



ASISTENTE VIRTUAL CHATBOT PARA LA WEB DE LA UNIVERSIDAD DEL VALLE
Desarrollo de Software

DIEGO ALEXANDER GIRALDO HOYOS
1310235-3743
diego.alexander.giraldo@correounivalle.edu.co

Oscar Bedoya
Doctor
oscar.bedoya@correounivalle.edu.co

Facultad de Ingeniería
Escuela de Ingeniería de Sistemas y Computación
Programa Académico de Ingeniería de Sistemas
Cali, Julio 5 de 2016

Resumen

Este trabajo de grado presenta el desarrollo de un asistente virtual para el sitio web de la Universidad del Valle. Un asistente virtual es un derivado de los buscadores que es capaz de analizar las peticiones del usuario en lenguaje natural y ofrecer respuestas también en lenguaje natural que satisfagan sus necesidades. Así, el asistente virtual tiene que cumplir con el propósito de mejorar la interacción del usuario con el sitio web, brindarle respuestas certeras y ayudarle a obtener enlaces a páginas dentro del dominio del sitio web de la universidad. El uso de un asistente virtual permite obtener información más rápido sobre las dependencias, profesores, asignaturas, entre otros. En particular, el asistente que se propone en este trabajo de grado tiene seis partes: interfaz, controlador, analizador gramatical, base de datos con el conocimiento de la universidad, chatbot y motor de búsqueda.

El asistente virtual que se propone está desarrollado en los lenguajes de programación PHP y JavaScript, por lo que es una aplicación web y se utiliza el lenguaje estructurado AIML para crear los patrones de coincidencia y las respuestas que dará. El asistente cuenta con una interfaz que le permite actualizar la información básica de la universidad y otra que le permite sincronizar su conocimiento con la información de programación académica. Además, posee el mecanismo 'Text-to-Speech' que le permite brindar una respuesta auditiva al usuario. El asistente también tiene un módulo de aprendizaje, que le permite adquirir conocimiento a partir de las afirmaciones hechas por el usuario que sostiene una conversación con él.

Para evaluar al asistente virtual se utilizaron tres criterios: interactividad, efectividad y actualidad. Los resultados de las pruebas indican que la precisión de la información en las respuestas obtenidas del asistente es mayor que la obtenida en los resultados del buscador de la universidad.

Contenido

INTRODUCCIÓN	5
1.1 Planteamiento y Formulación del Problema.....	5
1.2 Justificación del Problema.....	6
1.3 Objetivos	6
1.4 Resultados Esperados	7
1.5 Alcance	7
MARCO TEÓRICO	8
2.1 Program O [2]	8
2.2 Motor de Búsquedas [4].....	9
2.3 Analizador Gramatical.....	9
ANTECEDENTES.....	11
3.1 Asistente virtual Lucy [4].....	11
3.2 Asistente virtual Elvira [8]	12
3.3 Asistente virtual Sami [10]	14
3.4 Asistente virtual Cristina [12].....	15
3.5 Planteamiento semántico y pragmático para gestión de diálogos [13].....	15
3.6 Asesor de pregrado UFAQbot [14].....	18
3.7 Marco para la interoperabilidad semántica basada en agentes FRASI [15]	21
DESARROLLO DEL ASISTENTE VIRTUAL.....	24
4.1 Composición del asistente virtual.....	24
4.1.1 Interfaz.....	24
4.1.2 Motor de Procesamiento.....	25
4.1.3 Motor de Búsquedas	27
4.1.4 Analizador Gramatical.....	27
4.2 Procesamiento de oraciones.....	27
4.2.1 Patrones de coincidencia.....	27
4.2.2 Determinación de preguntas	28
4.2.3 Determinación de palabras clave	28
4.3 Conocimiento del Asistente virtual.....	29
4.3.1 Fuente de conocimiento personal	29
4.3.2 Fuente de conocimiento de información básica de la Universidad del Valle	29
4.3.2.1 Estructura de la información	29

4.3.2.2	Conocimiento precargado	31
4.3.2.3	Preguntas que responde el asistente virtual	33
4.3.3	Fuente de conocimiento del sitio web de la Universidad del Valle (www.univalle.edu.co).	33
4.4	Procesos	33
4.4.1	Proceso de cargar el conocimiento del asistente virtual	33
4.4.1.1	Fuente de conocimiento personal	33
4.4.1.2	La base de conocimiento de información básica de la Universidad del Valle (vista en la sección 4.3.2.1).....	34
4.4.1.3	Fuente de conocimiento de la web de la Universidad	36
4.4.2	Proceso de responder a las preguntas de los usuarios.....	36
4.5	Ejemplos de ejecución	37
4.5.1	Información personal.....	37
4.5.2	Base del conocimiento de la Universidad del Valle.....	39
4.5.3	Base de conocimiento del sitio web.....	40
PRUEBAS Y RESULTADOS		43
5.1	Implementación del asistente	43
5.2	Encuestas realizadas	44
5.2.1	Resultados de la evaluación del grado de precisión	44
5.2.2	Resultados de la evaluación del nivel de conocimiento	45
5.2.3	Resultados de la evaluación de actualidad	46
5.2.4	Resultados de la evaluación de interactividad	46
CONCLUSIONES		48
6.1	Conclusiones	48
6.2	Trabajo futuro	48
BIBLIOGRAFÍA		49
ANEXOS		50
Anexo 1. Preguntas que responde el asistente virtual.....		50
Anexo 2. Análisis y diseño		58

CAPÍTULO 1

INTRODUCCIÓN

Los asistentes virtuales son herramientas que surgieron a partir de la combinación de buscadores y chats. Un buscador es un sistema que permite realizar búsquedas de archivos en servidores valiéndose de palabras clave. Por su parte, un chat es un mecanismo de comunicación que permite el intercambio de información y opiniones entre dos o más personas. Un asistente virtual es un derivado de los buscadores que es capaz no solo de dar respuestas a palabras clave, sino que además cuenta con una estructura que le permite entender y responder a oraciones complejas a través del procesamiento de lenguaje natural. En la actualidad se tienen diferentes asistentes virtuales tales como Google Now, Siri y Speaktoit.

Actualmente, la Universidad del Valle cuenta con una página web en la que se puede encontrar información acerca de las dependencias que la componen, los cursos, profesores y estudiantes. La página <http://www.univalle.edu.co> cuenta además con un buscador que utiliza palabras clave y se encuentra disponible en la parte superior derecha que se indica en el recuadro verde de la Figura 1. En este trabajo de grado se propone desarrollar un asistente virtual que permita realizar consultas en lenguaje natural sobre la información de la Universidad del Valle.



Figura 1. Sitio Web de la Universidad del Valle. Tomado de [1]

1.1 Planteamiento y Formulación del Problema

La forma en que las personas estructuran sus consultas en los buscadores limita los resultados obtenidos debido a que la mayoría de usuarios no conocen estrategias de búsquedas o la estrategia utilizada por los buscadores para mostrar los resultados suele ser poco amigable y muchas veces al usuario le toca visitar varias páginas para obtener la respuesta deseada que cumpla los parámetros de precisión y actualidad de la información contenida en ella. En la página web de la Universidad del Valle se cuenta con extensa información de las diferentes facultades que la integran y los diferentes actores que la componen; además de la información concerniente a programas, actividades y/o eventos que se puedan llevar a cabo (información específica de la web de la Universidad), esta información muchas veces no está bien organizada y desactualizada. Por ejemplo, al realizar la consulta “pensum ingeniería sanitaria” en el buscador de la Universidad del Valle, se obtienen resultados referentes al programa académico, a la facultad de ingenierías y otros donde aparecen algunas de las palabras consultadas, la respuesta no es precisa y no resuelve la consulta.

El problema a tratar en esta tesis es la falta de una estrategia en los buscadores que permita interactuar con el usuario para hacer mejores búsquedas. Para ello, en este trabajo de grado se propone como solución desarrollar un asistente virtual chatbot para la web de la universidad. En general, un asistente virtual está compuesto de siete partes: un analizador gramatical, un motor de búsquedas, una base de datos, un módulo de aprendizaje, un chatbot, un controlador y una interfaz.

La pregunta orientada de este trabajo es ¿Cómo diseñar un asistente virtual que permita obtener información específica de la Universidad del Valle por medio de conversaciones expresadas en lenguaje natural?

1.2 Justificación del Problema

Justificación Social

En muchas ocasiones las personas se enfrentan al buscador de la Universidad del Valle queriendo hacer consultas específicas, y obteniendo por lo general resultados insatisfactorios. El asistente virtual que se propone desarrollar ayudará al usuario a encontrar la información que necesita e intentará interpretar lo que el usuario solicitó para obtener mejores resultados en sus búsquedas. También podrá ser de gran utilidad para estudiantes de primer semestre quienes llegan a la universidad sin conocer muchas cosas de ella, en este caso el asistente virtual tendría la capacidad de iniciarlos a este nuevo mundo.

Justificación Académica

Otro factor por el cual sería importante desarrollar este proyecto es que se motive el desarrollo de este tipo de herramientas, que cada día que pasa están aumentando sus fuerzas en la sociedad, y se pueden encontrar en muchos ámbitos con aplicaciones de servicio al cliente, educación, idiomas, marketing, entretenimiento, entre otros.

1.3 Objetivos

1.3.1. Objetivo General

Desarrollar un asistente virtual que permita realizar consultas en lenguaje natural sobre información de la Universidad del Valle.

1.3.2. Objetivos Específicos

1. Construir un Analizador gramatical que permite identificar las palabras clave de la búsqueda.
2. Desarrollar un chatbot que le permita a los usuarios, de manera conversacional, obtener información acerca de la Universidad.
3. Implementar un módulo de aprendizaje del chatbot que permita actualizar su conocimiento por medio del diálogo con el usuario.
4. Implementar una estrategia para la actualización de la información del chatbot.
5. Evaluar al asistente virtual a través de una encuesta que mida su nivel desempeño en tres aspectos: interactividad con el usuario, efectividad de búsquedas, actualidad en la información.

1.4 Resultados Esperados

Objetivo específico	Resultado o producto esperado
1	Implementación analizador gramatical
2	Configuración del chatbot adicionando patrones de coincidencia y conocimiento base de la universidad
3	Implementación del módulo de aprendizaje
4	Desarrollo de funcionalidad de extracción y actualización de información
5	Diseño piloto de la encuesta y documento con los resultados obtenidos por un grupo de estudiantes de diferentes dependencias académicas

1.5 Alcance

El Asistente Virtual, tendrá disponible conocimiento sobre:

- Universidad (Misión, Visión, Símbolos, Normativa, Dependencias, rector)
- Dependencias (descripción general, información de contacto)
- Facultades (ubicación, escuelas, estudios en el exterior, decano)
- Escuelas o Institutos (programas, docentes, eventos, grupos de investigación)
- Profesores (Oficinas, asignaturas que dictan, información de contacto)
- Programas académicos (Descripción, director, facultad a la que pertenece)
- Asignaturas (créditos, profesores que la dictan, si es validable, habilitable, válida para curso de verano)
- Eventos (fechas, descripción, información de contacto)

Para la obtención de este conocimiento se consultará con un grupo de personas que cumplen funciones administrativas en las diferentes dependencias académicas y se extraerá a partir del sitio web de la Universidad del Valle. Además, el asistente virtual cuenta con un módulo de aprendizaje que le permitirá alcanzar este y nuevo conocimiento.

El asistente virtual debe utilizar el mecanismo Text to Speech para dar una respuesta auditiva en caso de que el usuario así lo prefiera. Además, se establece que el asistente no tendrá búsqueda por voz ni incluirá avatar con expresiones faciales.

CAPÍTULO 2

MARCO TEÓRICO

En este capítulo se presentan algunos conceptos básicos relacionados con el diseño, desarrollo y despliegue de un asistente virtual para un sitio web. En la Figura 2 se muestra la composición general de un asistente virtual para un sitio web, donde se puede observar que se compone por un analizador gramatical, un motor de procesamiento (Program O), lenguaje AIML, módulo de aprendizaje y servidor de búsquedas. A continuación se describe cada uno de los elementos que hacen parte de un asistente virtual.

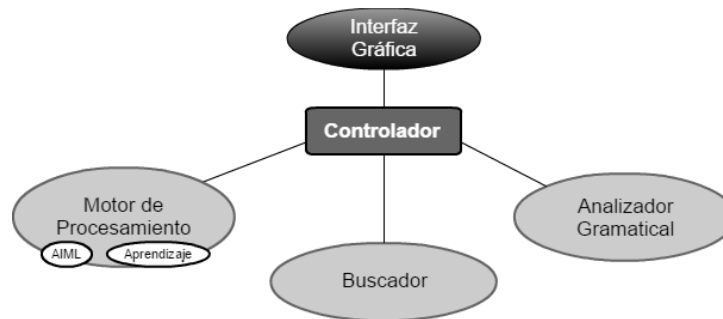


Figura 2. Composición general de un asistente virtual para un sitio web

2.1 Program O [2]

Program O es un motor de procesamiento de lenguaje AIML. Program O consta de un completo panel administrativo de chatbots donde se pueden configurar todas sus opciones, como importar los archivos AIML, modificarlos, configurar palabras mal sonantes y mal escritas, hacer pruebas, etc. Este motor de procesamiento está implementado en PHP con MySQL y puede ser instalado localmente en el servidor. Program O utiliza las librerías predeterminadas de A.L.I.C.E. lo que hace la configuración más rápida, solo basta con arrastrar los archivos AIML al servidor y generar los archivos SQL. En la Figura 3 se muestra el panel de administración de Program O.



Figura 3. Panel de administración de Program O.

Lenguaje AIML (Artificial Intelligence Mark-up Language) [3]

Es un lenguaje compatible con XML. Desde que se creó aproximadamente 15 a 20 años, ha sido utilizado como la herramienta principal para la creación de diferentes chatbots debido a su facilidad de uso y a la diversidad de comunidades que han surgido, mejorando los primeros chatbots y adaptándolos a diferentes ámbitos. Como AIML está basado en XML, tiene una estructura determinada, compuesta por etiquetas que permiten definir respuestas a ciertos patrones de entrada.

Las etiquetas que compone este lenguaje son: Las categorías (<category>), son las unidades fundamentales del conocimiento y contiene al menos un patrón y su respuesta. Los patrones (<pattern>), el texto ingresado por el usuario es comparado con el patrón. Las respuestas (<template>), contienen las salidas que dará el chatbot cuando la entrada coincide con el patrón. A la respuesta se le puede adicionar algunas etiquetas para personalizarla. Respuesta aleatoria (<random>), permite almacenar un conjunto de respuestas para un patrón y se retorna aleatoriamente una de estas. Especificación (<that> o <topic>), dotan de un contexto específico a la categoría, permitiendo dar diferentes respuestas a la misma pregunta dependiendo del contexto. En la Figura 4 se muestra la estructura de un archivo AIML. En este caso cuando la oración ingresada por el usuario coincida con el patrón "WHAT ARE YOU", el asistente virtual responde con el texto que está contenido dentro de la etiqueta <template>.

```
<category>
  <pattern>WHAT ARE YOU</pattern>
  <template>
    <think><set name="topic">Me</set></think>
    I am the latest result in artificial intelligence,
    which can reproduce the capabilities of the human brain
    with greater speed and accuracy.
  </template>
</category>
```

Figura 4. Ejemplo de estructura de un archivo AIML. Tomado de [3]

Además de las etiquetas básicas, Program O permite utilizar otras etiquetas para brindar mayor versatilidad y mejores funcionalidades, estas son: asignación de valores a variables (<set>) y obtención de dichos valores (<get>), etiquetas que pueden ser utilizadas durante la conversación para que el chatbot recuerde alguna información relevante. Condicional (<condition>), la que permite validar las variables almacenadas y dar respuestas dependiendo del contenido de éstas. Y por último, la etiqueta de aprendizaje (<learn>), con la cual se podrá crear nuevos patrones con sus template a partir del diálogo con el usuario.

2.2 Motor de Búsquedas [4]

Un motor de búsqueda es un software diseñado para obtener los recursos digitales de un sitio web, a partir de palabras clave. Estos recursos pueden ser páginas web, documentos, texto, imágenes, entre otros. Cuenta con dos elementos básicos:

- Crawler, es el programa encargado de recopilar toda la información de la web y la añade a una base de datos, este proceso es conocido como indexación.
- Índice, es la base de datos que contiene toda la información recogida por el crawler.

2.3 Analizador Gramatical

Las oraciones o sentencias están compuestas por un conjunto de tokens (palabras o signos de puntuación) los cuales a su vez tienen unos rasgos morfológicos, que determinan las funciones que cumple cada token libre de contexto (Adquiridos a través de un analizador morfológico). Los tageadores utilizan diccionarios etiquetados, conjunto de reglas morfológicas y ortográficas, de los cuales se pueden obtener probabilidades para poder identificar qué funciones morfológicas son más adecuadas para un token en la oración. En la Tabla 1 se muestra las funciones morfológicas de las palabras.

Nombre	Preposición	Poseitivo	Número	Indefinido	Modo
Adverbio	Artículo	Género	Numeral	Tiempo	Verbo
Adjetivo	Conjunción	Demostrativo	Persona	Pronombre	Conjugación

Tabla 1. Vocabulario análisis morfológico. Tomado de [5]

Algunos ejemplos de clases de palabras se describen a continuación [6].

- Nombres o sustantivos, se le designa a los objetos materiales o ideales de la realidad pensándolos con conceptos independientes. Ejemplo: profesor, decano, facultad.
- Adjetivos, se le designan a los objetos materiales o ideales de la realidad pensándolos con conceptos dependientes como calificativos, cualidades, estados, origen o procedencia de los nombres. Ejemplo: nuevo, antiguo.
- Artículos, anuncian al sustantivo. Ejemplo: el, los, las.
- Verbos, expresan comportamiento del sujeto o estados. Ejemplo: dictar, dirigir, presentar.
- Pronombres personales, indica las personas gramaticales sin usar nombres. Ejemplo: yo, tu, nosotros, ellos.
- Preposiciones, une elementos de distintas clases o función. Ejemplo: a, con, durante.
- Conjunciones, sirven como vínculo entre palabras y oraciones. Ejemplo: y, o.

En la Figura 5 se muestra un ejemplo de la salida de un analizador gramatical (Grampal) para la oración “decano de la facultad de ingeniería”, donde se han obtenido tres palabras pertenecientes a la categoría de Nombre (N), las cuales serían utilizadas como palabras clave para realizar la búsqueda en la base de conocimiento.

Grampal
Análisis de: [palabras](#) | [oraciones](#) | [textos](#) | [Generación de formas](#) | [Etiquetario](#) | [Autores](#)

Oración :

(Se utilizará la tabla léxica)

decano de la facultad de ingeniería

decano	categoría	N	lema	DECANO	rasgos	masculino singular
de	categoría	PREP	lema	DE		
la	categoría	ART	lema	EL	rasgos	femenino singular
facultad	categoría	N	lema	FACULTAD	rasgos	femenino singular
de	categoría	PREP	lema	DE		
ingeniería	categoría	N	lema	INGENIERÍA	rasgos	femenino singular

Figura 5. Ejemplo de una respuesta del analizador gramatical. Tomado de [7]

CAPÍTULO 3

ANTECEDENTES

Los asistentes virtuales se han convertido en una estrategia ampliamente usada en ambientes académicos y de fines comerciales. En este capítulo se presentan y describen algunos de los asistentes virtuales desarrollados hasta el momento.

3.1 Asistente virtual Lucy [4]

En la Universidad Complutense de Madrid (UCM) se realizó un trabajo muy similar al que se propone en este trabajo de grado, un asistente virtual para la web de la facultad de informática. El asistente virtual se conoce como Lucy y consta de cinco módulos que se interconectan a través de un controlador que maneja el flujo del sistema. En la Figura 6 se muestra un prototipo del asistente virtual.



Figura 6. Prototipo Lucy, Asistente virtual para la web de la facultad de Informática UCM. Tomado de [4]

Flujo de proceso en Lucy

El controlador recibe la pregunta del usuario, hace la solicitud de respuesta a un chatbot pasándole la pregunta. Si el chatbot tiene la respuesta, se la retorna al controlador y este a su vez a la interfaz para que la despliegue. En caso de que el chatbot no tenga la respuesta, se evalúa si la pregunta contiene tres palabras o menos, de ser así el controlador la trasmite a la base de datos. De encontrarse la respuesta en la base de datos se obtiene y se pasa al controlador y este la envía a la interfaz para que la despliegue. En el caso que la pregunta esté compuesta por más de tres palabras o que no haya devuelto respuesta la base de datos, la pregunta es enviada a un analizador gramatical (GRAMPAL) para obtener las palabras clave de la búsqueda y es reenviada a la base de datos para revisar si las palabras clave devuelven respuesta. Si hay nuevamente negativa de la base de datos, las palabras clave son enviadas a un buscador (SOLR) y se devolvería una lista con las URL's retornadas por el buscador. En caso de que el buscador no retorne nada, en la interfaz se desplegaría un mensaje de error.

Pruebas

Para probar la validez de las respuestas de Lucy, se realizó un cuestionario de 33 preguntas sobre la página Web de la UCM en donde 22 preguntas eran de recurrencia común y 11 no comunes. Estas

preguntas se evaluaron después de la primera y la última versión de Lucy. Para la evaluación se utilizó una escala de 0 a 3 en donde 0 significaba que no se había obtenido respuesta, 1 que la respuesta era poco precisa, 2 cuando la respuesta era medianamente precisa y 3 cuando la respuesta era precisa. Los resultados obtenidos muestran que hubo una mejora entre las dos versiones evaluadas. En la Tabla 2 y la Tabla 3 se muestran los resultados de la encuesta.

	Puntuación máxima posible	Puntuación obtenida	Relación	PORCENTAJE DE CALIDAD
Preguntas comunes	66	43	43 / 66	0.65
Preguntas poco comunes	33	13	13 / 33	0.39
TOTAL	99	56	66 / 99	0.56

Tabla 2. Resultado de la encuesta tras la primera versión de Lucy. Tomado de [4]

	Puntuación máxima posible	Puntuación obtenida	Relación	PORCENTAJE DE CALIDAD
Preguntas comunes	66	53	53 / 66	0.84
Preguntas poco comunes	33	23	23 / 33	0.69
TOTAL	99	76	76 / 99	0.76

Tabla 3. Resultado de la encuesta tras la versión final de Lucy. Tomado de [4]

3.2 Asistente virtual Elvira [8]

En la Universidad de Granada España se realizó un asistente virtual para la web de esta universidad. El asistente virtual se llama Elvira, está dotado de inteligencia artificial que le permite contextualizar, interpretar, reformular e intuir, realizando un procesamiento inteligente de la pregunta y puede responder en lenguaje natural escrito o hablado. Consta de tres módulos principales, Analizador de Lenguaje Natural (NLU, *Natural Language Understan­der*), Administrador de diálogos (DM, *Dialog Manager*), Generador de Comunicación (CG, *Communication Generator*). En la Figura 7 se muestra la arquitectura de Elvira.

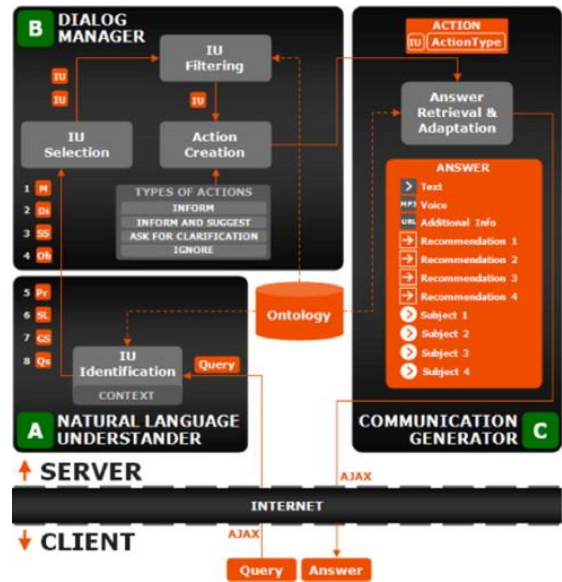


Figura 7. Arquitectura de Elvira. Tomado de [8]

Analizador de Lenguaje Natural (NLU, *Natural Language Understander*)

Es el módulo encargado de realizar la identificación de las unidades de información referidas en la pregunta ingresada por el usuario. Para ello se realizan dos pasos, el primero, reduce el lenguaje de la pregunta a un vocabulario específico para ser gestionado con mayor facilidad, permitiendo corregir algunos errores gramaticales, sustituir palabras con sinónimos o nombres complejos con abreviaturas. En el segundo paso se hace uso de archivos testigos para generar una lista con todas las unidades de información que se referencian en la pregunta e información contextual, utilizada para responder a la pregunta anterior del usuario.

Flujo del proceso

La pregunta se envía desde el cliente al servidor usando la tecnología AJAX (Asynchronous JavaScript And XML), la cual es recibida en el servidor y pasada por el Analizador de Lenguaje Natural (NLU). Luego, el analizador de lenguaje natural identifica las unidades de información a las que se refiere la pregunta y retorna una lista ordenada de acuerdo a las categorías. Esta lista se pasa al Administrador de diálogos (DM) que se encarga de generar las unidades de información que se responderá al usuario, filtrándolas para que sea lo más específica posible, es decir, para poder decidir qué se va a decir sobre ellas y qué tipo de acción se debe tomar. Por último, el Generador de Comunicación (CG), recibe la acción seleccionada y recupera y adapta la respuesta específica que se compone por el texto de respuesta, la URL del archivo de sonido de la respuesta, URL's de páginas web que contiene contenido complementario y una lista de recomendaciones o temas que se puede utilizar para continuar la conversación.

Evaluación y rendimiento

Elvira es accesible desde el 1 de febrero de 2010. En sus primeros 6 meses se registraron más de 35.000 visitas que han sido localizadas desde 93 países diferentes. Una forma de medir la eficacia y la calidad de Elvira es el tiempo que duraron sus visitantes en el sitio. Según los resultados publicados, los visitantes pasan en promedio 04:13 minutos con Elvira y visitan 6 páginas en promedio, evidenciando que realmente utilizan el sistema y encuentran la información que buscaban en un periodo de tiempo apropiado. Otra medida de calidad de las visitas ha sido las declaraciones de agradecimiento de los usuarios, recibiendo más de 600 agradecimientos expresamente por haberles ayudado. En cuanto a la fiabilidad de las respuestas de Elvira, tuvo una tasa de error de comprensión inicial del 25%, la cual se debió a que no se consideraron algunos temas en la base del conocimiento, pero en 6 meses se ha reducido casi al 10%. Elvira está disponible en la página web de la Universidad de Granada como se observa en la Figura 8.

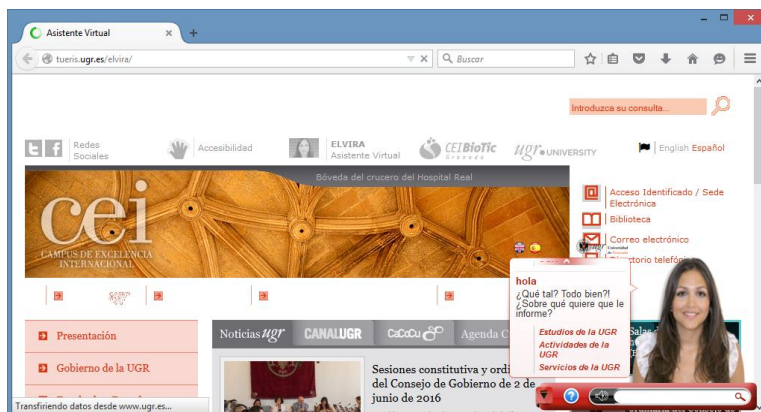


Figura 8. Sitio web de la universidad de granada. Tomado de [9]

3.3 Asistente virtual Sami [10]

Es un prototipo diseñado en la Universidad Nacional Mayor de San Marcos Perú (UNMSM) que está diseñado para dar soporte a temas relacionados con el Aula Virtual de la Universidad. Está dotado de inteligencia artificial y emplea la técnica de Razonamiento Basado en Casos (RBC) [11] para realizar las búsquedas en su base del conocimiento. En la Figura 9 se observa el ciclo de proceso realizado en la técnica de Razonamiento Basado en Casos.

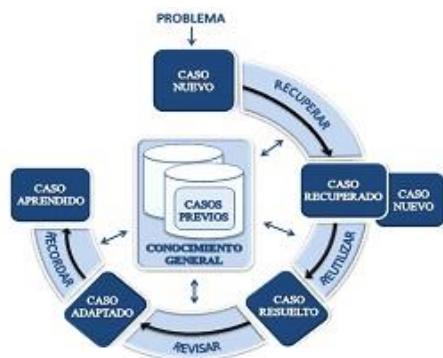


Figura 9. Ciclo del Razonamiento basado en Casos. Tomado de [11]

Flujo del proceso

La consulta escrita en lenguaje natural por el usuario se envía al servidor donde se le realiza un proceso de normalización. Este proceso consta de una sustitución en la que se realizan reemplazos a la entrada para ajustarla a los patrones almacenados en la base del conocimiento buscando solucionar algunos problemas ortográficos, uso de jergas, abreviaciones, excepciones, entre otras. Después de la sustitución se aplican una serie de funciones heurísticas procurando dividir la entrada en varias oraciones más simples y luego se ajusta eliminando los caracteres no permitidos en los patrones. Cuando ya se ha normalizado la entrada el siguiente paso consiste en la búsqueda en la base del conocimiento, la cual se hace por emparejamiento, buscando por profundidad una categoría que contenga el patrón que mejor se empareje con pregunta, retornado la respuesta para que sea desplegada en la interfaz de usuario.

Análisis y evaluación de resultados

Con el fin de evaluar el rendimiento y eficiencia de Sami, las pruebas se realizaron en 4 etapas donde Sami cuenta con diferentes cantidades de conocimiento debido a su evolución de una etapa a la siguiente y se enfrenta a una cantidad de usuarios cada vez mayor. En la Tabla 4 se muestra los resultados obtenidos en cada una de las etapas. Según los resultados obtenidos, se concluye que el porcentaje de preguntas sin ser resueltas se disminuye a medida que se aumenta el conocimiento del asistente virtual.

Etapas	Resumen	Cantidad de Usuarios	Cantidad de Diálogos	Cantidad de Archivos AIML	Cantidad de categoría	Preguntas sin Respuesta	% Preguntas sin respuesta
1	Configurar las variables iniciales del sistema SAMI e ingresarle "conocimiento genérico"	2	10	5	69	45	22,5
2	Retroalimentación de la Base de Conocimientos con la información de la primera etapa.	4	35	15	472	77	19,3
3	Incluir en la base de conocimiento temas relacionados al Aula Virtual y proporcionar a SAMI la capacidad de "reacción al usuario" mediante su interfaz.	15	190	20	544	81	5,4
4	Afinar la base de conocimiento y proporcionarla mayor carga de usuarios al sistema SAMI.	213	368	29	701	90	0,4

Tabla 4. Resultado de pruebas con Sami. Tomado de [10]

3.4 Asistente virtual Cristina [12]

En Hefagra, Hermandad Farmacéutica Granadina, se cuenta con un asistente virtual para su web que fue diseñado e implementado junto con Elvira por Virtual Solutions. Este asistente facilita la visita de los usuarios a la web de la farmacéutica siendo capaz de informarlo de cualquier servicio, producto o departamento de la empresa. Además, es capaz de aprender a partir de las consultas que le hacen y va mejorando en sus respuestas dando cada vez respuestas más precisas. El asistente virtual de Hefagra se muestra en la Figura 10.

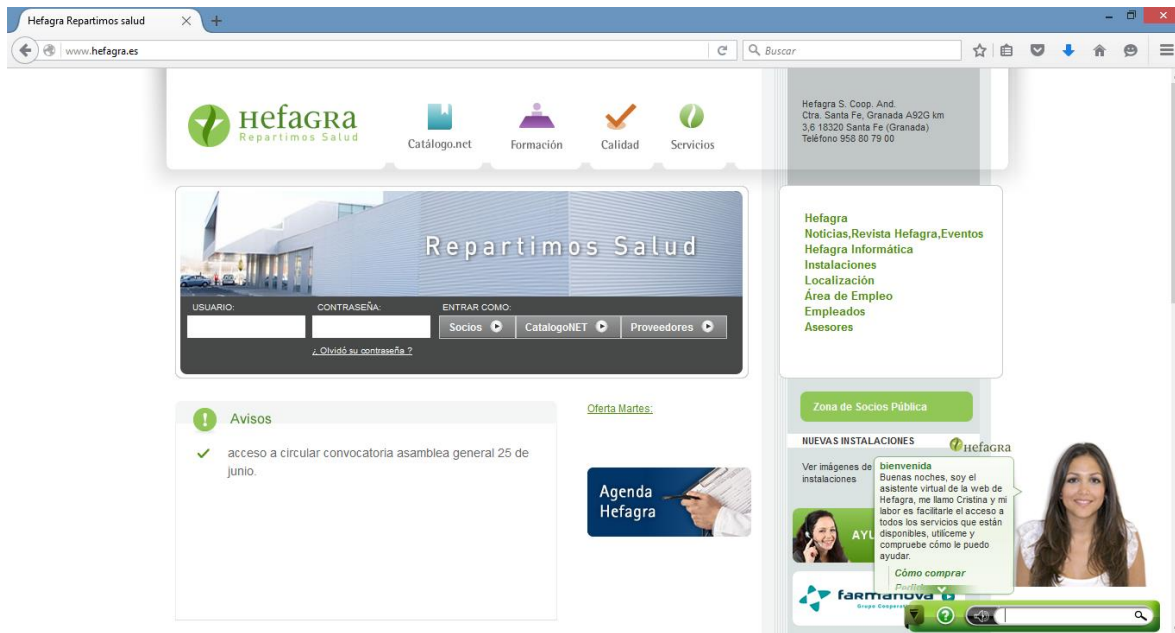


Figura 10. Sitio web de Hefagra. Tomado de [12]

3.5 Planteamiento semántico y pragmático para gestión de diálogos [13]

Es un prototipo diseñado para dar soporte a tiendas electrónicas ayudando a los usuarios a adquirir información sobre los productos y en el proceso de comprarlos. Cuenta con una gramática semántica que permite gestionar el diálogo con el usuario, además de dos mecanismos que le brindan mayor cobertura en la obtención de respuestas con información válida para el usuario.

Con el fin de que el asistente virtual responda a las preguntas del usuario de una forma más humana, éste debe almacenar los detalles proporcionados por el usuario, para no solicitarlos de nuevo. Además, debe ser capaz de anticiparse a las preguntas del usuario o aportar información sobre los productos en cada momento para que no tenga que visitar las páginas que lo contienen. Para cumplir con estas habilidades es necesario:

- Procesar las frases de usuario para extraer:
 - o Información relevante.
 - o Intenciones.
- Gestionar el desarrollo de diálogo manteniendo el hilo argumental lógico con las intenciones del usuario.
- Contar con un conocimiento exhaustivo sobre los productos de la tienda para poder aportar contenido a la conversación.

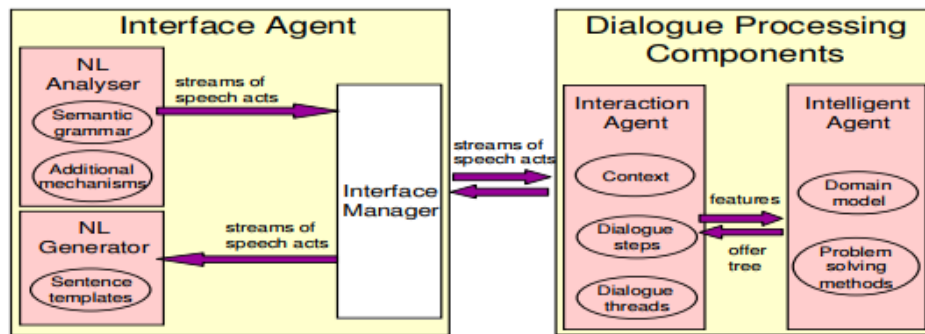


Figura 11. Arquitectura del asistente. Tomada de [13]

La Figura 11 muestra la composición del asistente virtual, el cual posee de dos componentes principales: un agente interfaz, el cual incluye los módulos para la interpretación de las intervenciones del usuario tanto desde el lenguaje natural como desde el uso de la interfaz gráfica; y los componentes de procesamiento del diálogo, los cuales son dos: el agente de interacción, encargado de gestionar el diálogo; y el agente inteligente, con el conocimiento sobre los productos de la tienda.

Flujo de proceso

El usuario interviene en el sistema de dos formas, la primera de ellas, es enviando una oración al asistente y la otra es interactuando con la interfaz gráfica. Esta intervención es recibida por el agente interfaz, el cual extrae la intención y el contenido de la intervención. A partir de la información extraída el agente interacción se encarga de obtener una intención de respuesta, que puede estar completa o requerir ser completada con información del agente inteligente. Ya estando la intención de respuesta completa, el agente interfaz la convierte en respuesta y la muestra al usuario coordinando la interfaz gráfica y el texto en lenguaje natural con el avatar del asistente.

Análisis intencional del lenguaje

Este análisis de intencionalidad permite al asistente identificar y transmitir contenido y la intencionalidad del usuario, para brindar respuestas con información correcta y completa.

Según los autores, se necesitó realizar un estudio previo donde se analizó un corpus artificial que contiene 202 consultas y diez diálogos completos sobre la compra de productos y desde donde se pudo extraer:

- La terminología propia de la tienda electrónica y las relaciones entre los términos y conceptos (ontología), como se observa en la Figura 12.

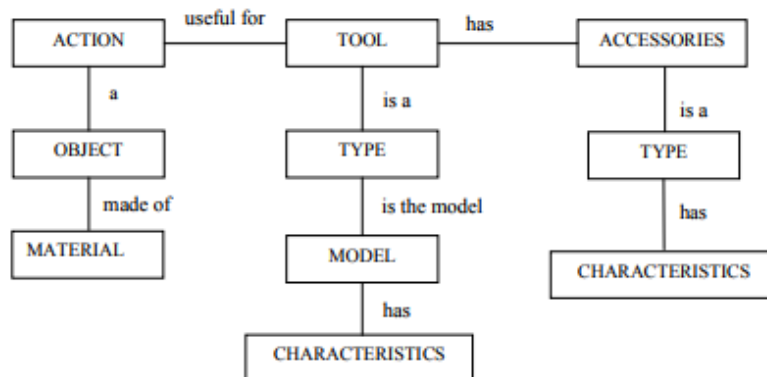


Figura 12. Ontología de la tienda electrónica. Tomada de [13]

- Con base en la ontología se hizo un estudio de las frases para extraer patrones de coincidencia.

De las intervenciones de los usuarios es necesario identificar la intencionalidad y extraer los datos concretos que detallan la conversación, con el fin de codificar este tipo de información se utilizan un conjunto de actos comunicativos:

- Aportando información: cuando se brinda un contenido.
- Solicitando información: cuando se realiza una pregunta.
- Cambiando la iniciativa: cuando uno cede la iniciativa del diálogo al otro.
- Dando órdenes: cuando uno manda a hacer algo al otro.
- Siendo educado: no son necesarios en una conversación en términos informal, aportando algo de carácter emocional.
- Nada para decir: algunas palabras no tienen intención concreta, este tipo de expresiones son carentes de significado.

Este asistente posee una gramática semántica, la cual es una representación del conocimiento que está compuesta por los patrones de coincidencia obtenidos desde el corpus etiquetados con los actos comunicativos, lo que permite realizar un análisis de las intervenciones del usuario para obtener resultados de manera intuitiva y directa, pero que no es completa debido a la riqueza propia del lenguaje, siempre se encontrarán frases desconocidas que no coincida con ningún patrón.

Con el fin de mejorar la cobertura se implementaron dos métodos que sacrifican la precisión:

- Técnica de relajación de las reglas gramaticales: consiste en saltarse u omitir algunas palabras buscando que el resto de palabras encaje en algún patrón, se da respuesta pero pierde precisión.
- Análisis de palabras del dominio: consiste en identificar las palabras en la frase que pertenecen al dominio y retornar información asociada. La cobertura es alta debido a que la gran mayoría de frases del usuario contienen dichas palabras, pero se pierde precisión al no aprovecharse el resto de información.

El asistente responde al usuario inicialmente utilizando la gramática semántica, pero si desde esta no se obtiene respuesta exitosa, se utilizan los resultados generados por los otros dos métodos.

Gestión del diálogo

La información proveniente de la intervención del usuario con el asistente se trata de tres maneras distintas:

- Se extrae información estática, los detalles de la conversación e información contextual, la cual se desecha cuando el diálogo termina.
- Los otros dos tipos de información son de carácter dinámico que indica cómo se desarrolla el diálogo.
 - o Pasos del diálogo: limita los siguientes pasos que el interlocutor puede seguir.
 - o Hilo de la conversación: el usuario y el asistente deben hablar de lo mismo, así el asistente intenta averiguar la intención del usuario para responder adecuadamente y la comunicación sea exitosa.

Generación del lenguaje

El asistente responde acorde a la entrada realizada por el usuario, para ello el agente de interacción analiza cuáles son los pasos del diálogo que se puede tomar y cuál es el hilo de conversación activo para generar los actos comunicativos para la respuesta con contenido que puede ser extraído del contexto o del agente inteligente, los cuales son enviados al agente interfaz para que los presente al usuario en lenguaje natural complementándolo con información de un conjunto de plantillas de frases preparadas para cada intención de manera que no sea repetitiva.

Evaluación

El proceso de evaluación constó de dos fases: la primera solo contenía la gramática semántica y la segunda fase después de agregar los dos mecanismos complementarios.

Para poder realizar una mejor comparación de los resultados obtenidos en ambas fases se elaboraron diez diálogos que contenían 68 intervenciones de usuario. Los resultados obtenidos se clasificaron en cuatro categorías: 'P' El resultado ofrecido es perfecto; 'Par' el resultado devuelto es correcto, pero parcial; 'N' el asistente no es capaz de analizar la frase de entrada; 'W' analiza mal la frase de entrada y responde mal.

Resultado	Eval. 1	Eval. 2	Diferencia
P	30.88 % (21)	41.17 % (28)	+ 10.29 %
Par	29.41 % (20)	36.76 % (25)	+ 7.35 %
N	38.23 % (26)	19.11 % (13)	- 19.12 %
W	4.41 % (3)	2.94 % (2)	- 1.47 %

Tabla 5. Resultado de pruebas. Tomado de [13]

En la Tabla 5 se muestran los resultados obtenidos en las evaluaciones donde se observa un aumento de 10,29% en el porcentaje de obtención de respuestas con información correctas ('P' y 'Par') con lo que se valida que la utilización de los métodos complementarios permitieron adquirir mayor cobertura en las respuestas. También se ha mejorado la precisión debido a que aumentó 7,35% en el porcentaje de respuestas con información correcta y completa ('P'). Hay que acotar que el asistente contó con un corpus artificial muy pequeño y los resultados mejorarían con un corpus real, más completo.

Fue implementado en un prototipo en el que integran el lenguaje natural, interfaz gráfica y avatar para responder al usuario.

3.6 Asesor de pregrado UFAQbot [14]

Es un sistema de pregunta-respuesta orientado a estudiantes universitarios de pregrado, que cumple con el objetivo de brindar información de admisión y sobre los cursos de la universidad.

Diseño

El sistema se basa en el chatbot Alicebot, lo que le permite responder a un conjunto de preguntas predefinidas, además cuenta con un conocimiento propio sobre la universidad, el cual es adicionado fácilmente con ayuda del lenguaje AIML.

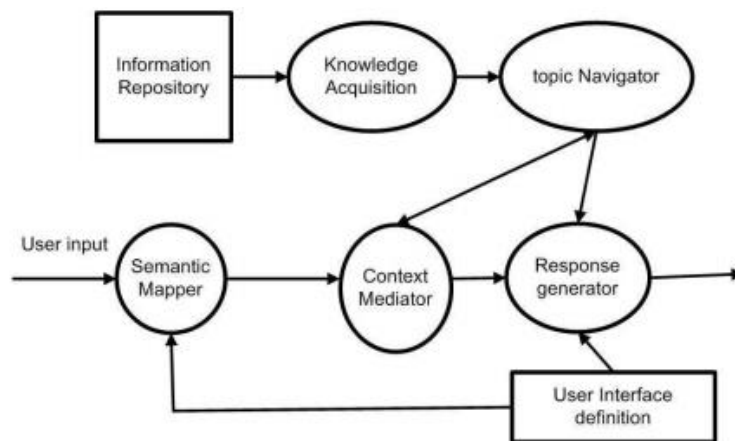


Figura 13. Composición del sistema. Tomada de [14]

Como se muestra la Figura 13 el sistema está compuesto por 5 parte principales: 'Semantic Mapper', es el encargado de mapear la entrada del usuario en elementos semánticos; 'Context Mediator', resuelve los conflictos semánticos; 'Topic Navigator', encargado de navegar por la fuente del conocimiento; 'Response Generator', genera la respuesta en lenguaje natural.

El 'Information Repository' se define como (E, e_0, L) , donde E corresponde a un conjunto de episodios, L a una lista de enlaces, que indica la relación entre dos episodios, y e_0 a un episodio especial correspondiente al mundo exterior. Con el conocimiento representado de esta manera se obtiene la probabilidad de que se dé un episodio a partir de otro, calculado la probabilidad de todos los relacionados con el episodio actual, se genera una respuesta con el de mayor valor, pero el flujo de la conversación depende principalmente de las entradas del usuario.

El 'Topic Navigator' es el encargado de hacer la transición del episodio actual al siguiente, este depende de la determinación de los elementos sintácticos realizada por el 'Semantic Mapper'. Pero si hay elementos sintácticos en conflicto el 'Context Mediator' genera una consulta, que el 'Response Generator' toma para construir un texto en lenguaje natural con el que solicita información adicional al usuario. Una vez se han determinado los elementos sintácticos con éxito, el 'Topic Navigator' obtiene el episodio correspondiente a este.

Cuando se ha completado la transmisión el componente 'Topic Finder' recolecta todos los enlaces que se originaron desde el episodio actual hasta sus vecinos y son enviados para que el 'Response Generator' genere la respuesta apropiada, combinando todas las contenidas en los nodos de información dependiendo de los enlaces y sus valores de preferencia.

Evaluación

El proceso de evaluación consistió en crear tres sistemas que difieren en sus bases conocimiento; BaseBot, es un chatterbot conversacional, utilizó los archivos AIML estándar que contiene 23920 entradas. UFAQBot(Dom eng) el cual contiene el conocimiento del BaseBot y además cuenta con 350 entradas que permiten responder a preguntas frecuentes sobre el sistema de admisión, cursos y la facultad. Este sistema utiliza el diseño presentado. UFAQBot(Rep), contiene las 350 entradas de preguntas frecuentes y 4298 entradas de conversación seleccionadas del BaseBot y modificadas con el diseño, para garantizar que el sistema responda siempre al usuario y para probar si limitando las entradas de conversación mejora el diseño del diálogo.

Se pudo observar en la experimentación que el UFAQBot retornaba el 65% de las consultas, debido a esto se probaron los efectos de los dos grados de conocimiento conversacional junto con el conocimiento del dominio. Para cumplir este propósito se contó con grupo de sujetos que produjeron el número de interacciones que se muestra en la **¡Error! No se encuentra el origen de la referencia.**, allí también se puede observar el desglose de las categorías contenidas en cada sistema.

Interactions for System Type	No of AIML Categories		
	Standard	UFAQBot	Total
3771 BaseBot	23,920	0	23,920
4552 UFAQBOT(Dom_Eng)	23,920	320	24,270
4005 UFAQBOT (rep)	4298	305	4698

Tabla 6. Distribución de categorías AIML. Tomado de [14]

Para medir el desempeño de los dos sistemas chatterbot que contienen conocimiento del dominio, se solicitó a estudiantes de último curso de tecnología de la información que interactuaran con los sistemas y que permanecieran enfocados en los temas del dominio. Se les pidió que hicieran el mismo número de preguntas por cada subtema y marcaran cada respuesta como satisfactoria o insatisfactoria. Las respuestas satisfactorias se basaban en la adecuación del asistente al contexto de la pregunta y las respuestas insatisfactorias si estaba fuera de contexto. Las respuestas se pueden observar en Tabla 7 y Tabla 8.

Conversation Context	Satisfactory	Unsatisfactory
Admission info	70%	30%
Course info	80%	20%
Faculty info	60%	40%

Tabla 7. Evaluación UFAQBot (Rep). Tomado de [14]

Conversation Context	Satisfactory	Unsatisfactory
Admission info	60%	40%
Course info	60%	40%
Faculty info	50%	50%

Tabla 8. Evaluación UFAQBot (Dom-Ing). Tomado de [14]

También se tuvieron en cuenta otras variables para medir los sistemas; Tasa de cambio de tema e índice de corrección.

La tasa de cambio de tema se define como el número de veces que en el diálogo se cambia de tema, dividido por número de conversaciones contenidas en el diálogo con un usuario. Cambio de tema incluye respuestas sin sentido y respuestas no deseadas por el usuario. La Tabla 9 muestra los resultados obtenidos.

System Name	No. of Conversation (session)	TSR
BaseBot	51	68.3%
UFAQBOT (Rep)	51	22.8%
UFAQBOT (Dom_Eng)	51	19.4%

Tabla 9. Evaluación de tasa de cambio de tema. Tomado de [14]

La tasa de corrección se define como el porcentaje de respuestas corregidas por el usuario, dividido por el número de interacciones mecanográficas en el sistema. La Tabla 10 muestra los resultados que se obtuvieron. Además, el usuario realizó una calificación subjetiva de satisfacción general, RS.

System Name	Correction Rate		
	Conversational	Information	RS
BaseBot	17.4%	38.4%	20
UFAQBOT(Dom_Eng)	16.1%	22.8%	60
UFAQBOT (rep)	20.3%	21.1%	70

Tabla 10. Evaluación de tasa de corrección. Tomado de [14]

Se encontró que el conocimiento específico de sistema obtiene mejor resultado para la relación de cambio de tema y la tasa de corrección que un BaseBot. Por otra parte, el UFAQBot (rep) muestra mejores resultados de satisfacción que el UFAQBot (Dom-Ing) lo que demuestra que si se aplica el diseño propuesto al diálogo conversacional, daría sesiones de diálogo máximas sobre un tema específico. Además, se encontró que el conocimiento del dominio justo con el conocimiento conversacional basado en el diseño propuesto produce el mejor resultado, 19.4% y produce el resultado 21.1% para el sistema que tiene diálogo conversacional simple y diálogo específico del tema.

3.7 Marco para la interoperabilidad semántica basada en agentes FRASI [15]

Es una implementación de mejora semántica orientado a un dominio para un sistema de pregunta-respuesta basado en tecnología chatbot. Cuenta con dos soluciones que le permiten la construcción de su conocimiento base y la obtención de respuestas, el primero es mediante la utilización de una ontología, para construir respuestas dinámicas como resultado de un proceso de inferencia sobre el dominio y rellenar automáticamente el conocimiento base con oraciones que se puedan derivar de la ontología, y el segundo es pre procesando las oraciones de usuario y reducirlas a una expresión más sencilla.

Diseño

La Figura 14 muestra la composición del sistema, donde se aprecia cómo la tecnología Alice estándar puede ser mejorada mediante la inclusión de nuevos módulos. El motor de diálogo se comunica con dos módulos principales, llamados E-SRAI (Enhanced Symbolic Reduction Module) y CYD a través de coincidencia de categorías AIML con etiquetas definidas.

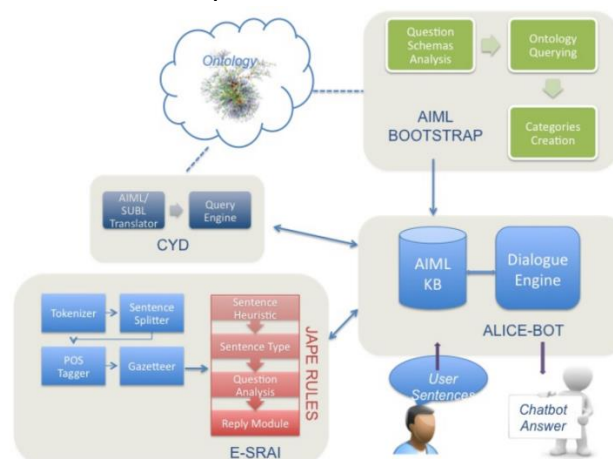


Figura 14. Composición de FRASI. Tomada de [15]

El E-SRAY, permite realizar un análisis más eficiente de cada frase del usuario, comprendiendo el tipo de pregunta y extrayendo variables particulares mediante reglas sintácticas, semánticas y pragmáticas, con el fin de reducir la frase ingresada a una estructura más sencilla. Utiliza las herramientas proporcionadas por OpenGate para definir las reglas de análisis. Debido a la complejidad de la identificación y el reconocimiento de los esquemas del lenguaje semántico, la definición del módulo de análisis lingüístico requiere la restricción del estudio de caso a un dominio limitado del conocimiento.

El CYD extiende el conocimiento del chatbot, permitiendo generar respuestas a las consultas semánticas dirigidas a la ontología del sentido común CYC. El lenguaje AIML es ampliado con etiquetas específicas que permiten consultar la ontología desde las reglas de base del conocimiento para producir la respuesta más adecuada para las preguntas del usuario. La posibilidad de explotar una ontología amplia, de sentido común, hace que el diálogo sea más fluido y reduce el número de respuestas por defecto necesarias en los chatbots tradicionales para llenar sus lagunas "culturales".

Estos dos módulos se ejecutan cada vez que el usuario ingresa una frase, primero la analiza con el E-SRAI y las respuestas se obtiene mediante la base conocimiento del chatbot KB, la cual puede activar la invocación del CYD para consultar la ontología.

Cuenta además con un módulo que se ejecuta fuera de línea llamado AIML Bootstrap, que tiene como objetivo poblar automáticamente la base de conocimiento del chatbot, a través de un proceso de consulta de una ontología por medio de un conjunto de esquemas de pregunta-respuesta predefinidas y creando un conjunto de categorías AIML, llenado patrones y esquemas de plantillas con los conceptos o relaciones derivadas de la ontología. Esto le permite al chatbot adaptarse rápidamente a diferentes dominios de aplicación. Este proceso se repite cada vez que se actualice la ontología para mantenerla sincronizada con la base de conocimiento.

Interacción entre módulos

Un ejemplo de interacción entre módulos durante un diálogo se muestra en la Figura 15. Allí se observa dos sentencias introducidas por el usuario; la primera es reconocida erróneamente como una afirmación, el módulo E-SRAI pre procesa la sentencia y da la respuesta al módulo AIML que identifica una coincidencia con los patrones almacenados y ejecuta las sentencias escritas en la sección de plantillas para construir las respuestas y comunica a la ontología para que almacene la información de contexto. La segunda sentencia enviada por el usuario es correctamente reconocida por el módulo E-SRAI como una pregunta formal para una definición y la traduce a una forma más sencilla. El motor AIML encuentra un patrón válido para la traducción y luego se consulta la ontología para generar la respuesta adecuada al usuario.

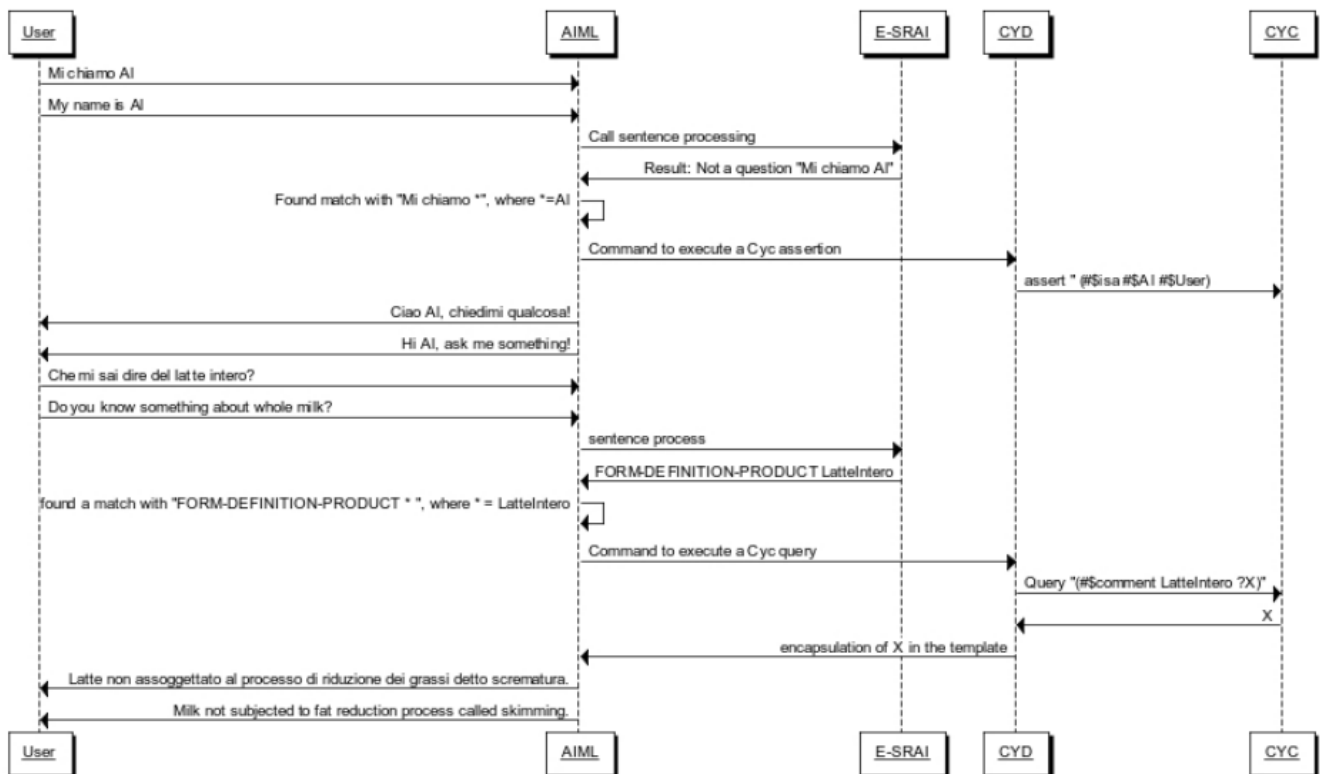


Figura 15. Diagrama de secuencia, comunicaci3n entre m3dulos. Tomada de [15]

CAPÍTULO 4

DESARROLLO DEL ASISTENTE VIRTUAL

4.1 Composición del asistente virtual

En este trabajo se desarrolló un asistente virtual que consta de seis partes: interfaz, es la página que permitirá la comunicación entre el usuario y el asistente virtual; controlador, es el encargado de gestionar la integración de los componentes del asistente virtual; analizador gramatical, permite descomponer oraciones, con el fin de determinar las palabras clave; chatbot, ejecuta la parte conversacional y de aprendizaje del asistente virtual; base de datos, allí se almacena el conocimiento base de la Universidad del Valle que el asistente conoce. Además, sirve de repositorio de los archivos AIML del chatbot; motor de búsqueda, permite hacer búsquedas a las cuales no se obtuvieron respuestas por medio del chatbot, ni de la base de datos. En la Figura 16 se muestra como está compuesto el asistente virtual.

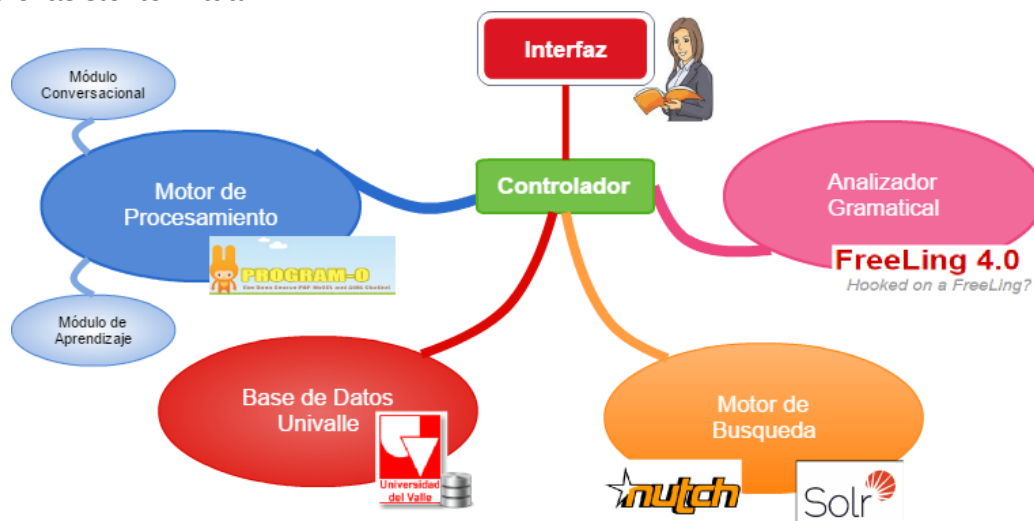


Figura 16. Composición del asistente virtual.

A continuación se describen cada una de las partes que lo conforman.

4.1.1 Interfaz

La interfaz cuenta con dos elementos:

- 1) Asistente virtual: se compone de un área donde se muestra la última iteración de la conversación (lo que dice el usuario y lo que responde el asistente virtual) también se encuentra un campo de inserción de texto en el que los usuarios pueden ingresar las preguntas que desean hacer, las cuales pueden ser enviadas presionando la tecla enter o dando clic en el botón de enviar. Además, cuenta con un botón 'Mostrar Log', el cual muestra el histórico de la conversación del usuario.
- 2) Un iframe: permite desplegar las páginas web retornadas por el asistente virtual y el histórico de la conversación. Hay que tener presente que por políticas de seguridad algunas páginas no pueden ser desplegadas en un iframe, pero el asistente virtual muestra los enlaces para que se puedan abrir en una nueva pestaña del navegador.

En la Figura 17 se muestra la interfaz gráfica y la ubicación de los elementos que la componen.

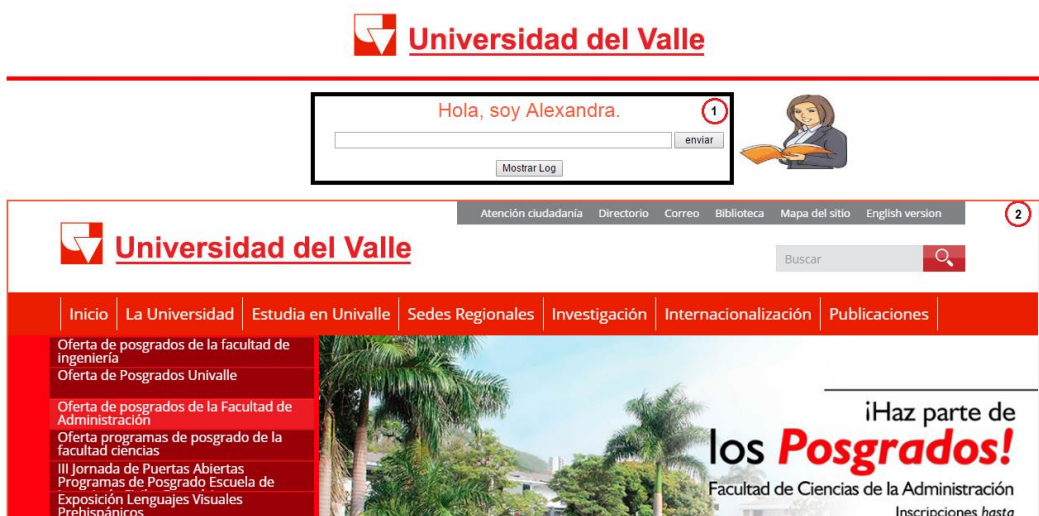


Figura 17. Interfaz asistente virtual.

4.1.2 Motor de Procesamiento

El motor de procesamiento es el módulo principal del asistente virtual. Este módulo se encarga de gestionar la parte conversacional, identificación de las preguntas y respuestas, y de la gestión del aprendizaje del asistente. Para su implementación se utilizó Program O, una aplicación web que brinda un completo panel administrativo para la gestión de los chatbots.

4.1.2.1. Módulo conversacional

El módulo conversacional se encarga de gestionar las preguntas y respuestas, haciendo uso de los patrones de coincidencia (ver sección 4.2.1) y de la información contenida en las fuentes de conocimiento (ver sección 4.3).

4.1.2.2. Módulo de Aprendizaje

El módulo de aprendizaje le permite al asistente virtual aprender a partir de las conversaciones que tenga con los usuarios. Esto se hace gracias a la utilización de la etiqueta '<learn>', a patrones de coincidencia y a la opción nueva de Program O, llamada 'Search/Edit AIML User', donde un usuario administrador se encarga de confirmar el conocimiento a adquirir para evitar inconsistencias con el conocimiento previo del asistente virtual.

Los patrones que se utilizaron en el módulo de aprendizaje son los siguientes:

- **Conceptos**
 - * es *
 - * son *
- **Ubicación**
 - * está *
 - * queda *
 - * se * en *
- **Composición**
 - * está conformada *
 - * está compuesta *
- * tiene * áreas académicas
- * de * son *
- **Función**
 - * funciona *
 - * es para *
 - * hacen *
 - * trabaja en *
 - * tiene por objeto *
- **Pertenece**
 - * pertenece * de *

El aprendizaje es posible a través de las conversaciones con el usuario, específicamente a través de las afirmaciones que hace. A continuación se muestran algunos ejemplos de afirmaciones que se pueden formar:

- El acuerdo 009 de 1997 **es** el reglamento estudiantil.
- El laboratorio **es** un lugar dotado de los medios necesarios para realizar investigaciones, experimentos, prácticas y trabajos de carácter científico, tecnológico o técnico.
- El centro de investigaciones territorio construcción y espacio **está** en el edificio 316, oficina 3003 en Meléndez.
- La facultad de ciencias naturales y exactas **está compuesta** por los departamentos de Biología, Física, Matemáticas y Química.
- El instituto de educación y pedagogía **tiene** 11 áreas académicas.
- El grupo de investigación Gravitación **tiene por objeto** la formación investigativa de sus integrantes, generando el desarrollo de sus capacidades y la difusión del conocimiento científico en áreas estratégicas de frontera en física.
- El restaurante de la Universidad **funciona** de 11:40 hasta las 2:30.
- El área de bienestar universitario **trabaja en** crear, mantener y consolidar un entorno universitario caracterizado por un ambiente intelectual, ético y estético de la comunidad universitaria.
- Las áreas del instituto de educación y pedagogía son Centro deportivo universitario, área de educación matemática y área de educación física y deportes
- El área de educación matemática **pertenece** al instituto **de** educación y pedagogía.

Las afirmaciones son convertidas a patrones que posteriormente permiten a nuevos usuarios obtener el conocimiento aprendido. A continuación se muestran algunos ejemplos de los patrones y respuestas que se generan:

- ¿Qué **es** el acuerdo 009 de 1997?
 - R/. **Es** el reglamento estudiantil.
- ¿Qué **es** un laboratorio?
 - R/. **Es** un lugar dotado de los medios necesarios para realizar investigaciones, experimentos, prácticas y trabajos de carácter científico, tecnológico o técnico.
- ¿Dónde **está** el centro de investigaciones territorio construcción y espacio?
 - R/. **Está** en el edificio 316, oficina 3003 en Meléndez.
- ¿Cómo **está compuesta** la facultad de ciencias naturales y exactas?
 - R/. **Está compuesta** por los departamentos de Biología, Física, Matemáticas y Química.
- ¿Cuántos grupos de investigación **tiene** el departamento de Física?
 - R/. **Tiene** 10 grupos de investigación.
- ¿Qué función tiene el grupo de investigación Gravitación?
 - R/. Trabajan en la formación investigativa de sus integrantes, generando el desarrollo de sus capacidades y la difusión del conocimiento científico en áreas estratégicas de frontera en física.
- ¿En que trabaja el área de bienestar universitario?
 - R/. **Trabaja en** crear, mantener y consolidar un entorno universitario caracterizado por un ambiente intelectual, ético y estético de la comunidad universitaria.
- ¿En qué horario **funciona** el restaurante de la Universidad?

- R/. **Funciona** de 11:40 hasta las 2:30.
- ¿A qué instituto pertenece el área de educación matemática?
 - R/. **Pertenece** al instituto **de** educación y pedagogía.

4.1.3 Motor de Búsquedas

El motor de búsquedas se integra con el asistente virtual con el fin de obtener contenido que dé respuesta a las preguntas que el usuario haga sobre la Universidad del Valle y él no pueda responder por no contar con estas respuestas en su conocimiento base sobre la Universidad. En su implementación, se hizo uso de la tecnología Lucene, que es un producto desarrollado por Apache, más precisamente se utilizaron las herramientas Apache Nutch y Apache Solr, las cuales son herramientas que se complementan a la perfección, el primero hace las veces de crawler y el segundo se encarga de la gestión de búsquedas. Las búsquedas se realizan con las palabras clave determinadas de las preguntas realizadas por el usuario, estas búsquedas se realizan en los índices de url, título y contenido, obtenidos por medio del crawler Apache Nutch, buscando obtener resultados más precisos a lo que el usuario desea.

4.1.4 Analizador Gramatical

El analizador gramatical se implementó con la librería Freeling. La cual cuenta con una amplia gama de herramientas para hacer análisis morfológico y tagger. Estas dos herramientas de Freeling se utilizaron en el procesamiento de las oraciones enviadas por el usuario para poder determinar si la oración corresponde a una pregunta o no (ver sección 4.2.2) y para determinar las palabras clave en la oración (ver sección 4.2.3).

4.2 Procesamiento de oraciones

El asistente virtual cuenta con tres mecanismos que le permiten identificar lo que el usuario desea decir y lo que debe responderle. Los mecanismos son los patrones de coincidencia, determinación de preguntas y determinación de palabras clave, los cuales se describen a continuación.

4.2.1 Patrones de coincidencia

Los patrones son cadenas de caracteres en los cuales se utilizan comodines (*) que brindan gran flexibilidad al poder ser sustituidos por una palabra o un conjunto de palabras. La idea de utilizar patrones es que permite hacer agrupaciones de oraciones y así el asistente virtual responda lo mismo a preguntas similares. El módulo encargado de realizar la coincidencia de patrones es el chatbot. A continuación se muestran algunos ejemplos:

Patrón 1.

*** PELICULA * GUSTA**

Este patrón puede ser utilizado para dar respuesta a varias preguntas del usuario:

¿Cuál es la película que más te gusta?

¿Qué película te gusta?

A pesar de que en este caso se tiene un patrón que puede resultar amplio, se busca obtener un patrón para el cual las expresiones del usuario encajen y tengan sentido para el asistente.

¿Dónde está la película que me gusta?

Como se puede ver ésta es una pregunta válida, pero la cual no tendría sentido que el usuario le haga al asistente.

Patrón 2.

CUÁL ES EL EMAIL DE *

Este patrón puede ser utilizado para dar respuesta a varias preguntas del usuario:

¿Cuál es el email de la Facultad de Ingenierías?

¿Cuál es el email de Oscar Bedoya?

¿Cuál es el email del Pregrado de Matemáticas?

La respuesta del asistente virtual en este caso varía dependiendo de la pregunta, ya que este patrón ayuda a identificar lo que el usuario está preguntando y devuelve la respuesta desde la base de conocimiento de la Universidad del Valle.

4.2.2 Determinación de preguntas

Para determinar si lo que dice el usuario es una pregunta, se busca en la oración palabras que cumplan con la función interrogativa. El tageador identifica dos tipos de funciones gramaticales (DT y PT) que pueden ser utilizadas en una oración que se refiere a una pregunta. Estas funciones son identificadas por Freeling bajo las siguientes etiquetas:

- DT: Determinante interrogativo.

- Qué
- Cuánto

- PT: Pronombre interrogativo.

- Quién
- Cuál
- Dónde
- Cuándo
- Cómo

4.2.3 Determinación de palabras clave

Las palabras clave son aquellas sobre las que se quiere enfocar la búsqueda. El tageador es el encargado de identificar las funciones que cumple cada palabra en la oración enviada por usuario, extrayéndose aquellas palabras que tienen funciones de sujetos y que son identificadas por:

NC: Nombres comunes

NP: Nombres propios

A continuación se muestran ejemplos de preguntas hechas al asistente y qué palabras son identificadas como clave y serán de gran utilidad en la búsqueda.

¿Cuál es la **misión** de la **Escuela** de **Ingeniería** de **Sistemas** y **Computación**?

¿Cuál es el **teléfono** de **Oscar Bedoya**?

¿Cuáles son los **laboratorios** de **Arquitectura**?

4.3 Conocimiento del Asistente virtual

El asistente virtual cuenta con tres fuentes de conocimiento, las cuales le permite responder preguntas de tipo personal, información básica de la Universidad del Valle e información contenida en su sitio web www.univalle.edu.co.

4.3.1 Fuente de conocimiento personal

Esta fuente de conocimiento permite tener una conversación más natural entre el usuario y el asistente como si se estuviera conversando con otra persona. El agente puede responder aproximadamente 130 preguntas. A continuación, se encuentran algunos ejemplos de preguntas que responde el asistente virtual:

- | | |
|---|----------------------------------|
| 1. ¿Cómo estás? | 12. ¿Cuál es tu comida favorita? |
| 2. ¿Tienes una casa? | 13. ¿Cuál es tu nombre? |
| 3. ¿Tiene mascotas? | 14. ¿Cómo te llamas? |
| 4. ¿Cómo se llama tu mascota? | 15. ¿Cuál es tu apellido? |
| 5. ¿Quieres salir conmigo? | 16. ¿Te gusta bailar? |
| 6. ¿Dónde vives? | 17. ¿Qué te gusta bailar? |
| 7. ¿Qué haces ahora? | 18. ¿Usted tiene hermanos? |
| 8. ¿Con quién vives? | 19. ¿Cómo estás? |
| 9. ¿Dónde trabajas? | 20. ¿Eres alta? |
| 10. ¿Desde hace cuánto trabaja? | 21. ¿Eres Hermosa? |
| 11. ¿Cuál es el color que más te gusta? | 22. Y muchas otras... |

4.3.2 Fuente de conocimiento de información básica de la Universidad del Valle

Permite dar respuestas a preguntas comunes sobre la Universidad del Valle, sobre sus dependencias y sus programas académicos.

4.3.2.1 Estructura de la información

El agente cuenta con información del área académica de la Universidad, tales como información de sus dependencias, programas académicos, personal, asignaturas, horarios de clases y de atención, e información sobre las áreas de investigación. A continuación se detalla la estructura de la información con que el agente dispone:

- Dependencias:
 - Nombre
 - Abreviatura
 - Tipo (Facultad, escuela, departamento, instituto)
 - E-mail
 - Teléfonos
 - Misión y visión
 - Página web
 - Ubicación
- Programas académicos:
 - Nombre
 - Abreviatura
 - Dependencia a la que pertenece

- Título que otorga
- E-mail
- Duración
- Créditos
- Jornada en que se dicta
- Perfil del egresado
- Pensum
- Lista de asignaturas
- Ubicación
- Sitio web
- Personal:
 - Nombres
 - Apellidos
 - Ocupación
 - Tipo (F -> funcionario, P->profesor, C->hora cátedra)
 - Dependencia en la que trabaja
 - Programa académico al que perteneces (Solo funcionarios)
 - E-mail
 - Teléfono
 - Oficina
- Asignaturas
 - Código
 - Nombre
 - Abreviatura
 - Dependencia que la dicta
 - Tipo (Básica y Electiva)
 - Créditos
 - Habilitable
 - Validable
 - Anual
 - Página Web o Descripción
- Prerrequisitos
 - Programa académico y Resolución
 - Asignatura
 - Prerrequisitos
- Clases
 - Profesor
 - Asignatura
 - Grupo
 - Horario
 - Semestre que se dicta. Ejemplo (2017 - 1)
- Disponibilidad
 - Profesor

- Oficina
- Horario de lunes a sábado
- Semestre para el cual es válida esta disponibilidad.
- Investigación
 - Tipo
 - Nombre
 - Abreviatura
 - Director
 - Contacto
 - Dependencia a la que pertenece
 - Grupo investigación al que pertenece
 - Líneas de investigación
 - Página web
 - Ubicación

4.3.2.2 Conocimiento precargado

El agente cuenta con información de diferentes dependencias de la Universidad del Valle, estas son:

- Facultad de Ingenierías
 - Escuela de ingeniería de sistemas y computación
 - Escuela de ingeniería sanitaria y ambiental
- Facultad de ciencias de la administración
- Facultad de ciencias naturales y exactas
 - Departamento de matemáticas
- Instituto de psicología

De estas dependencias se puede encontrar la información descritas anteriormente con algunas excepciones:

- Facultad de ciencias de la administración
 - E-mail
- Facultad de ciencias naturales y exactas
 - Misión y visión de la facultad

El agente cuenta con la información de los programas académicos de pregrado de las dependencias indicadas anteriormente. Los programas académicos registrados son:

- Administración de empresas
- Contaduría pública
- Comercio exterior
- Ingeniería de sistemas
- Tecnología en sistemas de información
- Ingeniería Sanitaria y Ambiental
- Tecnología en Ecología y Manejo Ambiental
- Tecnología en Manejo y Conservación de Suelos y Aguas
- Ingeniería Agrícola

- Psicología
- Matemáticas

Información faltante de los programas académicos:

- Facultad de Ingenierías
 - Escuela de ingeniería sanitaria y ambiental
 - Ingeniería Sanitaria y Ambiental
 - E-mail
 - Tecnología en Ecología y Manejo Ambiental
 - registro calificado
 - resolución
 - lista de asignaturas
 - Tecnología en Manejo y Conservación de Suelos y Aguas
 - registro calificado
 - resolución
 - lista de asignaturas
 - Ingeniería Agrícola
 - lista de asignaturas

El asistente contiene la información de contacto del personal de las dependencias antes mencionadas:

- Profesores nombrados
- Profesores de hora cátedra (hubo problemas para obtener la información de contacto como teléfono, oficina y/o email)
- Funcionarios

Además de información sobre la disponibilidad de los profesores de planta, que pertenecen a las siguientes dependencias:

- Escuela de ingeniería de sistemas y computación
- Facultad de ciencias de la administración

Se cuenta con información de las diferentes asignaturas, sus horarios de clase y sus prerrequisitos, pero no fue posible obtener el contenido programático de todas las asignaturas por políticas de las dependencias. En esos casos se maneja una breve descripción otorgada por los sistemas de Univalle.

También se adquirió conocimiento sobre las áreas de investigación de cada dependencia, obteniendo gran parte de información sobre:

- Centros de investigación
- Grupos de investigación
- Líneas de investigación
- Laboratorios

4.3.2.3 Preguntas que responde el asistente virtual

Algunas de las preguntas que responde el asistente son las siguientes:

- Abreviatura
 - Pregunta
 - ¿Qué significa la abreviatura **ABREV**?
 - Respuestas
 - Nombre dependencia
 - Nombre asignatura
 - Nombre programa académico
 - Nombre centro de investigación
 - Nombre grupo de investigación
 - Nombre laboratorio
 - Nombre línea de investigación
- Profesor que dicta una asignatura el semestre actual
 - Pregunta
 - ¿Quién dicta **ASIG** este semestre?
 - Respuesta
 - Nombre profesores

En el Anexo 1 se puede observar una lista de las preguntas que responde el asistente virtual sobre el conocimiento de la Universidad del Valle.

4.3.3 Fuente de conocimiento del sitio web de la Universidad del Valle (www.univalle.edu.co).

Permite al asistente dar información de la Universidad del Valle que pueda ser obtenida a partir de su sitio web, brindándole al usuario los enlaces para que pueda acceder al contenido requerido.

4.4 Procesos

El asistente virtual consta de dos procesos para su funcionamiento, el proceso de cargar el conocimiento y el proceso de responder las preguntas.

4.4.1 Proceso de cargar el conocimiento del asistente virtual

La información del asistente virtual se puede cargar a través de diferentes métodos dependiendo del tipo de información.

4.4.1.1 Fuente de conocimiento personal

La información personal se adiciona al asistente virtual haciendo uso de la opción de Program O llamada 'Upload AIML' donde se carga la información desde un archivo estructurado AIML.

En la Figura 18, se muestra una sección de un archivo AIML donde se configuran las respuestas que el asistente virtual da a preguntas sobre sus gustos de películas y actores. Debido a que varias preguntas pueden dar una misma respuesta, se hace uso de patrones base como 'ACTORFAV' y 'PELICULAFV'.

```

<category>
  <pattern>P * PELICULA * GUSTA</pattern>
  <template>
    <srxi>PELICULAFAV</srxi>
  </template>
</category>
<category>
  <pattern>P * ACTOR * GUSTA</pattern>
  <template>
    <srxi>ACTORFAV</srxi>
  </template>
</category>
<category>
  <pattern>ACTORFAV</pattern>
  <template>26|<random><li>Es <bot name="favoriteactor"/></li>
  <li>Mi actor favorito es <bot name="favoriteactor"/></li>
  <li>Me encanta <bot name="favoriteactor"/></li>
  <li>Me gusta mucho <bot name="favoriteactor"/></li>
  <li>Amo a <bot name="favoriteactor"/></li></random>
  </template>
</category>
<category>
  <pattern>PELICULAFAV</pattern>
  <template>31|<random><li>La película que mas me gusta es <bot name="favoritemovie"/>
  </li><li>Mi película favorita es <bot name="favoritemovie"/> es muy divertida</li>
  <li>Me encanta <bot name="favoritemovie"/></li></random>
  </template>
</category>

```

Figura 18. Preguntas sobre gustos.

4.4.1.2 La base de conocimiento de información básica de la Universidad del Valle (vista en la sección 4.3.2.1)

Esta información se carga a través de una opción adicionada a Program O llamada 'Add new knowledge', la cual permite desde unas plantillas de Excel cargar este conocimiento, extrayendo los datos almacenados en cada fila que corresponderían a un registro. Esta información se almacena en una base de datos. El proceso se muestra en la Figura 19.

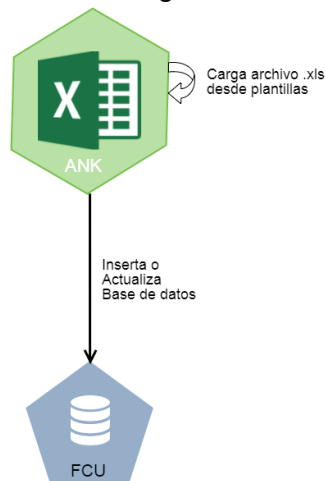


Figura 19. Proceso de carga de conocimiento básico de la Universidad del Valle

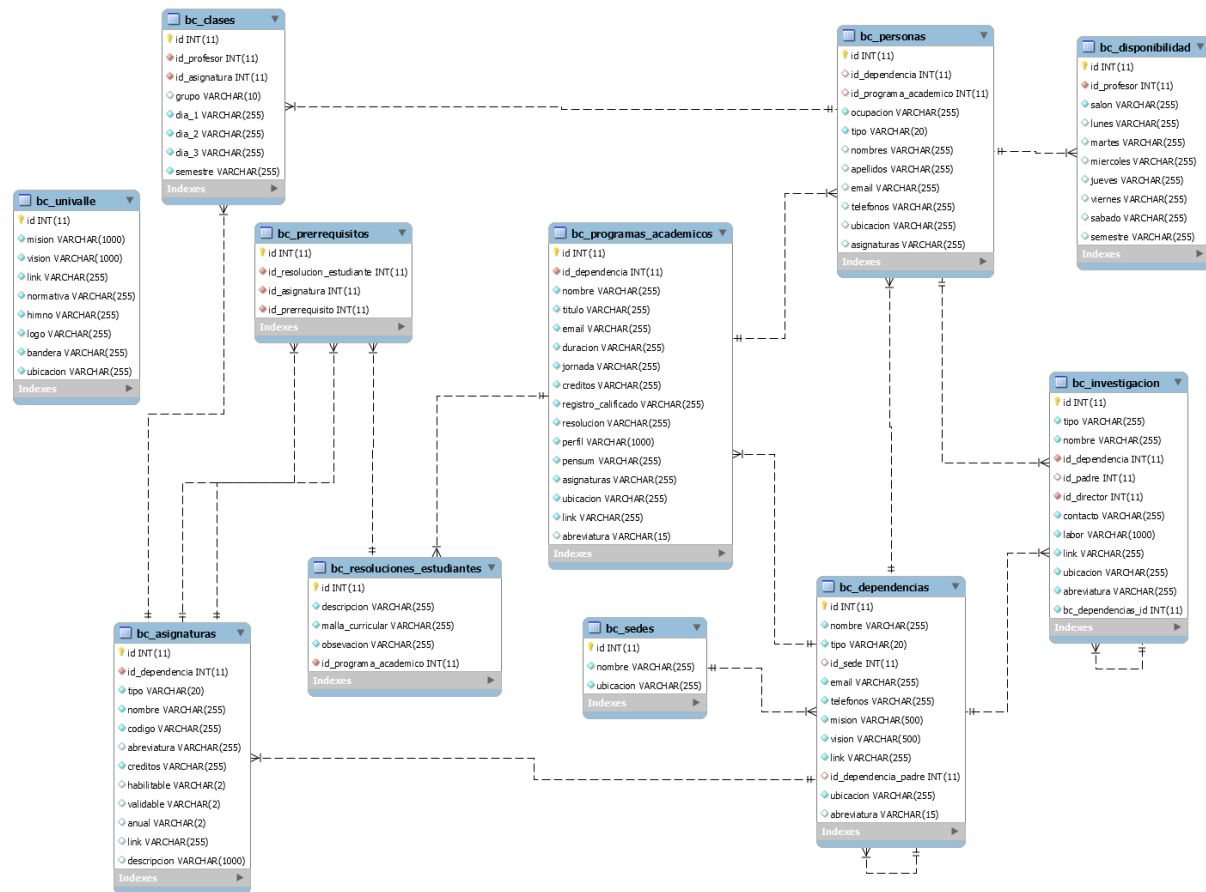


Figura 20. Base de datos de conocimiento del asistente virtual sobre la Universidad del Valle.

La Figura 20, representa el diseño de la base de datos que contiene el conocimiento de la Universidad del Valle. Se puede observar que cuenta con 11 tablas que contienen información de diferentes entidades que están presentes en la Universidad: *bc_univalle*, contiene información básica de la Universidad; *bc_sedes*, contiene información de las sedes; *bc_dependencias*, se almacena información de las facultades, institutos, escuelas y departamentos; *bc_programas_academicos*, contiene la información sobre los programas académicos de cada dependencia; *bc_resoluciones_estudiantes*, contiene las resoluciones y los pensum de los programas académicos; *bc_asignaturas*, almacena la información de las asignaturas; *bc_prerrequisitos*, aquí queda registrado que asignaturas tiene que haber aprobado el estudiante para poder ver otra ; *bc_clases*, registra los horarios de clase de cada asignatura; *bc_personas*, guarda la información de contacto de todas las personas que pertenecen a las dependencias (profesores y personal administrativo); *bc_disponibilidad*, almacena los horarios de atención de los profesores; y *bc_investigacion*, registra la información básica de las áreas de investigación de las dependencias.

Sincronización de programación académica (APS)

La Universidad del Valle cuenta con el sistema de registro académico el cual puede ser accedido por medio de un formulario en la cual los usuarios diligencian información como: código de la asignatura o selecciona una unidad académica, con esta información el sistema retorna los horarios de clase del semestre actual de una forma ordenada y estructurada. Con la ayuda de la librería CURL de Php, este trabajo puede realizarse omitiendo la función del formulario y enviando los parámetros necesarios por POST a la página del sistema, y por último se recorre de manera secuencial la

respuesta obtenida extrayendo los horarios de clase de las asignaturas del semestre actual y se actualizan en la base del conocimiento del asistente virtual. Para ello se adicionó una opción a Program O llamada 'Academic programming synchronization' donde se escoge la unidad académica que se desea sincronizar.

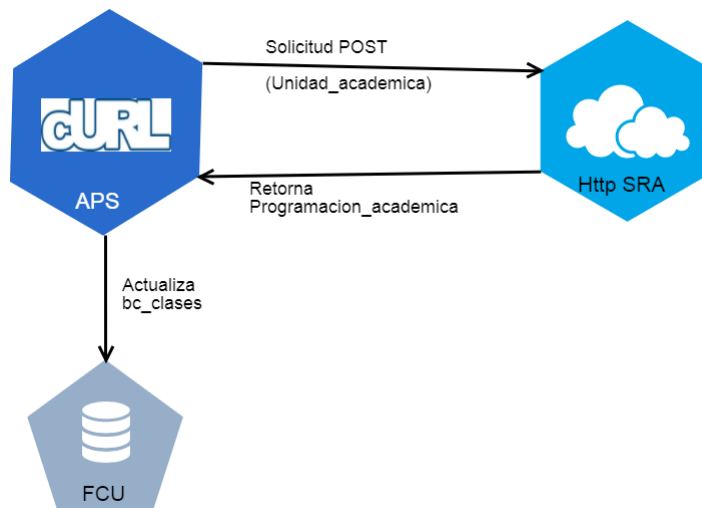


Figura 21. Proceso de sincronización de la programación académica.

La Figura 21, muestra el proceso que sigue esta funcionalidad, realiza una solicitud POST con la librería CURL al sitio web de sistema de registro académico SRA con el parámetro de la unidad académica seleccionada. El sitio responde la programación académica, de la cual es extraída los horarios de clases de las asignaturas y son actualizadas en el fuente de conocimiento básico de la universidad FCU, mas específicamente en la tabla bc_clases.

4.4.1.3 Fuente de conocimiento de la web de la Universidad

Esta última fuente de conocimiento se carga a través de la implementación de un buscador, el cual hace uso de las herramientas apache nutch y apache solr. Estas herramientas permiten hacer crawler, indexar y realizar consultas en el dominio de la Universidad. En total se indexaron 407 páginas web.

4.4.2 Proceso de responder a las preguntas de los usuarios

La Figura 22 muestra la secuencia seguida en la ejecución del asistente virtual:

1. El usuario envía la pregunta que desea realizar desde la interfaz
2. Se hace la consulta dependiendo de los patrones de coincidencia
 - a. Se encuentra el patrón
 - i. Si la respuesta contiene una consulta a la base de conocimiento de la Universidad del Valle
 1. Se realiza la consulta a la base de conocimiento
 - a. Si encuentra información en la base de conocimiento la retorna.
 - b. Sino, pasa al literal 2.b (no tiene almacenada la respuesta con respecto a la Universidad del Valle)
 - ii. Si la respuesta no contiene una consulta a la base de conocimiento de la Universidad del Valle
 1. Se devuelve la respuesta a la interfaz
 - b. Si no encuentra el patrón

- i. Se utiliza el buscador para obtener información
 1. Se obtienen las palabras claves de la pregunta del usuario
 2. Se realiza la búsqueda
 - a. Si obtiene información, devuelve las urls obtenidas como respuesta a la interfaz
 - b. Si no obtiene información, devuelve un mensaje a la interfaz indicando que no tiene respuesta a la pregunta
3. Se le da la respuesta al usuario.

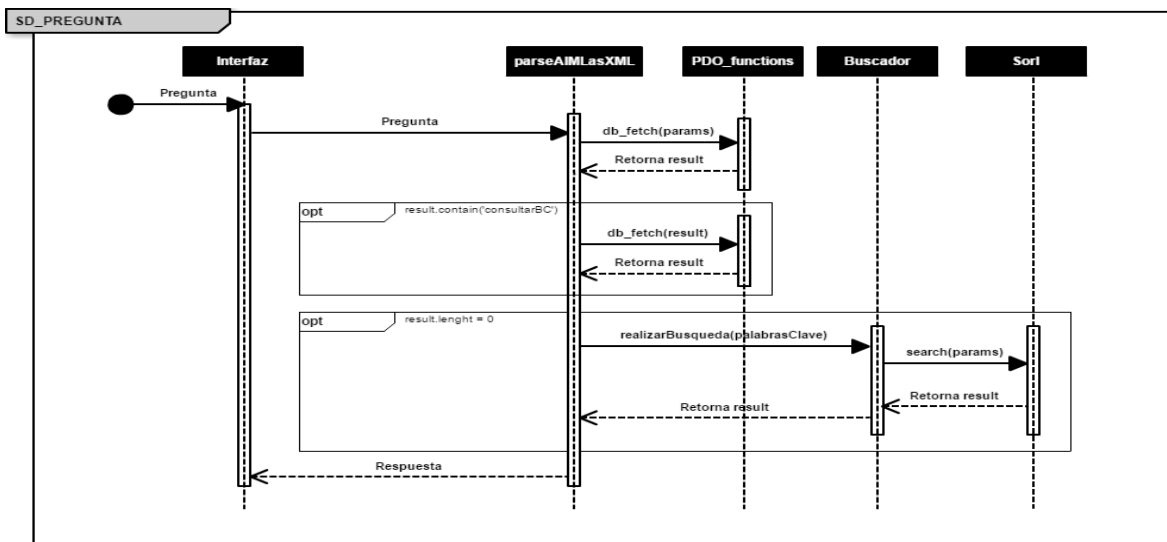


Figura 22. Diagrama de secuencia para la realización de una consulta.

4.5 Ejemplos de ejecución

Los siguientes son ejemplos de ejecución del asistente virtual para las tres fuentes de conocimiento.

4.5.1 Información personal

A continuación se puede observar un ciclo de conversación donde el usuario interactúa con el agente de manera natural, iniciando la conversación con un saludo, seguidamente hace un par de preguntas y finaliza dando las gracias.

La Figura 23, muestra la respuesta del asistente cuando el usuario le dice 'hola'.

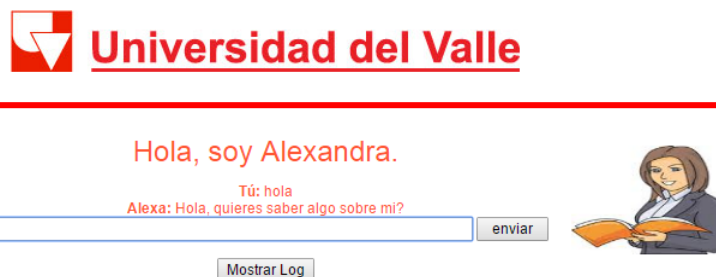


Figura 23. Ejemplo de ejecución 1, Hola.

La Figura 24, muestra la respuesta del asistente cuando el usuario le pregunta '¿dónde naciste?'.



Hola, soy Alexandra.

Tú: donde naciste?
Alexa: Soy de Cali

enviar

Mostrar Log



Figura 24. Ejemplo de ejecución 2, ¿Dónde naciste?

La Figura 25, muestra la respuesta del asistente cuando el usuario le pregunta '¿cuál es tu actor favorito?'.



Hola, soy Alexandra.

Tú: Cual es tu actor favorito?
Alexa: Me gusta mucho Brad pitt

enviar

Mostrar Log



Figura 25. Ejemplo de ejecución 3, ¿Cuál es tu actor favorito?

La Figura 26, muestra la respuesta del asistente cuando el usuario le dice 'Muchas gracias'.



Hola, soy Alexandra.

Tú: Muchas gracias
Alexa: Con mucho gusto, deseas saber sobre algo mas?

enviar

Mostrar Log



Figura 26. Ejemplo de ejecución 4, Muchas gracias.

4.5.2 Base del conocimiento de la Universidad del Valle

La Figura 27, muestra la respuesta del asistente cuando el usuario le pregunta '¿Cuál es la misión de la EISC?'.

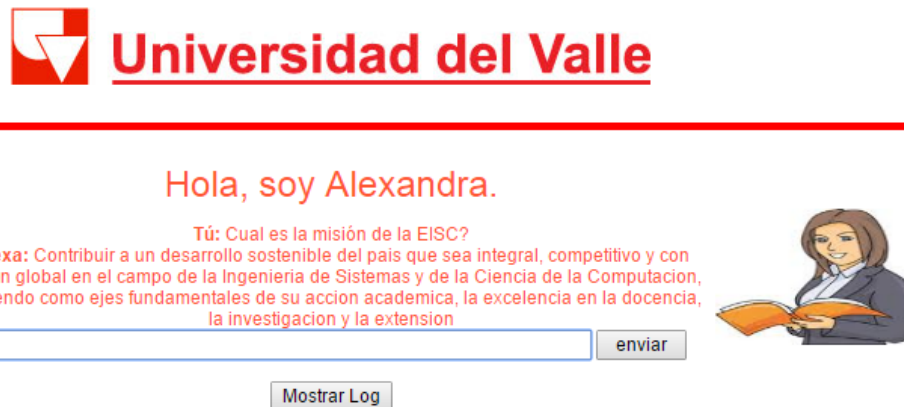


Figura 27. Ejemplo de ejecución 1, ¿Cuál es la misión de la EISC?

La Figura 28, muestra la respuesta del asistente cuando el usuario le pregunta '¿Cuál es la página web de la Facultad de Ingenierías?'.



Figura 28. Ejemplo de ejecución 2, ¿Cuál es la página de la Facultad de Ingenierías?

La Figura 29, muestra que el asistente solicita información complementaria cuando el usuario le pregunta '¿Qué prerequisites tiene la asignatura Calculo II?' y al usuario aportarla y completar la pregunta, da su respuesta.



Universidad del Valle

Hola, soy Alexandra.

Tú: que prerequisites tiene la asignatura Calculo II?
Alexa: Indicame el programa academico para poderte ayudar



Universidad del Valle

Hola, soy Alexandra.

Tú: Ingeniería de Sistemas
Alexa: el prerequisite de la asignatura calculo ii es 111050M - Calculo I para Ingenieria



Figura 29. Ejemplo de ejecución 3, ¿Que prerequisites tiene Calculo II?

4.5.3 Base de conocimiento del sitio web

La Figura 30, muestra que el asistente responde una lista de enlace extraídos por el motor de búsquedas cuando el usuario le pregunta '¿Cuáles son los diplomados de Arquitectura?'.



Universidad del Valle

Hola, soy Alexandra.

Tú: Cuales son los diplomados de Arquitectura?
Alexa: Se ha realizado la busqueda de diplomados Arquitectura



Sin titulo

<http://arquitectura.univalle.edu.co/index.php/category/extension/diplomados/>

Figura 30. Ejemplo de ejecución 1, ¿Cuáles son los diplomados de Arquitectura?

La Figura 31, muestra que el asistente responde una lista de enlaces extraídos por el motor de búsquedas cuando el usuario le pregunta '¿Cuál es la historia del Departamento de Física?'.



Universidad del Valle

Hola, soy Alexandra.

Tú: cual es la historia del Departamento de Física?
Alexa: Se ha realizado la busqueda de historia Departamento de Física



Historia
<http://fisica.univalle.edu.co/index.php/historia>

Departamento de Geografia/Facultad de Humanidades/Universidad del Valle
<http://geografia.univalle.edu.co/historia.html>

Figura 31. Ejemplo de ejecución 2, ¿Cuál es la historia del Departamento de Física?

La Figura 32, muestra que el asistente responde una lista de enlaces extraídos por el motor de búsquedas cuando el usuario le pregunta '¿Dónde encuentro el directorio de Univalle?'.



Universidad del Valle

Hola, soy Alexandra.

Tú: Donde encuentro el directorio de Univalle?
Alexa: No tengo información al respecto, talvez te pueda ser util alguno de los siguientes enlaces



Directorio / Universidad del Valle / Cali, Colombia
<http://www.univalle.edu.co/directorio-univalle>

Directorio telefónico / Universidad del Valle / Cali, Colombia
<http://www.univalle.edu.co/english/directorio/>

Figura 32. Ejemplo de ejecución 3, ¿Dónde encuentro el directorio de Univalle?

La Figura 33, muestra que el asistente responde una lista de enlaces extraídos por el motor de búsquedas cuando el usuario le pregunta '¿Cuál es el pensum de Arquitectura?'.



Universidad del Valle

Hola, soy Alexandra.

Tú: CUAL ES EL PENSUM DE ARQUITECTURA?
Alexa: En los siguientes enlaces puedes encontrar información que te sea útil

Pensum del pregrado en Arquitectura / Escuela de Arquitectura / Universidad del Valle
<http://arquitectura.univalle.edu.co/index.php/programa-academico-de-arquitectura/pensum-del-pregrado-en-arquitectura/>

Pensum Especialización en Administración de empresas de la construcción / Escuela de Arquitectura /
<http://arquitectura.univalle.edu.co/index.php/especializacion-de-administracion-de-empresas-de-la-construccion/pensum-especializacion-en-administracion-de-empresas-de-la-construccion/>



Figura 33. Ejemplo de ejecución 4, ¿Cuál es el pensum de Arquitectura?

CAPÍTULO 5

PRUEBAS Y RESULTADOS

En este capítulo se describe la implementación del asistente virtual y las pruebas que se realizaron, donde se evaluaron los aspectos de interactividad del asistente, efectividad y actualidad de la información en sus respuestas. Además, se discuten los resultados obtenidos.

5.1 Implementación del asistente

A continuación se dan algunos detalles de la implementación del asistente virtual:

- La metodología de desarrollo utilizada es XP (eXtrem Programming), bajo esta metodología se realiza el levantamiento de las historias de usuarios y se basa en la realización de ciclos cortos tras los cuales se obtiene retroalimentación. En el Anexo 2 se pueden observar los artefactos de análisis y diseño del asistente virtual.
- El asistente virtual se desarrolló en un ambiente web, utilizando el lenguaje de programación PHP y complementándolo con JavaScript y CSS. También se utilizó el lenguaje estructurado AIML, en el cual se crearon los patrones de coincidencia y respuestas del asistente.
- El proyecto fue desplegado en una máquina virtual con el sistema operativo Ubuntu y el servidor web apache.
- Para la implementación del analizador morfológico se utilizó Freeling, una herramienta que permite hacer análisis lingüístico de texto a través de funciones de procesamiento de lenguaje natural.
- El proyecto se basó en la implementación de Program O, el cual es el motor procesador del asistente virtual.
- La estructura de base de datos de Program O es necesariamente montada en un motor de base de datos MySql y así se decidió implementar la base de conocimiento del asistente (visto en la sección 4.3.2.1) en este motor.
- Para la implementación de “Text-to-Speech” es necesario el uso de la aplicación ‘espeak’ la cual está instalada predeterminadamente en Ubuntu 14, pero debe tenerse en cuenta si en otras distribuciones Linux está instalada o no.

Para la implementación del asistente virtual en un ambiente de pruebas fue necesario instalar:

- Ubuntu 14
- Program O modificado
- MySql 5.5
- Apache2
- Php 5
- Freeling 4.0
- Espeak

5.2 Encuestas realizadas

Para evaluar el asistente virtual se desarrolló una serie de preguntas que buscan medir el nivel de desempeño bajo tres aspectos: interactividad, efectividad y actualidad. Las preguntas relacionadas con el aspecto de interactividad, ayudan a determinar qué tan fácil es para el usuario el uso del asistente virtual. Las preguntas de efectividad, permiten medir la precisión que existe en la información contenida en las respuestas dadas por el asistente. Finalmente, el nivel de actualidad del asistente se evalúa por parte del usuario indicando si la información que provee el asistente es actual, o no. Las pruebas se desarrollaron en un ambiente en el que el usuario entabla una conversación con el asistente e intenta obtener información sobre la Universidad del Valle.

En la prueba participaron un total de 25 usuarios que pertenecen a la Universidad del Valle, los cuales eran de diferentes carreras y dependencias de la universidad. Los usuarios calificaron las respuestas que el asistente les brindaba indicando el grado de precisión, seleccionando una de las siguientes opciones: “No hubo respuesta”, “Respuesta imprecisa”, “Respuesta poco precisa”, “Respuesta medianamente precisa”, “Respuesta precisa” las cuales corresponden al nivel de precisión de la respuesta en una escala del uno al cinco respectivamente. Durante el proceso de evaluación de las preguntas se observó que los usuarios las calificaban como “respuesta imprecisa” a las respuestas erróneas y como “respuesta precisa” a las respuestas correctas, pero como “respuesta poco precisa” cuando el asistente daba respuestas que se podían encontrar en los resultados obtenidos pero había que ingresar y revisar en varias páginas, y por último, los usuarios calificaban como “medianamente precisa” cuando el resultado no era de los primeros dos en aparecer en la lista. Además, al final de la conversación los usuarios realizaron una evaluación general del asistente donde indicaban el nivel del conocimiento del asistente calificando si era “Muy bueno”, “Bueno”, “Regular”, “Malo” y “Muy malo”. El usuario también calificó la actualidad de la información donde indicaba si la información contenida en las respuestas era “Actual”, “Poco actual” o “No era actual”, y por último, el usuario evaluó la interactividad con el asistente respondiendo dos preguntas “¿Usted encuentra visualmente fácil el asistente?” y “¿Es intuitivo el inicio de la conversación?”. A continuación se muestran los resultados obtenidos de la prueba y su análisis.

5.2.1 Resultados de la evaluación del grado de precisión

Durante las pruebas realizadas con los 25 usuarios, se hicieron un total de 449 preguntas. En la Tabla 11 se muestra cómo se calificaron las respuestas dadas por el asistente para esas preguntas en términos de precisión. De acuerdo a los resultados obtenidos, las respuestas dadas por el asistente fueron precisas en un 53.23%. Se puede observar que las respuestas son precisas solo en la mitad de los casos, lo que es un buen resultado si se observan las tablas Tabla 12 y Tabla 14 donde se aprecia que las respuestas más imprecisas son obtenidas del buscador de la universidad con solo un 16,4% de respuestas precisas. De igual forma, los resultados indican que aún se debe mejorar mucho en la gestión de las respuestas y en el conocimiento del asistente.

Calificación	Cantidad	Porcentaje
No hubo respuesta	34	7,57%
Respuesta imprecisa	65	14,48%
Respuesta poco precisa	42	9,35%
Respuesta medianamente precisa	69	15,37%
Respuesta precisa	239	53,23%
Total	449	100%

Tabla 11. Evaluación individual de las respuestas del asistente

En las tablas Tabla 12, Tabla 13 y Tabla 14 se puede observar las evaluaciones de las respuestas discriminando la fuente de conocimiento desde donde fue resuelta por el asistente. A partir de los resultados se puede concluir que las respuestas son más precisas cuando viene de las fuentes de conocimiento propias del asistente y son más imprecisas cuando son obtenidas a través del buscador. Por ejemplo, cuando se pregunta por el horario del curso Matemáticas Discretas II grupo 04, el asistente responde a partir de su conocimiento con la respuesta precisa. Sin embargo, al hacer la misma consulta en el buscador de Univalle aparecen enlaces que hacen referencia a la programación académica de la sede Yumbo e incluso aparecen enlaces relacionados con una convocatoria de docentes.

Calificación	Cantidad	Porcentaje
No hubo respuesta	5	3,85%
Respuesta imprecisa	3	2,31%
Respuesta poco precisa	4	3,08%
Respuesta medianamente precisa	3	2,31%
Respuesta precisa	115	88,46%
Total	130	100%

Tabla 12. Evaluación individual de las respuestas de la base de conocimiento de Univalle

Calificación	Cantidad	Porcentaje
No hubo respuesta	16	12,31%
Respuesta imprecisa	12	9,23%
Respuesta poco precisa	2	1,54%
Respuesta medianamente precisa	7	5,38%
Respuesta precisa	93	71,54%
Total	130	100%

Tabla 13. Evaluación individual de las respuestas a preguntas personales

Calificación	Cantidad	Porcentaje
No hubo respuesta	13	6,88%
Respuesta imprecisa	50	26,46%
Respuesta poco precisa	36	19,05%
Respuesta medianamente precisa	59	31,22%
Respuesta precisa	31	16,40%
Total	189	100%

Tabla 14. Evaluación individual de las respuestas a preguntas sobre la Universidad

5.2.2 Resultados de la evaluación del nivel de conocimiento

Las tablas Tabla 15 y Tabla 16 muestran cómo el usuario calificó el nivel de conocimiento del asistente en el ámbito de la universidad y personal. Por ejemplo durante las pruebas se le hizo la pregunta ¿A qué te dedicas?, a lo que el asistente respondió “Soy un estudiante de la Universidad del Valle y trabajo de asistente virtual para ayudar a los usuarios a encontrar información sobre ella en su sitio web.” mostrando que tiene cierto conocimiento sobre sí mismo. Por otra parte, con respecto a la información de la universidad un usuario hizo la consulta ¿Dónde queda la facultad de Humanidades?, a lo que el asistente respondió “La facultad de humanidades está en el edificio 384, 2045 - Meléndez”. Los resultados muestran que el 72% de los usuarios se sintieron satisfechos con las respuestas obtenidas, al calificar el nivel como ‘Bueno’ y en algunos casos como ‘Muy bueno’, para ambos ámbitos.

Calificación	Cantidad	Porcentaje
Muy malo	0	0%
Malo	2	8%
Regular	5	20%
Bueno	15	60%
Muy bueno	3	12%
Total	25	100%

Tabla 15. Evaluación general del conocimiento base de la Universidad

Calificación	Cantidad	Porcentaje
Muy malo	0	0%
Malo	2	8%
Regular	6	24%
Bueno	12	48%
Muy bueno	5	20%
Total	25	100%

Tabla 16. Evaluación general sobre conocimiento personal del asistente

5.2.3 Resultados de la evaluación de actualidad

En la Tabla 17 se observa la calificación de la actualidad de la información en las respuestas del asistente. En general las respuestas dadas por el asistente presentaron información actualizada, aunque cabe anotar aquí que algunas respuestas no lo eran debido a las fuentes de información. Por ejemplo, un usuario preguntó: “Cual es el correo del profesor Víctor Manuel Parada”, a lo cual el asistente respondió: “El correo electrónico de VICTOR MANUEL es vicparas@univalle.edu.co” la cual es desactualizada debido a que la información fue recolectada de las sitios web de las diferentes dependencias y algunas se encontraban desactualizadas. La información actual es “victor.parada@correounivalle.edu.co”.

Calificación	Cantidad	Porcentaje
Respuestas no son actuales	0	0%
Respuestas son poco actuales	3	12%
Respuestas son actuales	22	88%
Total	25	100%

Tabla 17. Evaluación de la actualidad de las respuestas del asistente

5.2.4 Resultados de la evaluación de interactividad

En las tablas Tabla 18 y Tabla 19 se observa el resultado de la evaluación del aspecto de interactividad del asistente. El 100% de los usuarios respondieron positivamente a la pregunta “¿Usted encuentra visualmente fácil el asistente?” y el 96% de los usuarios respondieron positivamente a “¿Es intuitivo el inicio de la conversación?”, lo que indica que la comunicación con el asistente es fácilmente lograda.

Calificación	Cantidad	Porcentaje
No	0	0%
Si	25	100%
Total	25	100%

Tabla 18. Respuesta a: “¿Usted encuentra visualmente fácil el asistente?”

Calificación	Cantidad	Porcentaje
No	1	4%
Si	24	96%
Total	25	100%

Tabla 19. Respuesta a: “¿Es intuitivo el inicio de la conversación?”

CAPÍTULO 6

CONCLUSIONES

En este capítulo se presentan las conclusiones más importantes de este trabajo de grado. Además, se proponen trabajos futuros que se derivan de la realización del asistente virtual propuesto.

6.1 Conclusiones

A continuación se describen las conclusiones del trabajo de grado:

- El asistente virtual brinda respuestas más precisas que los resultados obtenidos por el buscador de la Universidad del Valle. Esto se debe a que la información de las páginas de la universidad no están totalmente actualizadas o que en ella no se encuentra cierta información importante para los estudiantes. Esto se evidencia en las tablas Tabla 12 y Tabla 14, donde la precisión de las respuestas del asistente virtual es del 88,4% para las preguntas sobre el conocimiento básico de la universidad y es del 16,4% para las respuestas redireccionadas al buscador de la universidad.
- El asistente virtual permite que el usuario realice búsqueda de información de manera más agradable en comparación con un buscador tradicional. Esto se debe a que el asistente permite que el usuario se comunique con él como si se tratara de otra persona. El comportamiento anterior fue evidenciado en la Tabla 19, en la cual el 96% de los usuarios evaluados indicaron que el inicio de la conversación con el asistente virtual era intuitivo y por ello fácil de comunicarse.
- El asistente virtual permite integrar otras herramientas que brindan información que no es fácilmente obtenida por medio de un buscador, ya que estas herramientas son aplicaciones web y no pueden ser indexadas. Por ejemplo, al solicitar información de la programación académica de una asignatura, el buscador no devolvería una respuesta precisa mientras que el asistente sí, debido a que se le puede incorporar este conocimiento y además, se integra con la aplicación web de programación académica.
- Las desventajas de este asistente son: Los patrones de coincidencia se tienen que ir adicionando a medida que el asistente no responde a las preguntas del usuario; y su conocimiento no se actualiza automáticamente, debido a que no se cuenta con fuentes confiables para extraer información.

6.2 Trabajo futuro

Hay varios aspectos sobre los cuales se puede trabajar en el futuro para mejorar el asistente virtual:

- Mejorar la interfaz gráfica, adicionándole una animación que simule expresiones faciales.
- Incorporar al asistente un mecanismo “Speech to Text”, el cual permitiría al usuario comunicarse con él por voz.
- Mejorar el módulo de aprendizaje, para que pueda obtener nuevo conocimiento desde las conversaciones con los usuarios e implementando nuevas técnicas.
- Integrar nuevas herramientas de la universidad que puedan ser de gran ayuda para la comunidad universitaria, como el mapa de la universidad, entre otras.
- Recopilar la información de todas las dependencias académicas de la universidad para cargarlo al asistente y ampliar su conocimiento básico de la universidad.
- Mejorar la sincronización de conocimiento de asistente, permitiéndole que se conecte los diferentes sistemas de información de la universidad por medio de servicios web que brinde más confianza en la obtención del conocimiento.

BIBLIOGRAFÍA

- [1] "Universidad del Valle Colombia" [Online]. Available: <http://univalle.edu.co> [Accedido: 27-mayo-2016]
- [2] "Program O." [Online]. Available: <http://www.program-o.com/>. [Accedido: 18-abril-2017]
- [3] A.L.I.C.E., "AIML: Artificial Intelligence Markup Language." [Online]. Available: <http://www.alicebot.org/aiml.html>. [Accedido: 18-abril-2017]
- [4] "Asistente Virtual (Chatbot) para la web de la Facultad de Informática" [Online]. Available: <http://eprints.ucm.es/33443/> [Accedido: 27-mayo-2016]
- [5] "Análisis morfológico." [Online]. Available: <http://roble.pntic.mec.es/~msanto1/lengua/1anamorf.htm>. [Accedido: 18-abril-2017]
- [6] "Elementos de gramática española" [Online]. Available: <http://repositorio.up.edu.pe/handle/11354/1528> [Accedido: 29-agosto-2017]
- [7] "Grampal." [Online]. Available: <http://cartago.illf.uam.es/grampal/>. [Accedido: 18-abril-2017]
- [8] E. M. Eisman, V. López, and J. L. Castro, "A framework for designing closed domain virtual assistants," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3135–3144, 2012.
- [9] "Asistente virtual de la Universidad de Granda España" [Online]. Available: <http://tueris.ugr.es/elvira/> [Accedido: 26-abril-2017]
- [10] S. M. CAMACHO JARA, "Un Sofbot de Charla Desarrollado con Técnicas de Razonamiento basado en Casos," *Rev. Investig. Sist. e Informática*, vol. 3, no. 4, pp. 20–28, 2006.
- [11] D. C. Morales, Y. R. Constanten, and Y. T. Rubio, "Sistema de Razonamiento Basado en Casos para la identificación de riesgos de software System of Case-Based identification Reasoning to support software risk," vol. 7, no. 2, pp. 95–112, 2013.
- [12] "Hefagra." [Online]. Available: www.hefagra.es/. [Accedido: 26-agosto-2017]
- [13] L. R. Aguado, A. M. Serramo and P. M. Fernández, "Planteamiento semántico y pragmático para la gestión de diálogos en asistentes virtuales". *Procesamiento de Lenguaje Natural*, 28, 2002, pp. 81-88.
- [14] S. Ghose and J. J. Barua, "Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor," 2013 International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, 2013, pp. 1-5.
- [15] A. Augello, G. Pilato, A. Machi, and S. Gaglio, "An approach to enhance chatbot semantic power and maintainability: Experiences within the frasi project," in *Semantic Computing (ICSC)*, 2012 IEEE Sixth International Conference on. IEEE, 2012, pp. 186-193.

ANEXOS

Anexo 1. Preguntas que responde el asistente virtual

- Abreviatura
 - Retorna
 - Nombre dependencia
 - Nombre asignatura
 - Nombre programa académico
 - Nombre centro de investigación
 - Nombre grupo de investigación
 - Nombre laboratorio
 - Nombre línea de investigación
 - Recibe
 - Abreviatura -> ABREV
 - Preguntas
 - ¿Qué es ABREV?
 - ¿Qué significa ABREV?
 - ¿Qué es la abreviatura ABREV?
 - ¿Qué significa la abreviatura ABREV?
- Cantidad
 - Retorna
 - Cantidad de escuelas
 - Cantidad de Facultades
 - Cantidad de sedes
 - Cantidad de departamentos
 - Cantidad de institutos
 - Recibe
 - Palabra -> 'Escuelas', 'Facultades' o 'Sedes'
 - Palabra -> 'Departamentos' o 'Institutos'
 - Preguntas
 - ¿Cuántas E-F-S tiene la Universidad?
 - ¿Cuántas E-F-S tiene la Universidad del Valle?
 - ¿Cuántos D-I tiene la Universidad?
 - ¿Cuántos D-I tiene la Universidad del Valle?
- Nombre
 - Retorna
 - Nombre de una escuela
 - Nombre de una Facultad
 - Nombre de un departamento
 - Nombre de un instituto
 - Nombre de un profesor
 - Recibe
 - Palabra -> 'Escuela', 'Facultad', 'Departamento' o 'Instituto'
 - Asignatura -> ASIG

- Programa académico -> PA
 - Grupo de investigación -> GI
 - Dependencia -> DEP
- Preguntas
 - ¿A qué F-E-D-I pertenece DEP?
 - ¿A qué F-E-D-I pertenece ASIG?
 - ¿A qué F-E-D-I pertenece PA?
 - ¿A qué F-E-D-I pertenece GI?
 - ¿A qué F-E-D-I pertenece la asignatura ASIG?
 - ¿A qué F-E-D-I pertenece el programa académico PA?
 - ¿A qué F-E-D-I pertenece la carrera PA?
 - ¿A qué F-E-D-I pertenece el grupo de investigación GI?
 - ¿Qué F-E-D-I-P dicta ASIG?
 - ¿Qué F-E-D-I-P dicta la asignatura ASIG?
 - ¿Quién dicta ASIG?
 - ¿Quién dicta la asignatura ASIG?
 - ¿Qué profesores dictan ASIG?
 - ¿Quién da ASIG?
 - ¿Quién da la asignatura ASIG?
 - ¿Qué profesores dan ASIG?
- Profesor que dicta una asignatura el semestre actual
 - Retorna
 - Nombre profesor
 - Recibe
 - Asignatura -> ASIG
 - Preguntas
 - ¿Quién dicta ASIG este semestre?
 - ¿Quién dicta este semestre ASIG?
 - ¿Quién da ASIG este semestre?
 - ¿Quién da este semestre ASIG?
- E-mail
 - Retorna
 - Email
 - Recibe
 - Palabra -> 'Correo electrónico' o 'E-mail'
 - Persona -> P
 - Preguntas
 - ¿Cuál es el CE-EMAIL de P?
 - ¿En qué CE-EMAIL puedo contactar a P?
 - ¿Dónde puedo enviar un CE-EMAIL a P?
- Nombre persona que tiene la ocupación
 - Retorna
 - Nombre de la persona
 - Recibe

- Ocupación -> Oc
 - Escuela -> E
 - Departamento -> D
 - Instituto -> I
 - Facultad -> F
- Pregunta
 - ¿Quién es la Oc de E-D-I-F?
- Teléfono
 - Retorna
 - Teléfono
 - Recibe
 - Facultad -> F
 - Escuela -> E
 - Departamento -> D
 - Instituto -> I
 - Persona -> P
 - Preguntas
 - ¿Cuál es el teléfono de E-D-I-F-P?
- Página web
 - Retorna
 - Link
 - Recibe
 - Palabras -> 'Página web' (PW) – 'Sitio web' (SW) – 'Link' (L) o 'Contenido' (C)
 - Palabra -> 'Universidad del Valle' (U)
 - Programa académico -> PA
 - Asignatura -> ASIG
 - Escuela -> E
 - Departamento -> D
 - Instituto -> I
 - Preguntas
 - ¿Cuál es el PW-SW-L-C de PA-ASIG-E-D-I-U?
- Ubicación dependencia
 - Retorna
 - Edificio Escuela, Departamento o Instituto
 - Recibe
 - Escuela -> E
 - Departamento -> D
 - Instituto -> I
 - Preguntas
 - ¿Cuál es la ubicación de E-D-I?
 - ¿Dónde está E-D-I?
 - ¿Dónde queda E-D-I?
 - ¿En qué edificio queda E-D-I?
 - ¿En qué edificio esta E-D-I?

- ¿Cuál es el edificio de E-D-I?
- Oficina persona
 - Retorna
 - Oficina
 - Recibe
 - Persona -> P
 - Preguntas
 - ¿Cuál es la oficina de P?
 - ¿Dónde ubico a P?
 - ¿Dónde está la oficina de P?
- Espacio de investigación
 - Retorna
 - Espacio Laboratorio, Grupo de investigación o Centro de investigación
 - Recibe
 - Palabra -> 'Laboratorio', 'Grupo de investigación' o 'Centro de investigación'
 - Nombre -> NOM
 - Abreviatura -> ABREV
 - Preguntas
 - ¿Dónde queda el LAB-GI-CI de NOM-ABR?
- Oficina de programa académico
 - Retorna
 - Oficina
 - Recibe
 - Programa académico -> PA
 - Preguntas
 - ¿Cuál es la ubicación del PA?
 - ¿Cuál es la oficina del PA?
 - ¿Dónde queda el PA?
 - ¿Dónde está la oficina del PA?
- Ubicación sede
 - Retorna
 - Dirección
 - Recibe
 - Sede -> S
 - Preguntas
 - ¿Dónde queda la S?
 - ¿Cuál es la ubicación de la S?
- Misión – Visión
 - Retorna
 - Misión o visión
 - Recibe
 - Palabras -> 'Misión' (M) o 'Visión' (V)
 - Palabra -> 'Univalle' (U)
 - Facultad -> F
 - Escuela -> E

- Departamento -> D
 - Instituto -> I
- Preguntas
 - ¿Cuál es la M-V de F-E-D-I-U?
- Duración programa académico
 - Retorna
 - Cantidad de semestres que dura la carrera
 - Recibe
 - Programa académico -> PA
 - Preguntas
 - ¿Cuál es la duración de PA?
 - ¿Qué duración tiene PA?
 - ¿Cuánto tiempo dura PA?
 - ¿Cuánto dura PA?
- Perfil profesional
 - Retorna
 - Perfil
 - Recibe
 - Programa académico -> PA
 - Título -> T
 - Preguntas
 - ¿Cuál es el perfil de PA-T?
 - ¿Cuál es el perfil de los estudiantes de PA?
 - ¿Cuál es el perfil de los estudiantes de PA de la Universidad del Valle?
 - ¿Cuál es el perfil de los estudiantes de PA egresados de la Universidad del Valle?
 - ¿Cuál es el perfil de los estudiantes PA que se gradúan de la Universidad del Valle?
- Título egresado
 - Retorna
 - Título
 - Recibe
 - Programa académico -> PA
 - Preguntas
 - ¿Cuál es el título otorgado por PA?
 - ¿Cuál es el título que reciben los estudiantes egresados de PA?
 - ¿Cuál es el título que reciben los estudiantes graduados de PA?
 - ¿Cuál es el título que reciben los estudiantes que se gradúan de PA?
 - ¿Qué título se otorga a los estudiantes de PA?
- Jornada de estudio
 - Retorna
 - Diurna o Nocturna
 - Recibe
 - Programa académico -> PA
 - Preguntas
 - ¿Cuál es la jornada de clases de PA?
 - ¿Cuál es la jornada del programa académico de PA?

- ¿Cuál es la jornada de PA?
- Asignaturas de la carrera
 - Retorna
 - Lista asignaturas
 - Recibe
 - Programa académico -> PA
 - Preguntas
 - ¿Cuáles son las asignaturas de PA?
 - ¿Qué asignaturas pertenecen a PA?
- Pensum de la carrera
 - Retorna
 - Pensum
 - Recibe
 - Programa académico -> PA
 - Preguntas
 - ¿Cuál es el pensum de PA?
 - ¿Cuál es la malla curricular de PA?
- Lista de tipos de personas
 - Retorna
 - Personas
 - Recibe
 - Escuela -> E
 - Departamento -> D
 - Instituto -> I
 - Preguntas
 - ¿Quiénes son los profesores contratistas de E-D-I?
 - ¿Quiénes son los profesores hora cátedra de E-D-I?
 - ¿Quiénes son los profesores de planta de E-D-I?
 - ¿Quiénes son los funcionarios de E-D-I?
 - ¿Quiénes son el personal administrativo de E-D-I?
- Código de asignaturas
 - Retorna
 - Código
 - Recibe
 - Asignatura -> ASIG
 - Preguntas
 - ¿Cuál es el código de ASIG?
 - ¿Cuál es el código de la asignatura ASIG?
- Cantidad de créditos
 - Retorna
 - Créditos
 - Recibe
 - Programa académico -> PA
 - Asignatura -> ASIG
 - Preguntas
 - ¿De cuántos créditos es ASIG-PA?

- ¿De cuántos créditos es la asignatura ASIG?
 - ¿De cuántos créditos es el programa académico PA?
 - ¿Cuántos créditos tiene ASIG-PA?
 - ¿Cuántos créditos tiene la asignatura ASIG?
 - ¿Cuántos créditos tiene el programa académico PA?
- Asignatura habilitable
 - Retorna
 - Si es habilitable o no
 - Recibe
 - Asignatura -> ASIG
 - Preguntas
 - ¿La asignatura ASIG es habilitable?
 - ¿ASIG es habilitable?
- Asignatura validable
 - Retorna
 - Si es validable o no
 - Recibe
 - Asignatura -> ASIG
 - Preguntas
 - ¿La asignatura ASIG es validable?
 - ¿ASIG es validable?
- Asignatura anual
 - Retorna
 - Si es anual o no
 - Recibe
 - Asignatura -> ASIG
 - Preguntas
 - ¿La asignatura ASIG es anual?
 - ¿ASIG es anual?
- Prerrequisitos
 - Retorna
 - Lista de prerrequisitos
 - Recibe
 - Programa académica -> PA
 - Asignatura -> ASIG
 - Resolución -> RES
 - Preguntas
 - ¿Qué prerrequisito tiene ASIG? -> Solicita programa académico
 - ¿Cuáles son los prerrequisitos de ASIG? -> Solicita programa académico
 - ¿Qué prerrequisito tiene ASIG de PA?
 - ¿Cuáles son los prerrequisitos de ASIG de PA?
 - ¿Qué prerrequisito tiene ASIG de la resolución RESOL?
 - ¿Cuáles son los prerrequisitos de ASIG de la resolución RESOL?
 - ¿ASIG tiene prerrequisitos?
 - ¿La asignatura ASIG tiene prerrequisitos?
- Asignaturas del profesor este semestre

- Retorna
 - Lista de asignaturas
- Recibe
 - Palabras -> 'Asignaturas' o 'Materias'
 - Profesor -> P
- Preguntas
 - ¿Qué ASIG-MAT dicta PROF este semestre?
 - ¿Qué ASIG-MAT da PROF este semestre?
 - ¿Qué ASIG-MAT dicta este semestre PROF?
 - ¿Qué ASIG-MAT da este semestre PROF?
 - ¿Este semestre que ASIG-MAT dicta PROF?
 - ¿Este semestre que ASIG-MAT da PROF?
- Trabajos de los grupos de investigación
 - Retorna
 - Trabajos
 - Recibe
 - Grupos de investigación -> GI
 - Preguntas
 - ¿Qué hacen en GI?
 - ¿En qué trabajan en GI?
 - ¿Qué hacen en el grupo de investigación GI?
 - ¿En qué trabajan en el grupo de investigación GI?
- Horario de clases de una asignatura
 - Retorna
 - Horario
 - Recibe
 - Asignatura -> ASIG
 - Grupo -> GR
 - Preguntas
 - ¿Cuál es el horario de ASIG? -> solicita el grupo
 - ¿Cuál es el horario de ASIG grupo GR?
 - ¿Qué horario tiene ASIG? -> solicita el grupo
 - ¿Qué horario tiene ASIG grupo GR?
- Disponibilidad de un profesor
 - Retorna
 - Horario
 - Recibe
 - Profesor – PROF
 - Día – DIA
 - Preguntas
 - ¿Qué disponibilidad tiene PROF?
 - ¿Qué disponibilidad tiene PROF el DIA?
 - ¿Qué disponibilidad tiene PROF el día DIA?
 - ¿Cuál es el horario de atención de PROF?
 - ¿Cuál es el horario de atención de PROF el DIA?
 - ¿Cuál es el horario de atención de PROF el día DIA?

- ¿En qué horario encuentro a PROF?
- ¿En qué horario encuentro a PROF el DIA?
- ¿En qué horario encuentro a PROF el día DIA?
- ¿Cuál es la disponibilidad de PROF?
- ¿Cuál es la disponibilidad de PROF el DIA?
- ¿Cuál es la disponibilidad de PROF el día DIA?

Anexo 2. Análisis y diseño

Historias de usuarios

1. Como usuario quiero ingresar nuevo conocimiento por medio de conversaciones.
2. Como usuario quiero obtener información de la Universidad del Valle.
3. Como administrador quiero confirmar el conocimiento adquirido de los usuarios.
4. Como administrador quiero agregar nuevo conocimiento a través de un archivo Excel.
5. Como usuario quiero que las búsquedas en el dominio de la Universidad se hagan extrayendo las palabras claves de la conversación.
6. Como administrador quiero obtener la programación académica del semestre y hacerla disponible desde el agente.
7. Como desarrollador considero importante identificar si la oración de la conversación es una pregunta o no (Ya que hay preguntas que se parecen mucho a afirmaciones).

Requisitos funcionales

1. El sistema genera respuestas (contiene varias respuestas a la misma pregunta)
2. El sistema evita responder las mismas preguntas (pero también tiene la opción de repetirlas por solicitud específica del usuario)
3. Actualización de conocimiento, como hay información que cambia cada semestre esta debe ser actualizada por cada facultad por medio de archivos.
4. El sistema permite conectarse a la web de programación académica y extraer información sobre los horarios de las asignaturas.
5. El sistema consulta el buscador de la universidad
6. El sistema consulta el directorio de la universidad
7. Abrir URL en un marco
8. Independencia de la morfología de los términos para las búsquedas

Requisitos no funcionales

1. Usabilidad, la interfaz debe ser sencilla e intuitiva, el usuario no debería tener dificultades para usarla por primera vez.
2. Estándar gráfico, debe tener el mismo estándar gráfico del sitio web de la universidad.
3. Concurrencia de usuarios, debe soportar a más de 30 usuarios sin verse afectado su rendimiento.
4. Multiplataforma, el sistema puede ser ejecutado desde cualquier navegador que soporte la ejecución de JavaScript.
5. Idioma, el sistema solo estará disponible en el idioma Español.

6. El sistema debe tener un tiempo máximo de respuesta de 5 segundos en cualquier tipo de consulta.

Casos de Uso

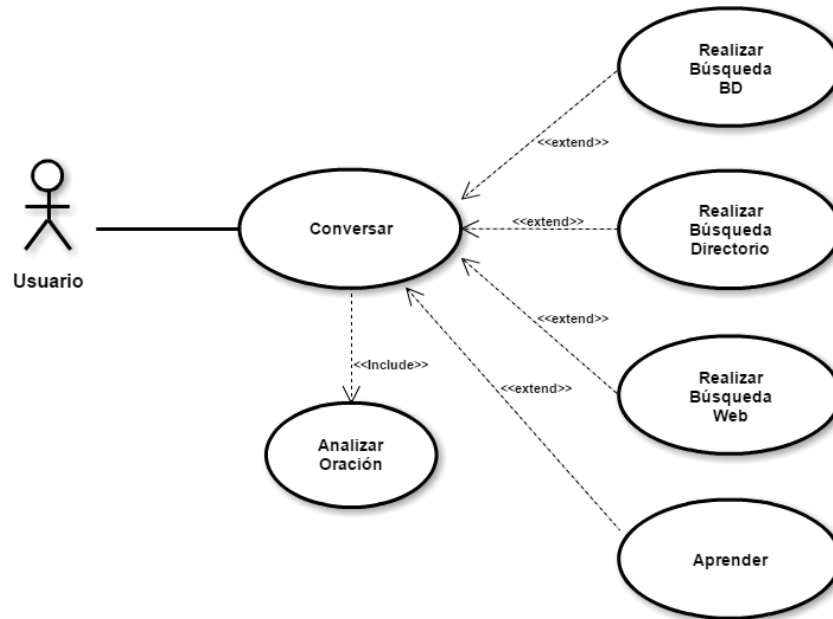


Figura 34. Diagrama de casos de uso del actor Usuario

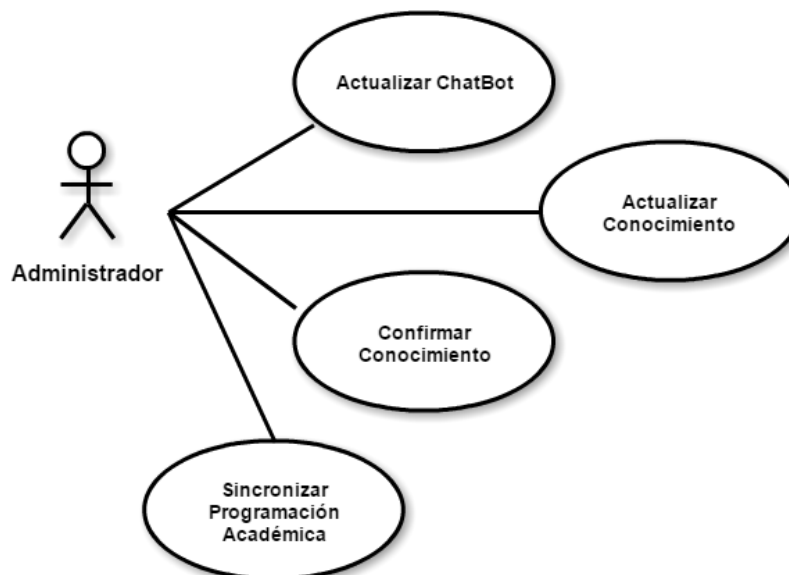


Figura 35. Diagrama de casos de uso del actor Usuario

CU-01	INICIAR CONVERSACION	
Dependencias		
precondición		
Descripción	El usuario debe escribirle al asistente.	
Secuencia normal	Paso	Acción
	1	El usuario escribe en el cuadro de diálogo.
	2	Detectar la pulsación de 'Enter' o clic en el botón.
	3	Se envía petición asíncrona al servidor.
Postcondición	El servidor recibe la oración enviada por el usuario.	
Excepciones	Paso	Acción

CU-02	RESPUESTA PREDEFINIDA	
Dependencias		
precondición	CU-01	
Descripción	El sistema deberá seguir los siguientes pasos, cuando la información solicitada está disponible en el chatbot.	
Secuencia normal	Paso	Acción
	1	Se analiza la oración que envió el usuario, se determina que es pregunta.
	2	Se busca patrones de coincidencia.
	3	Se obtiene respuesta.
Postcondición	Se muestra la respuesta al usuario.	
Excepciones	Paso	Acción
	2	No hay patrón de coincidencia

CU-03	RESPUESTA DE LA BASE DE CONOCIMIENTO DE UNIVALLE	
Dependencias		
precondición	CU-01	
Descripción	El sistema debe realizar los siguientes pasos y mostrará información acerca de Univalle almacenada en la base de datos.	
Secuencia normal	Paso	Acción
	1	Se analiza la oración que envió el usuario, se determina que es pregunta.
	2	Se busca patrones de coincidencia.
	3	Se consulta la base de conocimiento de Univalle.
	4	Se obtiene respuesta.
Postcondición	Se muestra la respuesta al usuario.	
Excepciones	Paso	Acción
	2	No hay patrón de coincidencia
	4	No hay respuesta

CU-04	BÚSQUEDA DIRECTORIO UNIVALLE	
Dependencias		
precondición	CU-03 no hay respuesta	
Descripción	El sistema podrá mostrará información de contacto acerca de personas que no conoce.	
Secuencia normal	Paso	Acción
	1	Se ejecuta en CURL la web del directorio.
	2	El sistema obtiene los resultados.
Postcondición	Se responde al usuario con los resultados obtenidos.	
Excepciones	Paso	Acción
	1	Si la web del directorio no responde se pasa al buscador.
	2	Si no se obtienen resultados.

CU-05	BÚSQUEDA WEB	
Dependencias		
precondición	CU-02 no encontró patrón de coincidencia, CU-03 falló, CU-04 no hubo resultados	
Descripción	El sistema verifica si hay palabras clave, y responderá al usuario.	
Secuencia normal	Paso	Acción
	1	Se analizará la oración, identificando las palabras claves.
	2	Petición asíncrona al buscador de Univalle.
	3	Se obtiene los resultados de búsqueda.
Postcondición	Se responde al usuario con los resultados obtenidos.	
Excepciones	Paso	Acción
	1	No hay palabras claves, se muestra al usuario que no se tiene la respuesta.

CU-06	APRENDER	
Dependencias		
precondición	CU-01	
Descripción	El sistema deberá seguir los siguientes pasos, cuando la oración enviada por el usuario es una afirmación.	
Secuencia normal	Paso	Acción
	1	Se analiza la oración que envió el usuario, se determina que es afirmación.
	2	Se busca patrones de coincidencia.
	3	Es almacenado temporalmente.
Postcondición	Se almacena el conocimiento temporalmente a la espera de ser confirmado.	
Excepciones	Paso	Acción
	2	No hay patrón de coincidencia

CU-07	ACTUALIZAR CONOCIMIENTO	
Dependencias		
precondición		
Descripción	El administrador deberá seguir los siguientes pasos para que el asistente actualice su conocimiento de Univalle por medio de archivos de Excel.	
Secuencia normal	Paso	Acción
	1	Upload archivo de Excel
	2	Seleccionar el tipo de información que va a cargar.
	3	Clic en botón 'add'.
	4	Se actualiza la base de datos.
Postcondición	Mensaje de resumen de información actualizada.	
Excepciones	Paso	Acción

CU-08	CONFIRMAR CONOCIMIENTO	
Dependencias		
precondición	CU-06	
Descripción	El administrador confirma que conocimiento adquirido en las conversaciones por el asistente virtual, es válido.	
Secuencia normal	Paso	Acción
	1	Busca el conocimiento.
	2	Modifica el patrón - respuesta (opcional)
	3	Clic en botón 'learn'.
Postcondición	El asistente virtual adquiere un nuevo conocimiento.	
Excepciones	Paso	Acción

CU-09	SINCRONIZAR PROGRAMACIÓN ACADÉMICA	
Dependencias		
precondición	CU-07 hace falta tener toda la información de dependencias, profesores y asignaturas	
Descripción	El sistema se conecta a programación académica, obtiene la información de las horas de clase y las almacena en su conocimiento.	
Secuencia normal	Paso	Acción
	1	El usuario selecciona una Facultad o Instituto.
	2	Clic en el botón 'synchronize'.
	3	El sistema se conecta a la web de programación académica.
	4	Se obtiene la programación del semestre de la unidad académica.
	5	Se actualiza el conocimiento del asistente virtual.
Postcondición	Se actualiza el conocimiento del asistente sobre los horarios de clase y le responde al usuario con un mensaje de resumen de la información actualizada.	
Excepciones	Paso	Acción