



Identificación del Trabajo	
Área:	Sistemas de información e informática
Categoría:	Alumno
Regional:	Santa Fe

WALLE: Asistente inteligente de ventas basado en sistemas de producción.

José Martín GONZÁLEZ, Fernanda Edith MIR, Facundo Nahuel PESOA

Lavaisse 610, Santa Fe, Facultad Regional Santa Fe, UTN

E-mail de autores: jgonzalez@frsf.utn.edu.ar, femir@frsf.utn.edu.ar, fpesoa@frsf.utn.edu.ar

Este trabajo ha sido realizado bajo la dirección la Dra. Ma. de los Milagros Gutiérrez y, del Dr. Jorge Roa en el marco del proyecto PID 4871 “Aplicación de técnicas de inteligencia artificial al diseño y análisis de procesos de negocio”. (2019 – 2019)

Resumen

Se presenta la implementación del prototipo de un chatbot con inteligencia artificial, diseñado para poder interactuar con clientes y vendedores de locales comerciales. El objetivo de este agente es ser un “asistente virtual”; es decir, brindar información y recomendaciones sobre productos a los clientes, como así también brindar información a los vendedores sobre los productos y sobre los gustos de clientes. Para que el chatbot pueda dar respuestas, se utilizan técnicas de procesamiento de lenguaje natural y se implementa un sistema de producción con encadenamiento hacia adelante para hallar la respuesta.

Palabras Claves: Chatbot; Inteligencia Artificial; Sistema de Producción; Procesamiento del lenguaje natural

1. Introducción

Este trabajo presenta el diseño e implementación de un chatbot que se desenvuelve en un entorno comercial. Este chatbot es un agente lógico, encargado de interactuar y mediar entre el vendedor y el cliente como un amigable “asistente virtual”. Los chatbot se han popularizado hoy en día, gracias al avance de la inteligencia artificial y principalmente del uso de técnicas de procesamiento de lenguaje natural, así como también de la proliferación de algoritmos capaces de transformar el lenguaje hablando (sonido) a lenguaje escrito (texto).

Albayrak et al.(2018). describen al chatbot como un software que puede conversar con personas, usando inteligencia artificial para ello. Este tipo de software es ampliamente usado para realizar tareas tales como: responder rápidamente a usuarios o clientes dándole la información que necesitan y ayudándolos a comprar el producto que están buscando y así poder mejorar la experiencia del cliente.

Con respecto a inteligencia artificial, si bien existen muchas definiciones como las propuestas por Russel y Norving (2004). y Nilson (2000)., en este trabajo se adopta la definición dada por Gershenfeld (2009)., donde se considera la inteligencia artificial como la disciplina que estudia la creación de máquinas inteligentes, cualquiera sea el concepto de inteligencia.

Dado que el chatbot presentado en este trabajo es un agente inteligente, debe representar el modelo básico de percepción y acción sobre un entorno. Se define que las percepciones del mismo son los mensajes enviados por el usuario a través de audio o texto, la cual se procesa mediante un componente de lematización. Luego, el agente, formula una posible respuesta a través de una máquina de inferencias por medio de la técnica de encadenamiento hacia adelante. Finalmente, el chatbot actúa sobre el entorno mediante una respuesta de audio y texto, acorde a lo solicitado.

2. Metodología

En este apartado se presenta la metodología utilizada en el modelado y desarrollo del chatbot. En primera medida, se evaluó el problema propuesto y se planteó una solución modelando para ello un agente. Posteriormente se evaluó, definió y acotó el vocabulario con el que se iba a trabajar, debido a que es necesario que el chatbot conozca el tema sobre el cuál deberá dar respuesta. Este paso requirió estudiar el dominio y encontrar las palabras claves y frases más usuales. Se destaca que si bien para este desarrollo se usó un vocabulario específico para una empresa de ventas de electrodoméstico, el diseño que se plantea es flexible, de manera que el agente pueda aprender de otros dominios y dar respuestas confiables sobre este. Por último, una vez definido el vocabulario se definieron las reglas que le permiten al agente encontrar la respuesta ante una pregunta del usuario.

2.1. Planteo de la solución

La Figura 1 muestra la arquitectura del agente chatbot, con los componentes necesarios para dar solución al problema propuesto. Dado que el chatbot debía poder interactuar de forma amigable con el entorno, se decidió utilizar un formato de entrada y salida Speech to Text o de entrada estándar. Por este motivo se utiliza un lenguaje que proporciona un soporte confiable a estas funcionalidades.

Para el procesamiento de las entradas, se realiza una lematización de la entrada a un dominio acotado del vocabulario que pueda comprender el agente. El dominio del chatbot está compuesto por predicados, constantes y palabras claves que son utilizadas por el sistema de producción para dar con una respuesta adecuada y obtener conocimiento.

2.2. Lematizador

En primera medida, se realizó un vocabulario básico del lenguaje que va a ser manejado por el agente en el entorno correspondiente. Se tomaron ejemplos de conversaciones donde participaban clientes con vendedores y se analizaron las frases más frecuentes para la obtención de un conjunto acotado de palabras claves útiles que le permita identificar la intención de la pregunta del usuario.

A través de la lematización, se utilizó un mecanismo que permite establecer la menor cantidad de palabras para formular una respuesta. Este proceso permite despojar la frase de las palabras que no son relevantes para entender el contexto y de esta manera mejorar el rendimiento del agente.

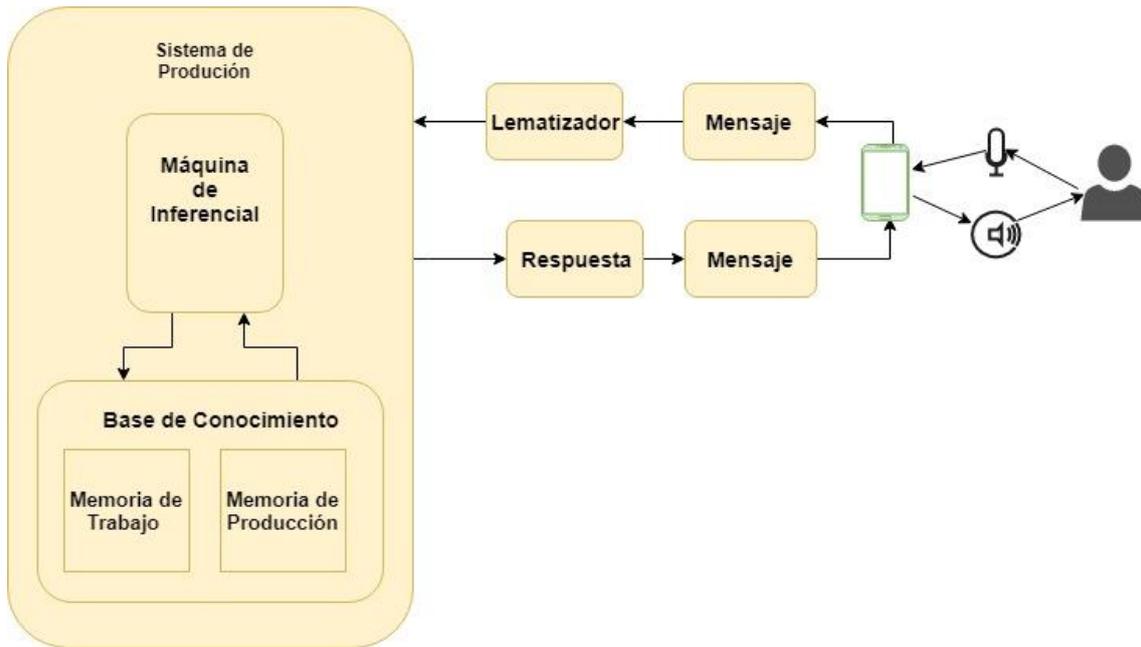


Figura 1. Esquema de componentes necesarios

En la Tabla I, se presentan algunos ejemplos de preguntas, sus posibles respuestas y las formas reducidas que tendrían las mismas para poder llegar a ser comprendidas y utilizadas posteriormente para realizar inferencias lógicas. En color verde se representa el lenguaje de interés para el usuario y en azul el lenguaje de interés para el vendedor.

Tabla I. Ejemplos del vocabulario básico del agente.

Pregunta	Posible respuesta	Palabra Clave
Qué PRODUCTO de MARCA tenés?	Los PRODUCTO MARCA disponibles son: Lista<modelo> MARCA? me quedan: LIST<MODELO>	que PRODUCTO MARCA tener
Cuál PRODUCTO de MARCA tenés?		cual PRODUCTO MARCA tener
Qué PRODUCTO de MARCA hay?		que PRODUCTO MARCA hay
Cuál PRODUCTO de MARCA hay?		cual PRODUCTO MARCA hay
Cuántas personas preguntaron por PRODUCTO?	Solo 1 persona Fueron 7 las consultas realizadas sobre PRODUCTO	cuanto preguntar PRODUCTO
Cuántas personas consultaron por PRODUCTO?		

Cuál es el PRODUCTO más barato que tenés?	El PRODUCTO más barato que tenemos es: MODELO y vale PRECIO El MODELO es el más económico.	cual ser PRODUCTO barato que tener
Tenés algún PRODUCTO barato?		tener PRODUCTO barato
Qué le gusta el cliente X?	Prefiere: MARCA Y Le gustan más: MARCA Y	que gustar CLIENTE
Qué prefiere X?		que gustar CLIENTE

2.3. Sistema de producción

Un Sistema de producción es una estrategia usada en inteligencia artificial para implementar sistemas expertos inteligentes basados en conocimiento. Está comprendido por dos grandes partes: la Máquina de Inferencias y la Base de Conocimiento.

Bruner (1957)., fue el primero en identificar la mente humana con una “Máquina de Inferencias” al referirse a su destreza para activar el conocimiento ya almacenado y utilizarlo para organizar e interpretar la nueva información entrante, a través de complejas relaciones abstractas no provenientes de los estímulos.

La Base de Conocimiento—precargada desde archivos externos—, contiene un conjunto de doce (12) reglas iniciales, como se evidencia en la Tabla II, que conforman la Memoria de Producción, y un conjunto de ciento veintitrés (123) literales positivos «Ground» que conforman la Memoria de Trabajo inicial que referencian al conjunto de conocimientos previos ejemplificados a continuación:

- esMarca(celular,samsung) → esMarca;celular;samsung
- esModelo(samsung,s10) → esModelo;samsung;s10
- esColor(s10,dorado) → esColor;s10;dorado
- esMarca(televisor,samsung) → esMarca;televisor;samsung
- esMarca(tablet,xview) → esMarca; tablet; xview
- esModelo(asus,vivobook) → esModelo;asus;vivobook

Tabla II. Reglas realizadas.

Número	Regla
1	esModelo(X,Y) ^ esColor(Y,Z) -> esColor(Y,Z) ^ add(esConsulta(C,elColor,Y))
2	esModelo(X,Y) ^ esPrecio(Y,Z) -> esPrecio(Y,Z) ^ add(esConsulta(C,elPrecio,Y))
3	esProducto(X) ^ esMarca(X,Y) -> esMarca(X,Y)

4	$esMarca(X, Y) \wedge esModelo(Y, Z) \wedge esDescripcion(X, Y, Z) \rightarrow esModelo(Y, Z) \wedge add(esConsulta(C, X, Y))$
5	$esConsulta(C, X, Y) \wedge esConsulta(C, Z, Y) \wedge esMarca(X, Y) \wedge esMarca(Z, Y) \rightarrow gusta(C, Y) \wedge add(gusta(C, Y))$
6	$esProducto(W) \wedge esMarca(W, X) \wedge esModelo(X, Y) \wedge esDescripcion(W, X, Y) \wedge esPrecio(Y, Z) \rightarrow esPrecio(Y, Z) \wedge add(esConsulta(C, masBarato, X))$
7	$esCliente(Z) \wedge esConsulta(Z, X, Y) \rightarrow esConsulta(Z, X, Y)$
8	$esProducto(X) \wedge esMarca(X, Y) \wedge esModelo(Y, Z) \rightarrow esModelo(Y, Z) \wedge add(esConsulta(C, X, Z))$
9	$esProducto(X) \wedge esMarca(X, Y) \wedge esModelo(Y, Z) \wedge esDescripcion(X, Y, Z) \wedge esCaracteristica(Z, W) \rightarrow esCaracteristica(Z, W) \wedge add(esConsulta(C, X, Z))$
10	$esModelo(Y, Z) \wedge esCaracteristica(Z, W) \rightarrow esCaracteristica(Z, W) \wedge add(esConsulta(C, Y, Z))$
11	$esProducto(Z) \wedge esConsulta(X, Z, Y) \rightarrow esConsulta(X, Z, Y)$
12	$esCliente(X) \wedge leGusta(X, Y) \rightarrow leGusta(X, Y)$

La Máquina de Inferencias, representada en la Figura 2, permitirá al chatbot adquirir nuevos conocimientos mediante la estrategia de razonamiento de encadenamiento hacia adelante, la cual se basa en datos. Debido a la naturaleza del problema que se presenta, donde el agente conoce con certeza cuáles son los datos de entrada pero no las conclusiones a las cuales llegará, es que se seleccionó esta estrategia en lugar de encadenamiento hacia atrás, la cual es guiada por objetivos que se intentan alcanzar desde la fundamentación de hipótesis a verificar.

El agente captará ciertas palabras que relacionará con los saberes de la memoria de trabajo y con las reglas de la memoria de producción a través de las diferentes fases.

2.4. Diseño del agente

El agente se desarrolló en una tecnología Android, lo cual facilitó la implementación del proceso de transformación de lenguaje hablado a texto y viceversa o Speak to Text y Text to Speak. Debido a la tecnología seleccionada, la cual solo permite el desarrollo de aplicaciones nativas en el lenguaje de programación Java, la solución del problema se planteó con un diseño de orientación a objetos. De este mismo diseño, se expone un diagrama de clases reducido de la implementación del prototipo, tal como se muestra en la Figura 3.

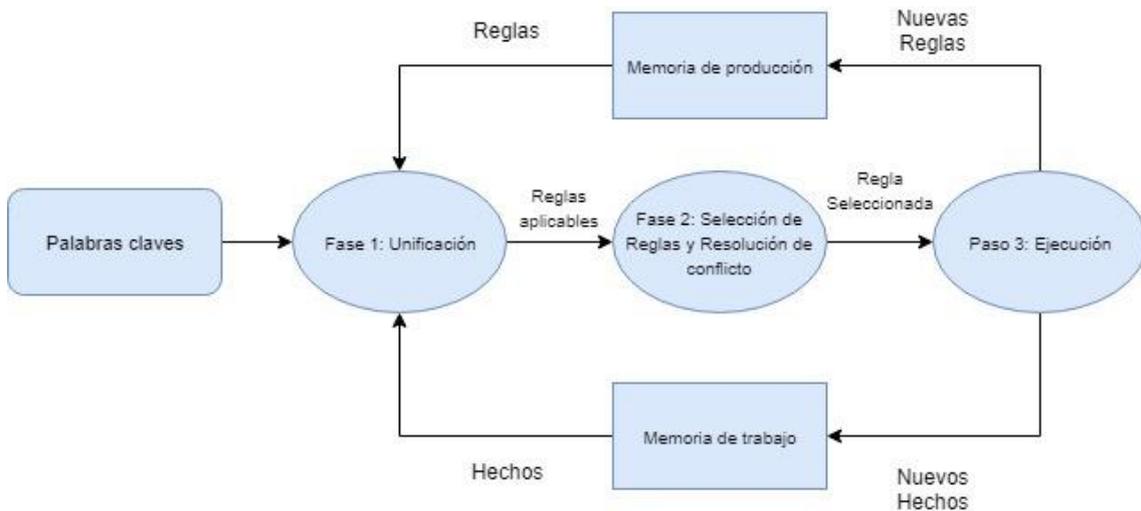


Figura 2. Esquema de Máquina de Inferencias.

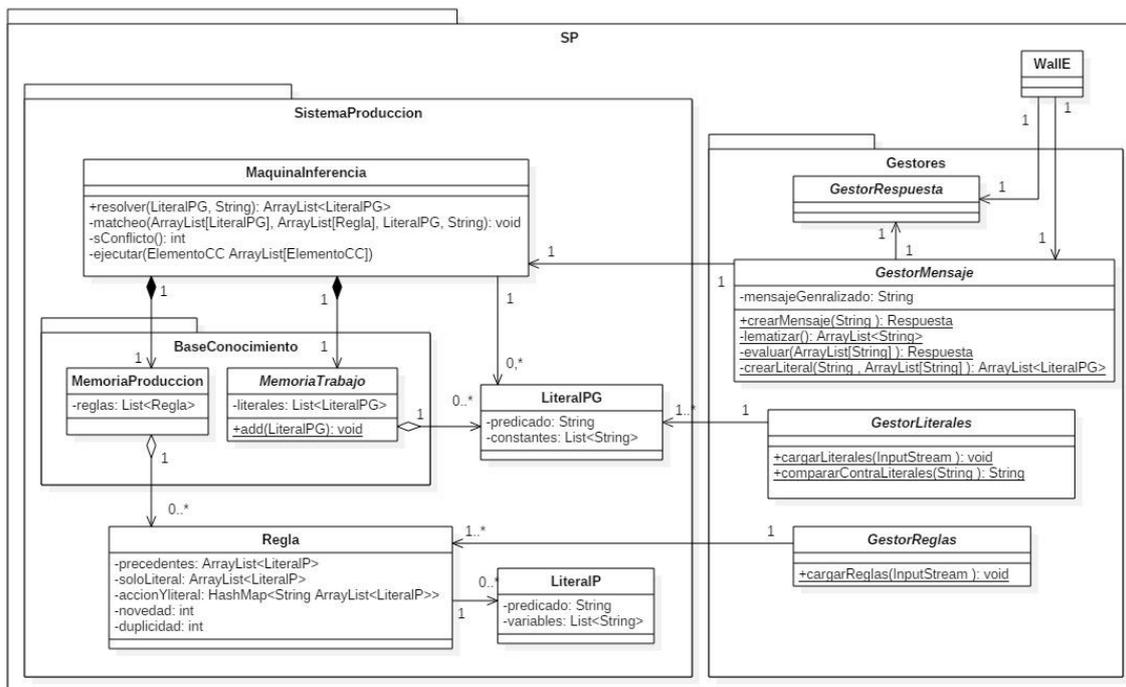


Figura 3. Diagrama reducido de clases

Como se observa, se optó por el uso de gestores que dividen la interacción del usuario de la máquina de inferencias que genera las posibles respuestas. De esta manera, se aporta un modelo ágil y modular.

3. Resultados

El prototipo desarrollado logró cumplir con los objetivos deseados; es decir, interpretar un mensaje hablado y lograr dar una respuesta adecuada en un tiempo acorde al modo en que se procesan los datos. Se definió e implementó un sistema de producción basado en objetos que codifica las reglas usadas por el chatbot para tomar decisiones. Además se desarrolló un vocabulario con el que el agente logra trabajar y comunicarse.

Para la evaluar el agente en forma cuantitativa y debido a que el procesamiento del mensaje atraviesa distintos componentes, su análisis se debe dividir en dos partes. Estas se desarrollarán posteriormente.

3.1. Entendimiento

Se evalúa la capacidad del chatbot de entender por medio de audio o mensajes lo que la persona le dice. En este caso, los problemas más significativos son respecto a la superposición del audio o al no entendimiento de una palabra.

La superposición del audio se da cuando el usuario realiza una consulta mientras el agente está contestando la pregunta anterior o hay demasiado ruido en el ambiente. Aquí es cuando el agente capta basura (palabras que no son dichas por el usuario) o no termina de captar la respuesta no transcribiendo de audio a texto algunas palabras. De las pruebas llevadas a cabo, se pudo obtener que en un 90% de los casos, el complemento de Google para Android que nos permite este traspaso de audio a texto logra capturar las palabras de manera correcta, por más que no todas pertenezcan al dominio del agente.

El no entendimiento de una palabra se puede dar por el uso de algún modismo por parte del usuario, por frases que no se encuentran bajo el dominio del agente o porque el usuario habla muy rápido. En estos casos es cuando se solicita la repetición o especificación de la pregunta, sucediendo eventualmente en un 10% de los casos.

3.2. Procesamiento

El procesamiento del mensaje en sí a través de su clasificación y su posterior paso por la máquina de inferencias para llegar a una posible respuesta es otra de las partes a evaluar.

De cinco (5) preguntas realizadas en una conversación, de un total de veinte siete (27) preguntas posibles —once (11) pertenecientes a distintos entornos, es decir que no buscan la misma respuesta—, solo una (1) dio una respuesta incorrecta.

Se evaluó los tiempos de ejecución de la máquina de inferencias con la memoria de trabajo inicial de ciento veintitrés (123) literales positivos. La máquina contiene el algoritmo de backtracking en la fase de unificación que suele tener un orden de complejidad algorítmica exponencial **(1)**.

En diez (10) de sus preguntas, la respuesta fue menor que cien (100) milisegundos. El caso que se evaluó detalladamente fue la pregunta: “Qué productos tenés?”, dado que tiene diez (10) palabras posibles a machear con las variables en este algoritmo. La respuesta a esta pregunta tarda doscientos diez (210) milisegundos y, por ser la de mayor tiempo de respuesta, se analizó lo que pasaría con la misma si la cantidad de palabras a machear con las variables de backtracking fueran incrementando hasta obtener cien (100) palabras. Con los resultados de los puntos intermedios, se visualizó en un gráfico y se notó que se aproximaban a una recta, por lo cual se calculó su recta de regresión lineal **(2)** para luego poder extrapolar y evaluar qué sucedería en caso de que se incrementaran las posibilidades de macheo.

$$O(m^n); \text{ donde } \begin{cases} m = \text{posibles decisiones} \\ n = \text{cantidad de ramas creadas} \end{cases} \quad (1)$$

$$y = 10,79x + 56,61 \quad (2)$$

4. Discusión

El trabajo con la tecnología Android aportó muchas facilidades al desarrollo de los componentes interactivos pero, por otro lado, no se encontraron extensiones que pudieran facilitar el uso del sistema de producción o la lematización de forma óptima en el lenguaje español. Por ello, se optó por desarrollarlas por completo, dado que se le otorgó prioridad a la posibilidad de brindar asistencia a un usuario al alcance de un dispositivo móvil o tablet.

Se buscó en primera medida conectar Java con Prolog en Android para poder implementar el sistema de producción de forma más sencilla, pero luego de varios intentos fallidos, se optó por descartar esta opción debido a que no se encontraron librerías que satisfagan las expectativas. Por tal motivo, se escogió desarrollar la primera fase, denominada unificación, utilizando backtracking —uno de los principios de Prolog— para la máquina de inferencias.

A pesar de que durante la definición de un vocabulario acotado se pudo haber optado por la inserción estática de respuestas, este modelo no era óptimo ni factible para las metas y objetivos propuestos, ya que perdía la flexibilidad aportada por cargar la memoria de trabajo y de producción de forma externa. Por otro lado, el planteo de un modelo general ayudó a una implementación fiel de cada uno de los componentes y las fases de un sistema de producción.

Si bien la aplicación se encuentra orientada a la venta de productos electrónicos, la meta del prototipo es que sea una aplicación flexible. Esto llevó al planteo de modelos más generales, lo cual le permite a un desarrollador modificar el menor número de componentes para que la aplicación se utilice en otro contexto, es decir, otros dominios sin afectar el funcionamiento del Sistema de Producción y solo por el mínimo intercambio de los archivos por donde se realiza la carga de datos externa que contienen las reglas y los hechos a utilizar provenientes de la memoria de producción y la memoria de trabajo.

5. Conclusiones

Hemos confirmado que es factible la implementación de un chatbot con una Máquina de Inferencias desarrollada en entorno Android. El agente tuvo una buena performance en el tiempo de respuesta al cliente y en la selección de un respuesta acorde a la/s pregunta/s generada/s.

Se planea continuar el desarrollo del prototipo alojando con posterioridad las memorias de producción y trabajo en una base de datos Firebase. Así se podrá realizar tanto la carga de información de productos de forma remota, como las disponibilidades y los cambios en precios.

Asimismo, se desea documentar detalladamente cada componente y dar una guía para la reutilización del chatbot en otros entornos comerciales, debido a que solo representa un cambio en los datos de entrada y otros cambios menores y no tan significativos en el gestor de respuesta.

Reconocimientos

Agradecemos el acompañamiento y orientación de los directores de proyecto, al Dr. Juan Leonardo Sarli por su compañía durante el proceso, y a los profesores de la UTN FRFSF por su dedicación en cada materia, lo cual ayudó a la formación de estos tres (3) estudiantes para integrar conocimientos adquiridos y nuevos en este proyecto.

Bibliografía

Albayrak, N., Özdemir, A., Zeydan, E. (2018). An overview of artificial intelligence based chatbots and an example chatbot application. 26th Signal Processing and Communications Applications Conference (SIU) Turquía: IEEE.

Russell, S., Norvig, P. (2004). La Inteligencia Artificial: Un Enfoque Moderno. Madrid, España: Prentice Hall.

Nilsson, N. (2000). Inteligencia Artificial. Una nueva síntesis. Madrid, España: Mc Graw Hill

Gershenfeld, N. (2009). Rethinking Artificial Intelligence. Web: <http://news.mit.edu/2009/ai-overview-1207>

Bruner, J. (1957). Going beyond the information given. En H. E. Gruber, K. R. Hammond y R. Jessor (Eds.), Contemporary approaches to cognition (pp. 41-69). Cambridge, MA: Harvard University Press.