

Gramáticas formales para el lenguaje natural

Curso 2021

Laboratorio

En esta tarea de laboratorio se analizarán los posibles aportes que puede brindar el análisis sintáctico para el problema de búsqueda de respuestas (Question Answering - QA).

Se trabajará en el análisis de diferentes ejemplos desde un punto de vista teórico, se realizarán pruebas con analizadores sintácticos disponibles para el español y, finalmente, se implementarán heurísticas que busquen mejorar la performance de un enfoque basado en modelos de lenguaje neuronales. Los ejemplos se incluyen al final de este documento.

En este trabajo se buscan respuestas a preguntas dentro de fragmentos de texto cortos, en algunos casos de una sola oración. Asumimos que estos fragmentos de texto se obtienen en una etapa previa de extracción de fragmentos relevantes para las preguntas que se quiere responder. El objetivo del trabajo es encontrar, dentro del fragmento dado, una respuesta a la pregunta, si existe alguna. El problema se define entonces como: dados un fragmento de texto F y una pregunta P , devolver el substring de F que contiene la respuesta a P , o un indicador de que no existe respuesta.

Parte a) Para cada par pregunta/texto (P_i/F_i) obtenga las respuestas que ofrece el modelo para QA entrenado sobre el modelo de lenguaje neuronal BETO (adaptación de BERT al español) utilizando el corpus SquAD (adaptación al español).

Parte b) Analice las respuestas indicando en qué casos considera que un análisis sintáctico podría mejorar los resultados. ¿Qué tipo de gramática considera más apropiada para lograr esto?

Parte c) Pruebe algunos ejemplos con los siguientes parsers, que se encuentran disponibles para el español: parser de constituyentes de Freeling, parser de dependencias de Freeling, parser basado en HPSG Parsur.

Parte d) Implemente un módulo que tome las salidas del modelo QA basado en BETO+SQuAD y aplique heurísticas basadas en información sintáctica para mejorar los resultados.

Recursos disponibles

Modelo entrenado para la tarea QA en español y ejemplo de uso:

https://colab.research.google.com/drive/1F-SK1uSTJOYWFIuS_tdMPj4Vns7FT3Ea

Ejemplo de uso del parser de dependencias de Spacy:

https://colab.research.google.com/drive/1FQ71hIb7ni2q1Dn8OwdO_I_fHcX6T-yT#scrollTo=nQiBjhtZTbdA

Ejemplos para analizar

Para cada bloque se deben probar todos los pares (pregunta, fragmento).

BLOQUE 1

P1.1: ¿Quién comió la manzana?

P1.2: ¿Qué comió el niño?

F1.1: El niño comió la manzana.

F1.2: La manzana fue comida por el niño.

F1.3: El niño comió la naranja.

F1.4: El perro corrió al niño que había comido la manzana.

F1.5: La abuela estaba buscando la manzana pero se la había comido el niño.

BLOQUE 2

P2.1: ¿A quiénes se les aplicará un descuento en el salario?

F2.1: Los trabajadores del sector público que tengan un sueldo líquido por encima de los 80.000 pesos tendrán descuentos escalonados de entre 5% y 20%, según el monto, que serán aplicados a su sueldo nominal.

F2.2: Los trabajadores del sector público tendrán descuentos escalonados de entre 5% y 20%, según el monto, que serán aplicados a su sueldo nominal.

F2.3: Los trabajadores del sector público no tendrán descuentos, pero los privados tendrán descuentos escalonados de entre 5% y 20%, según el monto, que serán aplicados a su sueldo nominal.

F2.4: Se evitará que los trabajadores del sector público tengan descuentos, pero los privados los tendrán.

F2.5: Los descuentos se aplicarán a los trabajadores del sector público que tengan un sueldo líquido por encima de los 80.000 pesos, pero no se aplicarán a los del sector privado.

BLOQUE 3

P3.1: ¿Quiénes aplicarán un descuento al salario?

F3.1: Los trabajadores del sector público tendrán descuentos en sus salarios, que serán aplicados por el gobierno y las intendencias.

F3.2: El gobierno y las intendencias aplicarán descuentos en los salarios a los trabajadores del sector público que tengan un ingreso nominal mayor a 80.000 pesos.

BLOQUE 4

P4.1: ¿Cuándo vence el plazo para presentarse al llamado de la ANII?

P4.2: ¿Qué monto financia la ANII?

P4.3: ¿Cuál fue el objetivo del desafío de la semana pasada?

P4.4: ¿Cuál es el objetivo del nuevo desafío?

F4.1: ANII lanzó ahora línea concursable para financiar el diseño y producción de respiradores para atender a pacientes con Covid-19. El llamado, en conjunto con el MIEM y Plan Ceibal, financia con hasta seis millones de pesos la fabricación de ventiladores pulmonares mecánicos para atender a 50 pacientes en cuidados intensivos que deberán estar en funcionamiento en mayo. Así como la semana pasada la Agencia Nacional de Investigación e Innovación (ANII) anunció el lanzamiento de "un desafío ANII", con el objetivo de financiar "el desarrollo y elaboración de un Kit de Diagnóstico del Covid-19" que deberían estar

"disponibles y en uso antes de mayo de 2020", este lunes se emitió el comunicado de un nuevo desafío, esta vez destinado al desarrollo y producción de ventiladores pulmonares mecánicos. Nuevamente, se trata de un llamado relámpago, ya que el plazo para presentarse vence el 31 de marzo, y está dirigido a "empresas y organizaciones que estén radicadas en el país, de forma individual o asociadas a otras organizaciones nacionales o extranjeras como empresas, centros tecnológicos o institutos de investigación públicos o privados". El desafío "planteado por el MIEM y ANII consiste en el diseño y producción de respiradores para atender a 50 personas", sostienen, y aclaran que deben ser "de fabricación rápida, completos, en funcionamiento comprobado y avalados por los estándares de las organizaciones médicas correspondientes (por ejemplo, usando impresoras 3D entre otras tecnologías 4.0)". El objetivo del llamado es claro: "se busca colaborar a la falta de respiradores en los CTI de los centros de salud del país para los pacientes graves afectados por coronavirus", razón por la cual se comunicarán los resultados el 3 de abril, "para que la producción comience inmediatamente después.

BLOQUE 5

P5.1: ¿En dónde empezó la pandemia?

P5.2: ¿Cuál es la capital de Hubei?

P5.3: ¿Cuándo empezó la pandemia?

P5.4: ¿Qué informaron los funcionarios de salud?

P5.5: ¿Quién informó sobre los pacientes de neumonía?

F5.1: Tengo un profuso trabajo y estudio de catástrofes globales, especialmente de origen astronómico, pero no experiencia en el análisis de pandemias. Desde esa perspectiva he querido hacer una contribución que intenta dar visualización a datos disponibles en la red, para poner en evidencia las diferentes tendencias en la evolución de la Covid-19 y utilizar mi experiencia para extraer algo de ellas. La epidemia de la Covid-19 comienza a desarrollarse en la ciudad de Wuhan, capital de la provincia de Hubei, China, sobre fines de diciembre de 2019. La primera aparición pública ocurre el 31 de ese mes, cuando funcionarios de salud chinos informaron a la Organización Mundial de la Salud (OMS) sobre un grupo de 41 pacientes con una misteriosa neumonía. La mayoría de los pacientes están conectados al mercado mayorista de mariscos de Huanan.