

Identificación de etiquetas semánticas para su uso en diálogos

Andrés Vázquez, David Pinto, Darnes Vilaríño

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación,
México

{andrex,dpinto,darnes}@cs.buap.mx

Resumen. En la interacción humano-robot la generación automática de diálogos es una parte importante. Los diálogos generados deben garantizar una conversación coherente entre el humano y el robot, se espera que la interacción sea lo más natural y eficaz que se pueda. En este trabajo se plantea el uso de entidades semánticas como elementos básicos que posteriormente serán utilizados para la realización de un módulo de comprensión del lenguaje como parte de un sistema de diálogos. Para poder identificar las entidades semánticas en los diálogos se evaluaron dos herramientas para reconocer entidades nombradas (NER), que utilizan como núcleo un clasificador basado en Campos Aleatorios Condicionales. Los resultados que se obtienen en este trabajo permiten afirmar que estos dos NER tienen buenos resultados en el reconocimiento de las entidades semánticas (tokens) pues de manera global superan el 84% de exactitud. El corpus utilizado en este trabajo es el corpus español DIHANA, el cual está compuesto de diálogos sobre un sistema de información de consultas telefónicas sobre horarios y precios de trenes de largo recorrido.

Palabras clave: sistema de diálogos, entidad semántica, reconocedor de entidad nombrada.

Identification of Semantic Tags for Use in Dialogues

Abstract. In the human-robot interaction the automatic generation of dialogues is an important part, the generated dialogues must guarantee a coherent conversation between the human and the robot, the interaction is expected to be as natural and effective as possible. In this paper, the use of semantic entities as basic elements is proposed, which will later be used for the realization of a language understanding module as part of a dialog system. To identify the semantic entities in the dialogues, two tools for recognizing named entities (NER) were evaluated, using as a nucleus a classifier based on Conditional Random Fields. The results obtained in this work allow us to affirm that these two NERs have satisfactory results in the recognition of semantic entities (tokens) because, overall, they exceed 84% accuracy. The corpus used in this

work is the Spanish DIHANA corpus, which is composed of dialogues about an information system of telephone queries about schedules and prices of long-distance trains.

Keywords: dialog system, semantic entity, named entity recognizer.

1. Introducción

A medida que mejora la comprensión del lenguaje y la tecnología de generación automática de diálogos, también aumenta el interés en la construcción de sistemas de conversación de usuario, que pueden ser utilizados para una variedad de aplicaciones tales como planificación de viajes, sistemas tutoriales o soporte técnico basado en chat. La descripción de algunos de estos sistemas desarrollados en los últimos años se puede encontrar en [1–8].

Los sistemas de diálogo hablado o sistemas conversacionales son una tecnología concebida para facilitar la interacción natural mediante el habla, entre una persona y una computadora, son una interfaz hombre-máquina capaz de reconocer y comprender una entrada hablada y reproducir una salida oral como respuesta. Los sistemas de diálogo consisten de una estructura modular en la que cada módulo se ocupa de determinadas tareas en interacción con todos los demás módulos, normalmente los módulos que conforman a un sistema de diálogo son: el reconocedor del habla, el módulo de comprensión del lenguaje, el gestor del diálogo y el módulo de generación de respuesta.

En este trabajo nos enfocamos únicamente al módulo de comprensión del lenguaje, en donde se maneja el concepto de entidad semántica y cómo reconocer estas entidades semánticas a través del uso de reconocedores de entidades nombradas (Named Entity Recognizer, NER) [9] y posteriormente sea más fácil el proceso de comprensión del lenguaje en la interpretación semántica (secuencia de unidades semánticas) y que en un futuro forme parte de un sistema de diálogo del tipo pregunta-respuesta.

El presente trabajo está estructurado de la siguiente manera: en la sección 2 se describe el corpus que se utilizó para hacer la prueba de los NER; en la sección 3 se describe el proceso general de un NER y se presentan los resultados obtenidos de dos herramientas NER utilizadas en este trabajo; finalmente, en la sección 4 se presentan las conclusiones y el trabajo a futuro.

2. Estudio de caso corpus DIHANA

Con la finalidad de evaluar la identificación de entidades semánticas en el uso de diálogos se utilizó el corpus DIHANA [10]. El corpus en español DIHANA está compuesto por 900 diálogos sobre un sistema de información de consultas telefónicas sobre horarios y precios de trenes de largo recorrido. Fue adquirido por 225 hablantes

Tabla 1. Ejemplo de algunos diálogos del corpus DIHANA.

No.	Diálogo
U0000	hola buenos días mira quería saber horario de trenes para ir a cuenca
U0001	sí que quería saber horarios de trenes para ir a cuenca
U0002	pues quiero salir el día treinta de junio
U0003	pues no gracias
U0004	hola buenos días quería saber horarios para ir a Barcelona
U0005	pues quiero salir el treinta de julio
U0006	pues que quiero ir el treinta de julio
U0007	sí efectivamente y quisiera ir en un tren que fuese rápido tengo que estar allí antes de las ocho de la tarde vamos
U0008	quisiera saber el horario de un tren que llegue allí antes de las ocho

U0000:hola buenos días mira quería saber horario de trenes para ir a cuenca	
hola buenos días:	cortesía
mira quería saber:	consulta
horario de trenes para ir:	<hora>
a cuenca:	ciudad_destino
U0001:sí que quería saber horarios de trenes para ir a cuenca	
sí:	<afirmacion>
que:	nada
quería saber:	consulta
horarios de trenes para ir:	<hora>
a cuenca:	ciudad_destino

Fig. 1. Dos diálogos y sus correspondientes entidades semánticas.

diferentes (153 hombres y 72 mujeres). Hay 6,280 turnos de usuario y 9,133 turnos del sistema. El tamaño del vocabulario es de 823 palabras. La cantidad total de señal de voz fue de aproximadamente cinco horas y media (véase la tabla 1).

La Fig. 1 muestra dos diálogos y sus correspondientes entidades semánticas.

La adquisición del corpus DIHANA se llevó a cabo por medio de un prototipo inicial, utilizando la técnica del Mago de Oz (WoZ). Esta adquisición solo se restringió a nivel semántico (es decir, los diálogos adquiridos están relacionados con un dominio de tareas específico) y no se restringió a nivel léxico y sintáctico (habla espontánea). En este proceso de adquisición, el control semántico fue proporcionado por la definición de escenarios que el usuario tenía que cumplir y por la estrategia WoZ, que define el comportamiento del sistema de adquisición.

3. Evaluación de reconocedores de entidades nombradas (NER)

En este trabajo se emplean dos reconocedores de entidades nombradas (NER) para reconocer entidades semánticas en los diálogos del corpus DIHANA y esto nos ayude en el proceso de comprensión del lenguaje en la interpretación semántica (secuencia de unidades semánticas) de los diálogos.

3.1. Descripción general del proceso de un NER

En el procesamiento de lenguaje natural, el reconocimiento de entidad nombrada es una tarea de extracción de información que busca ubicar y clasificar elementos en texto en categorías predefinidas, tales como personas, organizaciones, lugares, expresiones de tiempo y cantidades entre otros. Por ejemplo, en el siguiente texto [9]:

“Jim compró 300 acciones de Acme Corp. en 2006”

al aplicarle un NER se obtiene que se reconocen las entidades persona, organización y tiempo:

“[Jim] (persona) compró 300 acciones de [Acme Corp.] (organización) en [2006] (tiempo).”

El proceso de un NER normalmente consta de tres etapas:

Primera etapa. Corresponde a la preparación de los datos de entrenamiento, la cual se describe gráficamente en la Fig. 2, en donde inicialmente se tiene un corpus identificado como un “Archivo plano”, al cual se le aplica un método de “tokenización” y posteriormente se realiza el “etiquetado” del corpus de acuerdo con los requerimientos de cada NER.

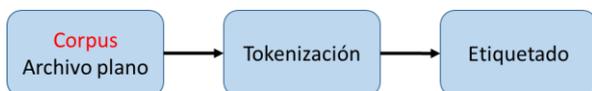


Fig. 2. Preparación de los datos de entrenamiento para un NER.

Segunda etapa. Esta etapa corresponde a la etapa de entrenamiento del NER, en donde se utiliza parte del corpus generado en la primera etapa (Training) y como resultado se obtiene un modelo de NER entrenado, el cual se describe en la Fig. 3.

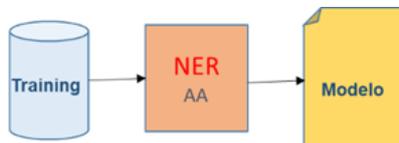


Fig. 3. Etapa de entrenamiento de un NER.

Tercera etapa. Corresponde a la evaluación del modelo obtenido en la etapa anterior y es aquí donde se ve el desempeño del modelo, verificando la precisión de las entidades semánticas identificadas, como se muestra en la Fig. 4.

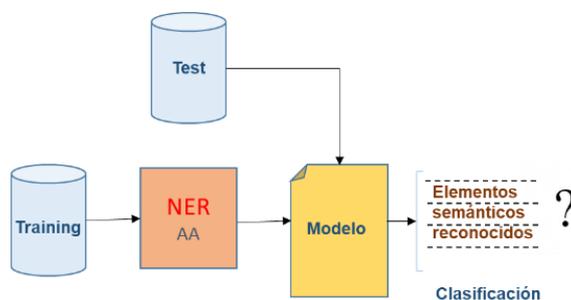


Fig. 4. Etapa de evaluación de un NER.

Tabla 2. Descripción de los archivos de entrenamiento y prueba.

Característica	Archivo de entrenamiento	Archivo de prueba
Intervenciones	6279	4878
Etiquetas	30	28

Tabla 3. Precisión obtenida en la identificación de tokens.

NER	No. Tokens	Precisión
Stanford	10,485	0.84
Xingdi (Eric) Yuan	10,485	0.89

3.2. Resultados experimentales

Las dos herramientas NER que se utilizaron en este trabajo fueron: el NER de Stanford [11] y el NER de Xingdi (Eric) Yuan [12]. Estas dos herramientas utilizan el modelo de secuencia de campos aleatorios condicionales (CRF) para la clasificación.

En ambos modelos se utilizaron los mismos archivos de entrenamiento y de prueba del corpus DIHANA, los cuales se describen en la Tabla 2.

Las intervenciones en cada archivo son diferentes, así como las etiquetas que se describen son etiquetas diferentes que pueden aparecer más de una vez en los archivos de entrenamiento y de prueba.

Los resultados globales de la precisión obtenida en los dos NER se muestran en la Tabla 3, aquí la precisión es el número de tokens de un tipo dado que el modelo identificó correctamente, entre el número total de tokens que el modelo predijo que era de ese tipo.

Si bien, se puede observar que el etiquetador de entidades nombradas Xingdi (Eric) Yuan obtiene el mejor resultado global, es muy importante analizar el resultado obtenido para cada una de las clases semánticas clasificadas. Así, en la Tabla 4, se han incluido dichos resultados, los cuales procedemos a analizar a continuación.

CRFClassifier tagged 10485 words in 1 documents at 141.85 words per second.

Entity	P	R	F1	TP	FP	FN
<afirmacion>	0.9730	0.9665	0.9698	433	12	15
<duracion>	0.0000	0.0000	0.0000	0	0	2
<hora>	0.8152	0.8113	0.8132	172	39	40
<hora_llegada>	0.9412	0.9412	0.9412	16	1	1
<hora_salida>	0.8571	0.6667	0.7500	6	1	3
<negacion>	0.8561	0.9154	0.8848	238	40	22
<no_entendido>	0.5000	0.1429	0.2222	1	1	6
<precio>	0.7429	0.7280	0.7354	182	63	68
<tipo_tren>	0.5833	0.7000	0.6364	7	5	3
albacete	0.0000	1.0000	0.0000	0	1	0
barcelona	1.0000	1.0000	1.0000	2	0	0
ciudad	0.7778	0.6563	0.7119	21	6	11
ciudad_destino	0.9789	0.9754	0.9772	278	6	7
ciudad_origen	0.9872	0.9625	0.9747	154	2	6
clase_billete	0.7000	0.6364	0.6667	21	9	12
coletilla	0.7612	0.7846	0.7727	102	32	28
consulta	0.8003	0.8082	0.8042	573	143	136
cortesía	0.8537	0.8140	0.8333	105	18	24
cã;ceres	0.0000	0.0000	0.0000	0	0	1
cã;diz	0.0000	0.0000	0.0000	0	0	1
fecha	0.9263	0.9337	0.9300	352	28	25
gasteiz	0.0000	0.0000	0.0000	0	0	1
granada	0.0000	1.0000	0.0000	0	2	0
hora	0.8750	0.8537	0.8642	175	25	30

Fig. 5. Inicio de la ejecución del NER Stanford.

ciudad_destino	0.9789	0.9754	0.9772	278	6	7
ciudad_origen	0.9872	0.9625	0.9747	154	2	6
clase_billete	0.7000	0.6364	0.6667	21	9	12
coletilla	0.7612	0.7846	0.7727	102	32	28
consulta	0.8003	0.8082	0.8042	573	143	136
cortesía	0.8537	0.8140	0.8333	105	18	24
cã;ceres	0.0000	0.0000	0.0000	0	0	1
cã;diz	0.0000	0.0000	0.0000	0	0	1
fecha	0.9263	0.9337	0.9300	352	28	25
gasteiz	0.0000	0.0000	0.0000	0	0	1
granada	0.0000	1.0000	0.0000	0	2	0
hora	0.8750	0.8537	0.8642	175	25	30
la_coruña	1.0000	0.5000	0.6667	1	0	1
lleida	0.0000	0.0000	0.0000	0	0	1
logroño	0.0000	1.0000	0.0000	0	0	1
m_llegada	0.6962	0.7051	0.7006	55	24	23
m_salida	0.8492	0.8492	0.8492	107	19	19
madrid	0.0000	1.0000	0.0000	0	1	0
mañana	0.0000	0.0000	0.0000	0	2	1
no	1.0000	1.0000	1.0000	83	0	0
nombre_atributo	0.0000	0.0000	0.0000	0	0	1
not	0.0000	0.0000	0.0000	0	2	9
numero_relativo_orden	0.6250	0.7143	0.6667	5	3	2
pamplona	0.0000	0.0000	0.0000	0	0	1
precio	0.6667	0.4000	0.5000	2	1	3
pues	0.0000	0.0000	0.0000	0	0	1
san_sebastiã;n	0.0000	1.0000	0.0000	0	1	0
sã-	1.0000	1.0000	1.0000	131	0	0
tarragona	1.0000	0.5000	0.6667	1	0	1
tipo_tren	0.6458	0.6596	0.6526	62	34	32
tipo_viaje	0.5067	0.5135	0.5101	76	74	72
valladolid	0.0000	0.0000	0.0000	0	0	1
viernes	0.0000	0.0000	0.0000	0	0	1
Totals	0.8494	0.8462	0.8478	3361	596	611

andrex@arcturus:~/stanford-ner-2017-06-09\$ █

Fig. 6. Final de la ejecución del NER Stanford.

Tabla 4. Resultados obtenidos de la evaluación de los dos NER.

Etiquetas semánticas	Frecuencia Training	Frecuencia Test	Exactitud NER Stanford	Exactitud NER Xingdi (Eric) Yuan
<hora_llegada>	73	73	94.12%	100.00%
fecha	1678	1655	92.63%	98.63%
ciudad_destino	611	601	97.89%	98.36%
<afirmacion>	541	527	97.30%	97.41%
hora	978	949	87.50%	97.03%
ciudad_origen	323	310	98.72%	95.98%
<precio>	917	866	74.29%	94.44%
<hora>	618	578	81.52%	93.53%
consulta	1935	1758	80.03%	90.85%
<negacion>	344	302	85.61%	87.79%
numero_relativo_orden	24	21	62.50%	87.50%
m_salida	186	159	84.92%	85.48%
cortesia	253	215	85.37%	84.98%
m_llegada	136	115	69.62%	84.56%
coletilla	225	188	76.12%	83.56%
tipo_viaje	395	328	50.67%	83.04%
<tipo_tren>	65	52	58.33%	80.00%
tipo_tren	246	187	64.58%	76.02%
clase_billete	94	70	70.00%	74.47%
ciudad	44	30	77.78%	68.18%
<hora_salida>	38	20	85.71%	52.63%
precio	15	6	66.67%	40.00%
O	471	171	34.15%	36.31%
not	12	2	0.00%	16.67%
<duracion>	8	0	0.00%	0.00%
<no_entendido>	22	0	0.00%	0.00%
nombre_atributo	4	0	0.00%	0.00%

Las etiquetas semánticas “<hora_llegada>”, “fecha”, “ciudad_destino”, “<afirmación>”, “hora” y “ciudad_origen” son las que han obtenido el mayor grado de exactitud durante el proceso de identificación automática con un valor que supera el 95%.

Consideramos que este resultado se encuentra fundamentado en dos cosas: primeramente, en la cantidad de datos usados durante la fase de entrenamiento, pero también a la inherente naturaleza de las palabras del lenguaje natural asociadas a dichas entidades semánticas, pues existe cierta uniformidad para expresar, horarios y fechas, y en cuanto a los nombres de las ciudades, se considera que el número es ciertamente limitado, lo cual facilita el proceso de identificación.

Aun así, existen diferencias importantes entre los dos NER evaluados, que pueden ser vistas a detalle en la Tabla 4. La conclusión obtenida es que el NER Xingdi (Eric) Yuan es más estable que el proporcionado por Stanford.

4. Conclusión y trabajo futuro

El objetivo principal de este trabajo fue el de evaluar la tarea de comprensión del lenguaje, asociada a la generación de diálogos en lenguaje natural. Hemos utilizado el concepto de reconocimiento de entidades nombradas para hacer frente al problema de traducción de sentencias del lenguaje natural a entidades semánticas. Se evaluaron dos sistemas que utilizan como núcleo un clasificador basado en Campos Aleatorios Condicionales. Los resultados que se obtienen en el presente trabajo permiten afirmar que estos dos NER tienen buenos resultados en el reconocimiento de las entidades semánticas (tokens) pues de manera global superan el 84% de exactitud. En particular, es el NER Xingdi (Eric) Yuan el que produce los resultados más confiables y estables, por lo que consideramos importante utilizarlo próximamente durante la fase de comprensión del lenguaje.

Como trabajo futuro se tiene contemplado tomar las entidades semánticas reconocidas y representar los diálogos mediante inferencia gramatical con la finalidad de inducir una gramática de los componentes semánticos que apoye en la generación de gramáticas más complejas que incluyan el concepto de pregunta-respuesta y empatamiento de contenido con su correspondiente generación de respuesta validada.

También se tiene contemplado probar otros dos NER con el enfoque de redes neuronales recurrentes (LSTM) y convolucionales con la finalidad de verificar si su rendimiento es mejor que aquellos basados en campos aleatorios condicionales.

Agradecimientos. Este trabajo de investigación ha sido parcialmente respaldado por la beca CONACyT # 80286, bajo el Programa de Doctorado en Ingeniería del Lenguaje y del Conocimiento (LKE) de la Benemérita Universidad Autónoma de Puebla.

Referencias

1. Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T., Hetherington L.: JUPITER: A Telephone-Based Conversational Interface for Weather Information. *IEEE Transactions on Speech and Audio Processing*, 8(1) (2000)
2. Glass, J., Weinstein, E.: Speech builder: facilitating spoken dialog system development. In: *Proceedings of EUROSPEECH*, pp. 1335–1338 (2001)
3. Córdoba, R., San-Segundo, R., Montero, J.M., Colás, J., Ferreiros, J. Macias-Guarasa, J., Pardo, J.M.: An interactive directory assistance service for Spanish with large vocabulary recognition. In: *Proceedings of EUROSPEECH*, pp. 1279–1282 (2001)
4. Hurtado, L., Blat, F., García, F., Grau, S., Griol, D., Sanchis, E., Segarra, E., Torres, F.: Sistema de diálogo para el Proyecto DIHANA. *Procesamiento del Lenguaje Natural*, 35, pp. 453–454 (2005)
5. Griol, D., Hurtado, L., Sanchis, M., Segarra, E.: Dos aproximaciones basadas en reglas para la gestión del diálogo. *Procesamiento del Lenguaje Natural*, 35, pp. 213–220 (2005)
6. Pietquin, O.: Inverse Reinforcement Learning for Interactive Systems. *ACM (MLIS'13)* (2013)
7. Lison, P.: Model-based Bayesian Reinforcement Learning for Dialogue Management (2013)

8. Barlier, M., Perolat, J., Laroche, R., Pietquin, O.: Human-Machine Dialogue as a Stochastic Game. In: Proceedings of the SIGDIAL 2015 Conference, Prague, Czech Republic. pp. 2–11 (2015)
9. WIKIPEDIA: https://en.wikipedia.org/wiki/Named-entity_recognition (2018)
10. Benedí, J., Lleida, E., Varona, A., Castro, M., Galiano, I., Justo, R., López de Letona, I., Miguel, A.: Design and acquisition of a telephone spontaneous speech dialogue corpus in spanish: Dihana. In: Fifth International Conference on Language Resources and Evaluation (LREC), pp. 1636–1639 (2006)
11. NER Stanford: <https://nlp.stanford.edu/software/CRF-NER.html> (2018)
12. NER Xingdi: <https://github.com/xingdi-eric-yuan/nerpp> (2018)