



INSTITUTO POLITECNICO NACIONAL “IPN”



ESCUELA SUPERIOR DE COMPUTO “ESCOM”

LANDEROS CORTES PEDRO JONAS

MACHINE LEARNING

PRACTICA 3

INDICE

Introducción	2
Desarrollo.....	3
Flujo del programa	3
Ejecución del programa.....	3
Explicación de código	4
Anexos	5
Conclusiones	8

INDICE DE FIGURAS

Figure 1. Estado 0	3
Figure 2. Estado 60	4
Figure 3. Estado entre 1-57.....	4
Figure 4. Estado 59	4
Figure 5. Diagrama de flujo.....	5
Figure 6. Conjunto de fallback keywords	6

Introducción

En procesamiento de lenguaje natural (PLN) existen dos grandes familias de enfoques: Los basados en reglas y los basados en aprendizaje automático (machine learning). Los primeros, como el que se implementa en esta práctica, se apoyan en patrones explícitos, en esta práctica expresiones regulares, y en una maquina de estados finitos para lograr identificar valores útiles para el contexto en el cual se estén desarrollando. Los segundos emplean modelos entrenados con datos para generalizar a partir de ejemplos.

En esta práctica, se construye un chatbot temático sobre **Duolingo** con un pipeline clásico de reglas: normalizar los datos de entrada (eliminar acentos y pasar todo a minúsculas), detección de texto mediante expresiones regulares reforzadas con raíces (la parte fundamental de una palabra que contiene su significado original), sinónimos y control conversacional mediante estados, los cuales son definidos mediante un conjunto de datos previamente definidos y delimitados para una mejor interacción con el usuario.

Aunque aquí no se entrena un modelo de ML, el sistema sirve como referencia para comprender que problemas resuelve bien un enfoque determinista y cuando empiezan a ser útiles los modelos de ML en donde hay distintas variaciones entre entradas, y existe una mayor ambigüedad.

Desarrollo

Flujo del programa

1. **Normalización de la entrada:** Cada mensaje que entra se pasa por una función que recorta espacios, elimina acentos, y convierte a minúsculas. Con esto se reduce la variabilidad superficial de una misma palabra con distintas configuraciones y se facilita la detección de patrones.
2. **Detección con raíces y sinónimos:** El sistema cuenta con docenas de patrones específicos que capturan sinónimos, variaciones y conjugaciones comunes, si un patrón matchea, se asigna un número asociado a su correspondiente respuesta.
3. **Fallback por palabras clave:** Si ningún regex captura la entrada, se usa un “router” por palabras clave que intenta mapear términos frecuentes a estados relevantes, permitiendo que, ante entradas no previstas, se tenga una mayor cobertura de respuestas.
4. **Máquina de estados finitos:** El estado 0 es el estado inicial de entrada, lee la consulta y busca hacer el match con el regex correspondiente para posteriormente avanzar al estado correspondiente. Los estados 1-57 devuelven respuestas estáticas predefinidas, el estado 58 es el menú de continuación para repetir el ciclo, el estado 59 es la salida y el estado 60 es un estado que cubre el caso de no match y reformula la pregunta (este estado devuelve al estado 0).

Ejecución del programa

A continuación, se muestran las distintas etapas del programa.

La siguiente imagen es el estado inicial (0) al ejecutar el chatbot:

¡Hola! Soy el Chatbot no oficial sobre Duolingo. ¿En qué puedo ayudarte hoy?
Cuéntame tu duda (stories, racha, planes, precio, idiomas, etc.):

Figure 1. Estado 0

La siguiente imagen muestra el estado 60, el cual representa que no hubo ningún match y devuelve al estado 0:

```
¡Hola! Soy el Chatbot no oficial sobre Duolingo. ¿En qué puedo ayudarte hoy?  
Cuéntame tu duda (stories, racha, planes, precio, idiomas, etc.): asdasdasd  
No logré entender tu consulta. ¿Podrías intentarlo de otra forma? (prueba: 'precio', 'idiomas', 'planes', 'racha')  
Cuéntame tu duda (stories, racha, planes, precio, idiomas, etc.):
```

Figure 2. Estado 60

La siguiente imagen muestra una respuesta de los estados 1-57, la cual esta relacionada con los idiomas presentes en Duolingo:

```
Cuéntame tu duda (stories, racha, planes, precio, idiomas, etc.): asdasdasd  
No logré entender tu consulta. ¿Podrías intentarlo de otra forma? (prueba: 'precio', 'idiomas', 'planes', 'racha')  
Cuéntame tu duda (stories, racha, planes, precio, idiomas, etc.): que idiomas hay en duolingo  
Duolingo ofrece más de 100 cursos. Desde inglés, hay ~39 idiomas (p. ej., español, francés, navajo, klingon). También puedes aprender catalán desde español y hay inglés desde más de 25 idiomas base.  
¿Necesitas ayuda con algo más? (escribe 'salir' para terminar):
```

Figure 3. Estado entre 1-57

La siguiente imagen muestra el estado 59, el cual representa la salida:

```
¡Hola! Soy el Chatbot no oficial sobre Duolingo. ¿En qué puedo ayudarte hoy?  
Cuéntame tu duda (stories, racha, planes, precio, idiomas, etc.): quien creo duolingo  
Duolingo fue creado por Luis von Ahn, Severin Hacker y el equipo de Duolingo.  
¿Necesitas ayuda con algo más? (escribe 'salir' para terminar): salir  
¡Gracias por contactarme! Fue un placer ayudarte.
```

Figure 4. Estado 59

El programa cubre un total de 57 preguntas, las cuales se obtuvieron a través de la página oficial de FAQ de Duolingo y de Duolingo Wiki (un espacio para que los usuarios respondan preguntas generadas por otros usuarios),

Explicación de código

Configuración y utilidades: SHOW_DEBUG controla mensajes de depuración; dbg() los imprime cuando está activo. normaliza() aplica recorte, NFD y minúsculas para uniformar la entrada.

Patrones base: salir_REGEX detecta comandos de salida; hola_REGEX agrupa saludos y “ayuda/menú/opciones”.

Patrones de intención (q1_REGEX ... q57_REGEX): conjuntos de regex temáticas que incluyen **raíces y sinónimos**, cubriendo creadores, soporte, modelo de negocio, A/B testing, cambios de curso, racha, ligas, stories, podcasts, privacidad, gestión de cuenta, suscripciones, prueba, reembolsos, etc.

Atajos y fallback: qPRECIO_RE, qIDIOMAS_SUELTO_RE, qQUE_IDIOMAS_PUEDO_RE capturan consultas breves. FALBACK_KEYWORDS con route_by_keywords() asigna estados por palabras clave si no hubo match principal.

Bucle principal (run_chatbot_duolingo):

- **Estado 0** enruta por la lista ordenada de (regex, destino); si falla, prueba el fallback; si coincide “salir”, va a 59; de lo contrario, 60.
- **Estados 1–57, 61–62** devuelven respuesta y saltan a **58**.
- **Estado 58** repite el menú “¿algo más?” con la misma lógica de ruteo.
- **Estado 59** termina la sesión.
- **Estado 60** solicita reformulación y vuelve a 0.

Anexos

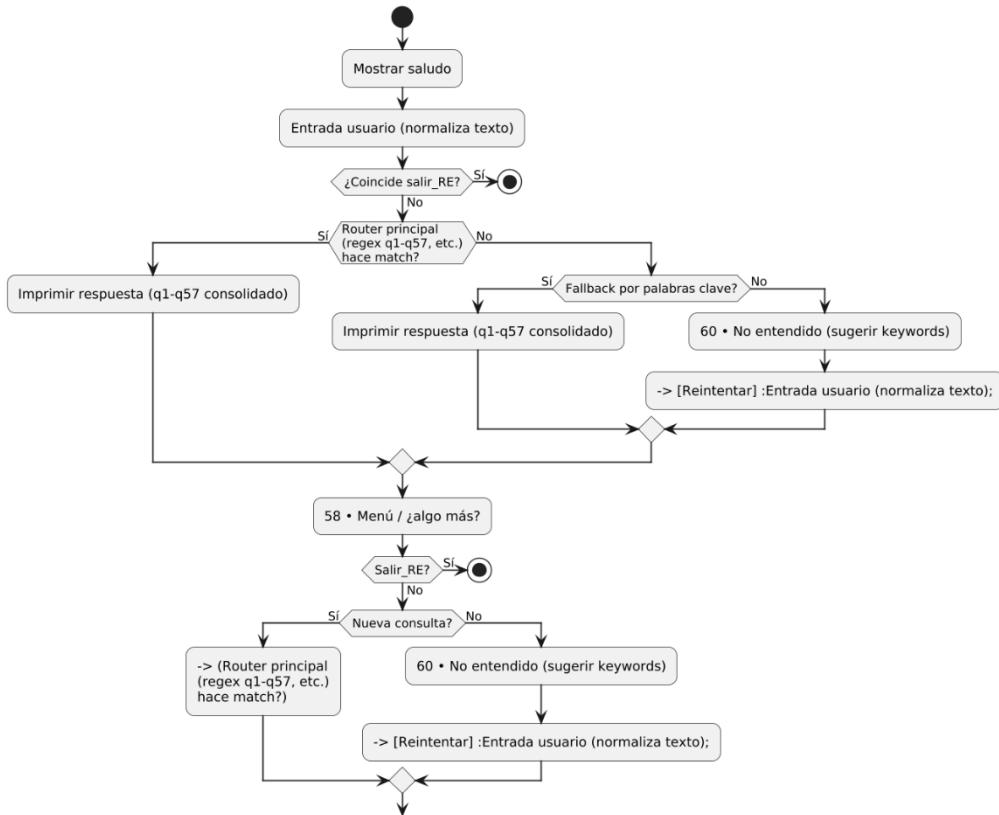


Figure 5. Diagrama de flujo

```

FALLBACK_KEYWORDS = [
    (re.compile(r"precio|costo|valor|cu[aá]nto (cuesta|sale)|tarifa"), 61),
    (re.compile(r"\bidiomas?\b|\blenguajes?\b"), 62),
    (re.compile(r"\bplanes?\b|\bsuscrip|\bgratis\b|\bfamiliar\b"), 42),
    (re.compile(r"cancel(ar|aci[oó]n)|dar de baja"), 35),
    (re.compile(r"reembols|devoluci[oó]n"), 36),
    (re.compile(r"privad"), 27),
    (re.compile(r"\bsegu(i|ir)|bloque|report"), 23),
    (re.compile(r"\belimin|borr|quitar .*cuenta"), 30),
    (re.compile(r"\breinici|reset.*curso"), 17),
    (re.compile(r"storie|historia(s)?\b"), 21),
    (re.compile(r"podcast"), 22),
    (re.compile(r"racha|streak"), 19),
    (re.compile(r"\bligas?\b|\bdivisiones?\b"), 20),
    (re.compile(r"\bmax\b|\bduolingo max\b"), 34),
]

```

Figure 6. Conjunto de fallback keywords

FAQ

1. ¿Quién creó Duolingo?

Duolingo fue creado por Luis von Ahn, Severin Hacker y el equipo de Duolingo.

2. ¿Dónde puedo obtener ayuda oficial?

En el Centro de Ayuda de Duolingo, en la Duolingo Wiki, o en el área de discusiones de la comunidad.

3. ¿Cuál es el modelo de negocio de Duolingo?

Duolingo gana dinero con el Duolingo English Test, las suscripciones de Súper Duolingo/Duolingo Max y compras dentro de la aplicación.

4. ¿Cada cuánto salen nuevas actualizaciones?

Cada pocas semanas. Algunas funciones se prueban, cambian o desaparecen según los resultados.

5. ¿Por qué Duolingo no tiene X función opcional?

Añadir demasiadas funciones haría la app confusa y difícil de mantener, por eso priorizan la simplicidad.

6. ¿Por qué yo tengo una función que mis amigos no tienen (o al revés)?

Duolingo hace pruebas A/B: lanza funciones solo a un grupo de usuarios y después decide si las mantiene.

7. Tenía una función y ya no la tengo, ¿por qué?

Si una función no muestra buenos resultados, puede retirarse o ponerse en pausa para mejorarla.

8. ¿Por qué cambió mi curso?

Porque Duolingo actualiza los cursos constantemente para hacerlos más efectivos y alinearte con el nuevo contenido.

9. ¿Cuántos idiomas ofrece Duolingo?

Alrededor de 39 idiomas desde inglés, incluyendo español, francés, navajo o klingon. También se puede aprender catalán desde español e inglés desde más de 25 idiomas base.

10. ¿Cuántos idiomas puedo aprender a la vez?

No hay límite. Puedes estudiar varios en paralelo.

11. ¿Qué idiomas inventados enseña Duolingo?

Klingon, Esperanto y Alto Valyrio.

12. ¿Qué lenguas en peligro puedo aprender en Duolingo?

Gaélico escocés, navajo y hawaiano.

13. ¿Duolingo tiene idiomas africanos?

Sí: suajili y zulú. El curso de xhosa se retiró en 2023.

14. ¿Qué idioma debería aprender en Duolingo?

Depende de tu objetivo: español o mandarín para el CV, francés para viajes, o klingon/alto valyrio por diversión.

15. ¿Cómo empiezo otro curso en Duolingo?

Haz clic en la bandera, selecciona "+" y elige el idioma.

16. ¿Cómo cambio de idioma en la aplicación?

Haz clic en la bandera y selecciona el curso que quieras practicar.

17. ¿Cómo reinicio un curso en Duolingo?

Debes eliminarlo desde tu perfil y volver a agregarlo.

18. ¿Cómo elimino un idioma en Duolingo?

Entra en Perfil → Configuración → Cursos → selecciona idioma → Quitar.

19. ¿Qué es la racha?

Es el número de días seguidos que completas lecciones. Sirve como motivación para practicar a diario.

20. ¿Qué son las Ligas y divisiones?

Son competencias semanales donde compites con otros usuarios al ganar EXP.

21. ¿Qué cursos tienen Stories (historias)?

Desde inglés: francés, español, portugués, alemán, italiano y japonés.

Para aprender inglés: alemán, español, francés, italiano, portugués, turco, ruso, ucraniano, hindi, coreano, chino y japonés.

Además, existe el proyecto comunitario Duostories con miles de historias extra.

22. ¿Qué idiomas tienen podcasts en Duolingo?

Cuatro: español (para angloparlantes), francés (para angloparlantes), inglés para hispanohablantes e inglés para lusohablantes.

23. ¿Cómo busco, sigo o bloqueo usuarios en Duolingo?

Para seguir: entra al perfil y pulsa "Seguir".

Para dejar de seguir: pulsa "Siguiendo".

Para bloquear: Perfil → Lista de amigos → Usuario → Bloquear.

Para reportar: Perfil → "Reportar usuario".

24. ¿Dónde veo a mis amigos en Duolingo?

En tu Perfil, tanto en la app como en la web, aparece tu lista de amigos y sus estadísticas.

25. ¿Qué pasa si no puedo seguir a alguien?

Verifica tu correo, completa una lección si tu cuenta está inactiva o revisa si eres menor de 13 años o tienes el perfil privado.

26. ¿Cómo envío felicitaciones en Duolingo?

Cuando un amigo alcanza un objetivo, la app te permite enviar un mensaje de apoyo.

27. ¿Cómo hago privado mi perfil en Duolingo?

Entra a Configuración de privacidad y desmarca "Volver público mi perfil".

28. ¿Cómo cambio mi nombre de usuario o correo electrónico?

En Configuración, edita nombre de usuario o correo y guarda cambios. Si ya están en uso, deberás modificarlos.

29. Tengo problemas para acceder a mi cuenta, ¿qué hago?

Restablece tu contraseña en duolingo.com/forgot_password. Si entraste con Google o Facebook, usa ese correo.

30. ¿Cómo elimino mi cuenta y mis datos?

Solicita la eliminación en la Bóveda de datos de Duolingo. Puede tardar hasta 30 días.

31. ¿Qué hago si mis datos se vieron comprometidos?

Si recibiste un aviso oficial, cambia tus contraseñas y protege tus cuentas.

32. ¿Qué es Súper Duolingo?

Una suscripción que incluye: sin anuncios, vidas ilimitadas, práctica personalizada, repaso de errores, desafíos sin límites e intentos ilimitados en niveles legendarios.

33. ¿Qué es el Plan Familiar de Súper Duolingo?

Permite compartir Súper Duolingo con hasta 6 miembros.

34. ¿Qué es Duolingo Max?

Incluye todo de Súper Duolingo y funciones con IA: Explica mi respuesta, Juego de roles y Videollamadas con Lily.

35. ¿Cómo cancelo mi suscripción a Súper Duolingo?

Desde la plataforma de compra (Google Play, Apple). Eliminar la app o tu cuenta no cancela la suscripción.

36. ¿Puedo pedir un reembolso de mi suscripción?

En general no. Depende de Apple o Google Play.

37. ¿Cómo uso un código promocional en Duolingo?

Entra a duolingo.com/redeem, introduce el código y sigue los pasos.

38. ¿Se puede aprender lengua de señas en Duolingo?

No. Se recomienda recurrir a clases locales o recursos online.

39. ¿Duolingo añadirá más idiomas?

En 2022 se añadieron criollo haitiano y zulú. Actualmente se enfocan en mejorar los cursos existentes y alinearlos con marcos CEFR/ACTFL.

40. ¿Se puede aprender completamente un idioma solo con Duolingo?

No. Es un gran inicio, pero no reemplaza la práctica real y el estudio adicional.

41. ¿Alguien logró fluidez solo con Duolingo?

Algunos usuarios han alcanzado un nivel B2 en idiomas como español o francés, pero la mayoría necesita apoyo adicional.

42. ¿Qué planes ofrece Duolingo?

Tres: Gratis, Súper y Súper Familia.

43. ¿Qué incluye el plan Gratis?

Acceso al contenido de aprendizaje básico de los cursos (con anuncios y sin las ventajas de Súper).

44. ¿Qué incluye Súper Duolingo?

Sin anuncios, vidas ilimitadas, práctica de habilidades, repaso de errores, desafíos sin límites y Nivel legendario sin límites.

45. ¿Qué incluye Súper Familia?

Todos los beneficios de Súper Duolingo para hasta 6 usuarios.

46. ¿Puedo probar Súper Duolingo gratis?

Sí, hay prueba gratis de 1 semana.

47. ¿Cómo funciona la prueba gratis de Súper?

Día 0: acceso completo · Día 5: recordatorio · Día 7: se cobra si no cancelas.

48. ¿Cuándo me cobran después de la prueba?

En el día 7. Debes cancelar al menos 24 horas antes para evitar el cargo.

49. ¿Hay compromiso de permanencia?

No. Puedes cancelar en cualquier momento.

50. ¿Cancelar la app cancela mi suscripción?

No. Debes cancelarla desde la tienda donde la compraste.

51. ¿El Plan Familiar afecta mi progreso o racha si entro o salgo?

No, tu progreso y racha se mantienen.

52. ¿Cuántas personas pueden usar Súper Familia?

Hasta 6 usuarios.

53. ¿Súper Duolingo tiene anuncios?

No. Solo el plan Gratis incluye anuncios.

54. ¿Mi suscripción ayuda a mantener Duolingo gratuito?

Sí. Tu suscripción financia la educación gratuita para millones de usuarios.

55. ¿En qué plataformas puedo usar Duolingo y gestionar mi suscripción?

En iOS (App Store) y Android (Google Play).

56. ¿Cuántos cursos hay disponibles en Duolingo?

Más de 100 cursos de idiomas.

57. ¿Qué beneficios de Súper me ayudan a progresar más rápido?

Vidas ilimitadas, práctica personalizada, repaso de errores, nivel legendario sin límites y sin anuncios.

Conclusiones

El desarrollo de esta práctica me permitió comprender de manera mas clara cómo funciona un sistema conversacional basado en reglas, este chatbot no aprende automáticamente de ejemplos, sino que depende de expresiones regulares y una maquina de estados para guiar la conversación. El hecho de normalizar entradas y trabajar con raíces de palabras permitió que el programa reconociera diferentes formas de una misma consulta lo que hizo al sistema mas flexible dentro de sus límites.

Considero que la experiencia fue útil porque me mostro la forma de como un chatbot se puede construir sin la necesidad de modelos complejos, simplemente organizando reglas claras y un flujo de estados bien definidos para lograr un sistema determinista.