



Universidad de Valladolid

Escuela de Ingeniería Informática

TRABAJO DE FIN DE GRADO

Grado en Ingeniería Informática
Mención en Computación

Asistentes virtuales: estudio del estado de la cuestión y desarrollo de un prototipo

Autor:

D. Pablo Valdunciel Sánchez

Tutor:

Dr. César González Ferreras

CONTENIDOS DE LA PRESENTACIÓN

1

INTRODUCCIÓN

2

PLAN
DE
PROYECTO

3

TECNOLOGÍAS
PARA LA
CONSTRUCCIÓN
DE
CHATBOTS

4

WORD EMBEDDINGS
-
SIMILARIDAD
SEMÁNTICA

5

CONSTRUCCIÓN
DE LOS
PROTOTIPOS

6

EVALUACIÓN
CON
USUARIOS

7

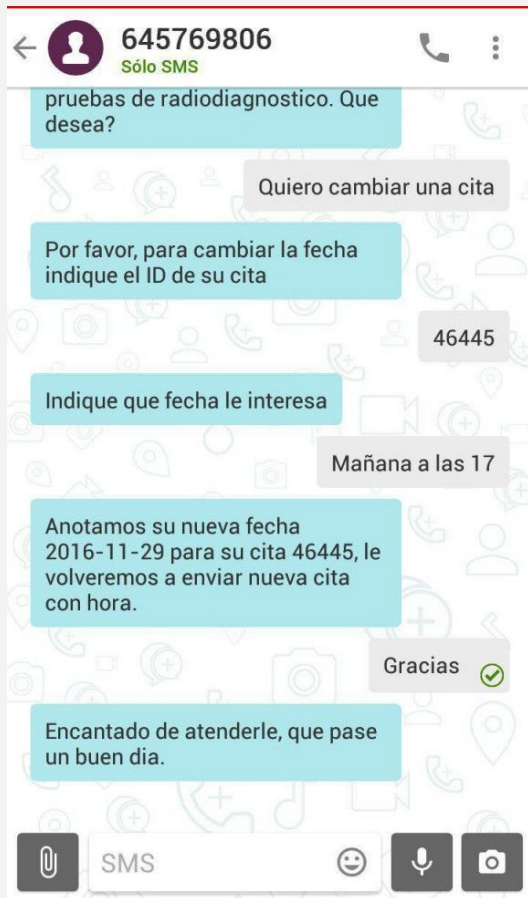
CONCLUSIONES
Y
TRABAJO
FUTURO



01

INTRODUCCIÓN

¿QUÉ ES UN ASISTENTE VIRTUAL O CHATBOT?



“programas que utilizan tecnología diseñada para simular una interacción conversacional con usuarios humanos, y que pueden incluir procesos automatizados ejecutados como resultado de esas interacciones”



PROCESOS AUTOMATIZADOS EN 2022

75 %
SERVICIOS
DE
SALUD

90%
BANCA



OBJETIVOS

1. Aprender los conceptos fundamentales de un chatbot.
2. Estudiar y comparar las tecnologías existentes para la construcción de chatbots.
3. Estudiar los conceptos básicos de la vectorización de palabras (*word embeddings*).
4. Desarrollo de 2 prototipos que sirvan como prueba de concepto de este tipo de sistema.
 - a. Con *Framework*
 - b. Sin *Framework* (librerías de PLN)
5. Evaluar y comparar los dos prototipos desarrollados.



02

PLAN DE PROYECTO



PLAN DE PROYECTO

METODOLOGÍA

Scrum of One

1 *sprint* = 2 semanas

8 *sprints*

PROTOTIPADO

Prototipado vertical

Prototipado en paralelo

GESTIÓN DE RIESGOS

- 10 riesgos identificados
- Descriptores cualitativos de impacto y probabilidad

Impacto\Prob.	Bajo	Moderado	Significativo	Alto
Alto	Ri05		Ri01, Ri03, Ri07, Ri08, Ri09	
Significativo	Ri02, Ri06, Ri10			
Moderado				Ri04
Bajo				

- Planes de Protección
- Planes de Contingencia

Impacto\Prob.	Bajo	Moderado	Significativo	Alto
Alto				
Significativo	Ri07	Ri09	Ri08	
Moderado	Ri06, Ri10	Ri03		
Bajo	Ri01, Ri02, Ri05			Ri04

PLAN DE TAREAS

Tareas principales

1. Elaborar Plan de Proyecto.
2. Obtención de conjuntos de pares pregunta-respuesta.
3. Obtención de datos para evaluar la similaridad semántica entre oraciones.
4. Estudio de las tecnologías para la construcción de chatbots.
5. Estudio de los algoritmos de vectorización de palabras.
6. *Construcción del prototipo 1.
7. *Construcción de prototipo 2.
8. Evaluación con usuarios de los prototipos.
9. Otros.

Calendario

480 horas
16 semanas
30 horas /semana
20 de enero - 8 de mayo

Estimado



458 horas y 10 minutos
15 semanas
30-36 horas/semana
20 de enero - 30 de abril

Real

PRESUPUESTO

6.974'75 €

03

TECNOLOGÍAS PARA LA CONSTRUCCIÓN DE CHATBOTS



APLICACIONES DE MENSAJERÍA



WhatsApp Business API

Asociado a empresa real
Posibles costes
Formas de interacción algo limitadas



Messenger Platform

Asociado a una página de Facebook
Amplias formas de interacción



Telegram Bot API

Creación muy sencilla
Amplias formas de interacción



WeChat

Estricto control del gobierno chino
Uso limitado a China



Slack RM API

Entorno profesional
Grupos

FRAMEWORKS



Amazon Lex

Integrado con AWS

Integración automática con FM, Slack y Twilio

Sólo disponible en inglés



Watson Conversation Services

Interfaz potente

Soporta 10 lenguajes

Traductor incorporado

No permite diálogos basados en ML (no FAQs)



Dialogflow

Soporta 20 lenguajes

Integración automática de diferentes canales

Knowledge connectors

Reconocimiento de entidades basado en
diccionarios



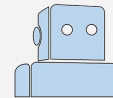
Azure Bot Service

Reconocimiento de entidades basado en ML

Soporta 15 lenguajes

QnA Maker

Funcionalidades gratuitas limitadas



Wit.ai

Soporta 80 lenguajes

Código abierto

API sencilla y fácil de usar

Requiere mayor esfuerzo de desarrollo

LIBRERÍAS DE PLN



Python

Funcionalidades exclusivas del **inglés**

No soporta *word embeddings*

No soporta el entrenamiento sobre **GPU**

Sin modelos pre-entrenados para **análisis del discurso o revisión ortográfica**

spaCy

Python / Cython

Documentación comprensible y con ejemplos

Soporta **53 idiomas**

Modelos pre-entrenados en **11 idiomas**

Word embeddings de **GloVe** integrados

Soporta el entrenamiento sobre **GPU**

Sin modelos pre-entrenados para **análisis del discurso o revisión ortográfica**

Spark NLP

Java / Scala / Python

Soporta **5 idiomas**

Análisis del discurso

Revisión ortográfica

Word / sentence embeddings (GloVe, W2V, BERT, ELMO)

Entrenamiento sobre GPU

Stanford CoreNLP

Java

Soporta **6 idiomas**

Análisis del discurso

Revisión ortográfica

SUTime

¿ Word embeddings ?
Verbosa, más compleja



04

SIMILARIDAD SEMÁNTICA ENTRE PARES DE ORACIONES



WORD EMBEDDINGS Y SIMILARIDAD SEMÁNTICA

Similaridad semántica

“Métrica definida sobre un conjunto de documentos o términos, donde la idea de distancia entre ellos se define en base la igualdad de significado o contenido semántico”

Word embedding

“Nombre que se le da a un conjunto de lenguajes de modelado y técnicas de aprendizaje automático en Procesamiento del Lenguaje Natural (PLN) donde las palabras del vocabulario son vinculadas a vectores de números reales”

Algoritmos basados en predicción para generar *word embeddings*

Word2Vec
(W2V)

GloVe

FastText
(FT)

Medidas de similaridad

Average
+
Similitud Coseno
(AVG)

Smooth Inverse Frequency
+
Similitud Coseno
(SIF)

(-1) * Word Mover's Distance
(WMD)

COMPARACIÓN DE MÉTODOS

CONJUNTOS DE VECTORES

Modelo	Vectores pre-entrenados	Vocabulario	Dimensión de los vectores	Idioma
Word2Vec	GoogleNews-vectors-negative300	3 millones	300	Inglés
GloVe	glove.840B.300d	2,2 millones	300	Inglés
FastText	crawl-300d-2M	2 millones	300	Inglés

DATOS

Corpus	Nº de instancias
STS12 (test)	3108
STS13 (test)	1500
STS14 (test)	3750
STS15 (test)	3000
STS16 (test)	1186
SICK (relatedness)	9927

- 6 conjuntos de datos
- Similitud entre oraciones puntuada entre 0 y 5 (*Gold Standards*)
- Toolkit SentEval

COMPARACIÓN DE MÉTODOS



TEST SOBRE RANKINGS

1. Cálculo del Coeficiente de correlación de Pearson.
2. Para cada conjunto, ordenar los métodos según el Coef. de Pearson (ranking).
3. Asignar la posición del ranking a cada método.
4. Calcular los rankings promedios.

Test de Friedman

→ Existen diferencias significativas.

Test de Bonferroni-Dunn

→ **FT+SIF** es significativamente mejor que *W2V+WMD*, *FT+WMD*, *GLOVE+WMD*, *GLOVE+AVG* y *GLOVE + AVG*

Posición	Ranking promedio	Método
1	1	FT + SIF
2	2,33	W2V + SIF
3	3	GLOVE + SIF
4	4,33	W2V + AVG
5	4,67	FT + AVG
6	6,67	W2V + WMD
7	7	FT + WMD
8	7,83	GLOVE + AVG
9	8,17	GLOVE + WMD

05

CONSTRUCCIÓN DE LOS PROTOTIPOS



DESCRIPCIÓN DE LOS PROTOTIPOS

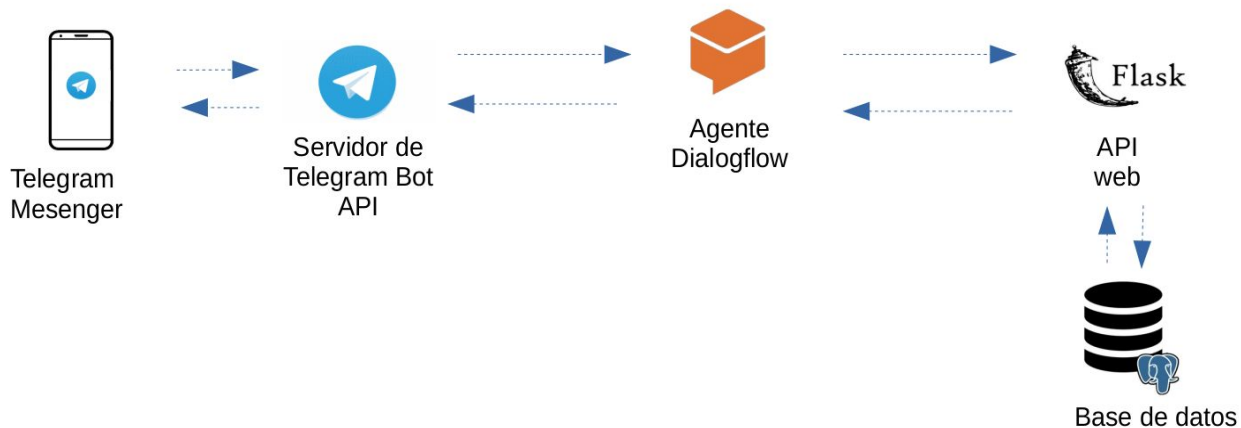
Asistente virtual para el servicio de Relaciones Internacionales de la Universidad de Valladolid (UVa)

Funcionalidades del chatbot *RelintBot*

1. Responder a las preguntas frecuentes sobre estudiar en la Universidad de Valladolid presentes en la web del servicio de Relaciones Internacionales.
2. Conversación casual (*smalltalk*): saludar, despedir, responder a un agradecimiento.
3. Concertar citas en la oficina de Relaciones Internacionales especificando: asunto, fecha y hora.
4. Gestión de idiomas: inglés y español
5. Logging y monitorización.



RelintBot2



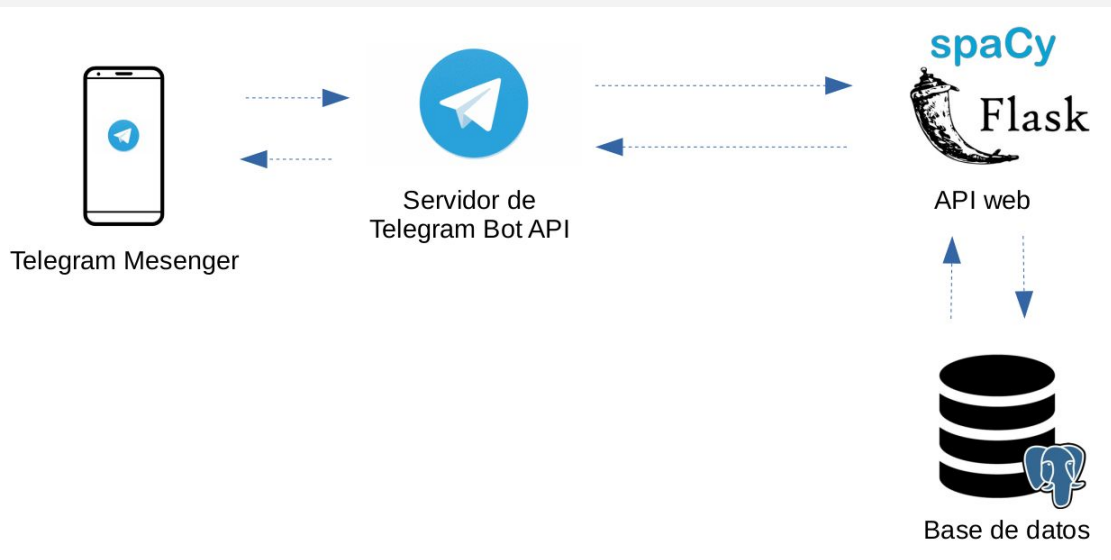
Esfuerzo estimado: 60 horas
Esfuerzo real: 45 horas y 30 minutos

1. Responder a FAQs → **Knowledge connector**
2. Conversación casual → Módulo **Small talk**
3. Concertar cita → *intent* Appointment con parámetros y conexión webhook con **API externa**.
4. Idiomas: **inglés**
5. Logging y monitorización incluidos en Dialogflow

The slide features three faint, light-gray hexagonal outlines: one in the top-left corner, one in the top-right corner, and one in the bottom-left corner. A large, solid black triangle is positioned on the right side of the slide, pointing towards the center.

Demostración RelintBot2

RelintBot1



Esfuerzo estimado: 140 horas.

Esfuerzo real: 154 horas y 10 minutos.

1. Base de conocimiento → basada en el cómputo de la similaridad semántica a partir de *word embeddings*.
2. Intenciones → componente ***TextClassifier***
3. Entidades → componente ***EntityRuler***
4. Idiomas → **inglés** y **español**
5. Logging → registro de mensajes en la BD y logs en un fichero de texto
6. Monitorización → extensión *Flask Monitoring Dashboard*

The slide features three faint, light-gray hexagonal outlines: one in the top-left corner, one in the top-right corner, and one in the bottom-left corner. A large, solid black triangle is positioned on the right side of the slide, pointing towards the center.

Demostración RelintBot1

06

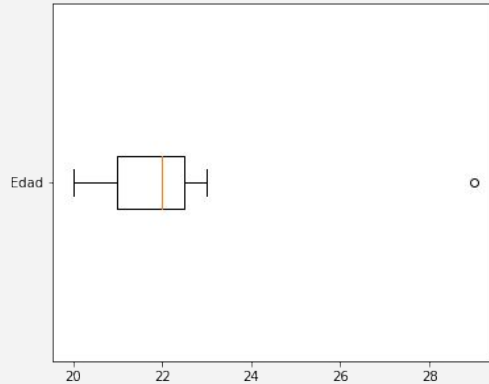
EVALUACIÓN CON USUARIOS



USUARIOS

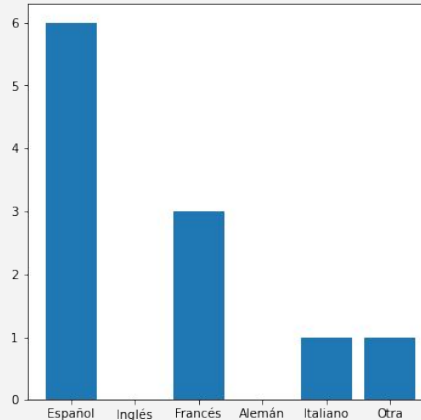
En el proceso de evaluación participaron 12 usuarios. Finalmente, se consideraron únicamente las evaluaciones de **11 usuarios**.

Edad



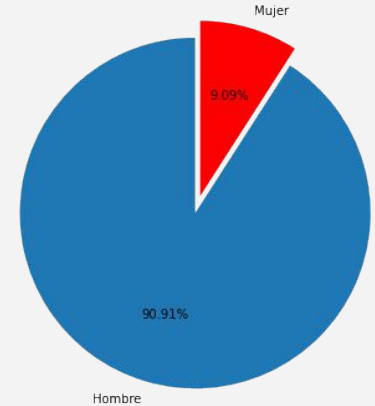
Media = 22'27 años
Mediana = 22 años
Mínimo = 20 años
Máximo = 29 años

Lengua Materna



¿ Nivel de inglés ?

Sexo



Un único usuario Mujer

TAREA 1

Resolver 5 dudas relacionadas con estudiar en la Universidad de Valladolid realizando preguntas al chatbot *Relint*

Prototipo	Media	Desviación típica	Mínimo	Máximo
RelintBot1	3,81	0,98	3	5
RelintBot2	4,45	0,93	2	5

→ El p-valor obtenido para el contraste de hipótesis sobre la igualdad de medias es **0,135**.

Fallos RelintBot1

1. El chatbot proporcionó la respuesta a otra pregunta muy similar
2. Detección incorrecta de la intención “Appointment”

Fallos RelintBot2

1. El chatbot proporcionó la respuesta a otra pregunta muy similar

TAREA 2

Concertar una cita en la oficina de Relaciones Internacionales al día siguiente a las doce de la mañana para obtener información sobre cursos de español.

Prototipo	Completada	Completada (%)	Número medio de mensajes	Número medio de caracteres por mensaje
RelintBot1	8	72,73 %	6,38	183,00
RelintBot2	8	72,73 %	3,25	170,38

Fallos RelintBot1

1. Imposibilidad en ocasiones de reconocer formatos no estándares de fechas
2. Imposibilidad en ocasiones de reconocer el motivo de la cita.

Fallos RelintBot2

1. Imposibilidad de expresar que la hora deseada ya estaba ocupada o que en la fecha y/o hora solicitadas la oficina estaba cerrada.

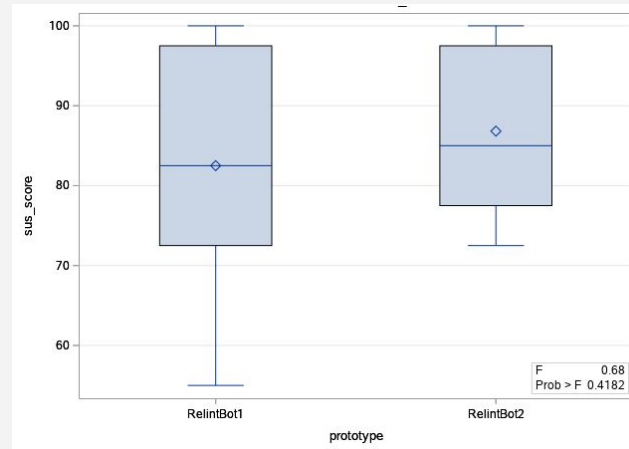
TAREA 3

Obtener la dirección de la oficina de Relaciones Internacionales de la Universidad de Valladolid.

Prototipo	Completada	Completada (%)	Número medio de mensajes	Número medio de caracteres por mensaje
RelintBot1	11	100 %	1,12	42,38
RelintBot2	11	100 %	1	41,36

USABILIDAD - SUS

System Usability Scale (SUS) - el usuario debe contestar a cada pregunta indicando su grado de acuerdo o desacuerdo, variando entre 1 y 5



Prototipo	Media	Desviación típica	Mínimo	Máximo
RelintBot1	82,50	14,14	55,00	100,00
RelintBot2	86,81	10,01	72,50	100,00

→ El p-valor obtenido para el contraste de hipótesis sobre la igualdad de medias es **0,4182**.

07

CONCLUSIONES y TRABAJO FUTURO



CONCLUSIONES

- ➔ La utilización de un *framework* a la hora de construir un chatbot reduce notablemente el esfuerzo de desarrollo. Como inconvenientes de la utilización de un *framework* destaca la funcionalidad limitada a la hora de trabajar en un idioma distinta al inglés y el poco o nulo control sobre los modelos utilizados.
- ➔ La construcción de un chatbot fuera de un *framework*, haciendo uso de librerías de PLN, tiene sentido únicamente si se conocen las bases de este tipo de sistemas y si los procesos que se desean automatizar son en cierta medida complejos o particulares.
- ➔ La tecnología actual permite construir sin problemas asistentes virtuales para automatizar procesos de diferente complejidad, pudiendo expresarse y comprender el lenguaje a un nivel muy cercano al de una persona. No obstante, desarrollar un sistema de este tipo no es sencillo y requiere haber trabajado previamente en este área.

TRABAJO FUTURO - MEJORAS

- ➔ **Dos niveles de intención:** establecer dos niveles de intención, agrupando diferentes intenciones con cierta similaridad por temas.
- ➔ **Desambiguación:** proporcionar a los usuarios las opciones posibles cuando su intención no esté clara y puede corresponderse con varias (después de la primera clasificación).
- ➔ **Opciones cerradas:** a la hora de reconocer entidades, cuando una entidad tenga un número reducido y claro de alternativas, proporcionar una lista para seleccionar una de las alternativas en lugar de permitir la entrada de texto libre.
- ➔ **Limitar la insistencia:** limitar la insistencia del chatbot, de manera que cuando el usuario fracase en proporcionar la información varias veces, el proceso en curso sea cancelado automáticamente.



Universidad de Valladolid

Escuela de Ingeniería Informática

TRABAJO DE FIN DE GRADO

Grado en Ingeniería Informática
Mención en Computación

Asistentes virtuales: estudio del estado de la cuestión y desarrollo de un prototipo

Autor:

D. Pablo Valdunciel Sánchez

Tutor:

Dr. César González Ferreras