amplio de posibles consultas con las que el *skill* inicia la conversación con el usuario, entre las que encontramos “¿cómo estás?”, “¿cómo te ha ido el día?”, “¿cómo fue el día de ayer?”, etc. De esta forma se inicia una conversación con el usuario, solicitándole más información si las respuestas son frases muy cortas. Una vez obtenido el mensaje, y procesado, la *skill* proporcionará una respuesta al usuario y preguntará si quiere decir algo más. En caso de reconocer un estado neutro, se volverá a realizar otra pregunta para obtener otro mensaje para analizar. En cada interacción con el usuario el *skill* hace una llamada al sistema

de backend, que implementa el detector de emociones.

El último paso de cada iteración con el *skill* consiste en dar una respuesta al usuario. Esta variará en función de la emoción que se haya reconocido, aportando un mensaje que ayude al usuario, le informe acerca de la emoción que siente o haga que la *skill* empatice con él. A continuación se muestra la respuesta que se dará dependiendo de la emoción detectada:

- **Alegría**. Se refuerza esa emoción ya que es positiva.
- **Tristeza**. Se transmite un mensaje que ayuda a animar al usuario.
- **Enojo**. Se transmite un mensaje que ayuda a calmar al usuario.
- **Sorpresa**. Se transmite un consejo o un mensaje que ayuda al usuario a relajarse.
- **Neutro**. Se intenta animar al usuario a que proporcione más información para poder reconocer una emoción.

Para cada emoción detectada se han definido un conjunto amplio de respuestas que aleatoriamente interactúan con el usuario. Además, toda la interacción, así como la emoción detectada en cada caso, quedan almacenados para su posterior evaluación.

#### 3.2 Backend

Para el sistema de clasificación de emociones se ha utilizado un enfoque basado en lexicones, haciendo uso de LIWC, Emolex y SEL.

Se realizó una evaluación de estos tres lexicones utilizando la versión en español del corpus EmoEvent (Plaza del Arco et al., 2020) filtrando las cuatro emociones utilizadas en este proyecto, obteniendo un total de 4025 tweets: 858 que expresan enojo, 1816 alegría, 1008 tristeza y 344 sorpresa. Cada registro fue categorizado por el sistema de clasificación de emociones creado, con el objetivo de obtener la misma emoción que indicaba el dataset. Los resultados más interesantes se obtuvieron con la combinación de lexicones SEL y LIWC, en la que SEL sería el recurso principal ya que tiene términos para todas las emociones y LIWC serviría para reforzar las emociones tristeza y enojo, ya que es la combinación con mejor porcentaje de acierto respecto al fallo, además de tener el mayor número de tweets en los que se acierta

la emoción y de las que menos tweets etiquetados como neutros tienen. A nivel general, esta combinación obtuvo unos resultados de un 66,19% de precisión.

El contenido que expresa el usuario, obtenido en formato textual gracias al módulo STT (*Speech To Text*) de Alexa, es procesado y se buscan coincidencias de los términos que forman la oración procesada. Se lleva un contador para cada emoción, al que se va sumando el peso de cada término que represente dicha emoción y que esté presente tanto en la oración como en el lexicon. Un término será clasificado en la emoción para la que obtenga una mayor puntuación. Se tomará como resultado de la clasificación un estado neutro cuando no se encuentra ninguna coincidencia entre los términos de la oración y los lexicones o en el caso de que el valor del contador de varias emociones sea el mismo.

#### 4 Conclusiones

En este artículo se presenta Smile, un asistente virtual inteligente con detección de emociones desarrollado como *skill* para Amazon Alexa. Este sistema combina dos de las áreas de la Informática con más auge en la actualidad, los AVI y el PLN. Con esta aplicación se pretende mantener un diálogo con el usuario de forma natural, dándole una respuesta al mismo en función del estado anímico detectado a través de varios recursos léxicos.

#### Agradecimientos

Este trabajo ha sido parcialmente financiado por el Fondo Social Europeo, la Administración de la Junta de Andalucía (DOC\_01073), el Fondo Europeo de Desarrollo Regional (FEDER), el Ministerio de Ciencia, Innovación y Universidades (Beca FPI-PRE2019-089310), el proyecto LIVING-LANG (RTI2018-094653-B-C21) del Gobierno de España y el proyecto InLife FGCSIC/PSL-INTERREG/FEDER.

#### Bibliografía

Canales, L. y P. Martínez-Barco. 2014. Emotion detection from text: A survey. En *Proceedings of the workshop on natural language processing in the 5th information systems research working days (JISIC)*, páginas 37–43.

Ekman, P. 1999. Basic emotions. *handbook of cognition and emotion*. Wiley, New York, páginas 301–320.

Hayes, A. 2017. *Amazon Alexa: A Quick-start Beginner's Guide*. CreateSpace Independent Publishing Platform.

Mehrabian, A. 1996. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292.

Mohammad, S. M. y P. D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational intelligence*, 29(3):436–465.

Plaza-del-Arco, F. M., M. T. Martín-Valdivia, L. A. Ureña-López, y R. Mitkov. 2020. Improved emotion recognition in spanish social media through incorporation of lexical knowledge. *Future Generation Computer Systems*, 110:1000–1008.

Plaza-del Arco, F. M., M. D. Molina-González, S. M. Jiménez-Zafra, y M. T. Martín-Valdivia. 2018. Lexicon adaptation for spanish emotion mining. *Procesamiento del Lenguaje Natural*, 61:117–124.

Plaza del Arco, F. M., C. Strapparava, L. A. Ureña Lopez, y M. Martin. 2020. EmoEvent: A multilingual emotion corpus based on different events. En *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association.

Plutchik, R. 1980. A general psychoevolutionary theory of emotion. En *Theories of emotion*. Elsevier, páginas 3–33.

Rangel, I. D., G. Sidorov, y S. S. Guerra. 2014. Creación y evaluación de un diccionario marcado con emociones y ponderado para el español. *Onomazein*, (29):31–46.

Russell, J. A. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Tausczik, Y. R. y J. W. Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Yadollahi, A., A. G. Shahraki, y O. R. Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):1–33.