Hola, soy Miguel Ángel Navarro Arenas, estudiante de SAR. Veremos el apartado dos, respuestas a preguntas basadas en el conocimiento.

Gran parte de la información se encuentra codificada en texto web, pero podemos extraer su información de maneras más estructuradas para responder preguntas, las cuales utilizan bases de datos. Un sistema basado en este tipo de respuestas presenta dos problemas: traducir el lenguaje natural en un formato lógico entendible por los sistemas informáticos y contestar a estas preguntas de manera correcta

Como podemos ver en el ejemplo, a la izquierda tendríamos la representación natural de las preguntas y a la derecha cómo serían interpretadas por la máquina. Los analizadores semánticos son los encargados de realizar este cambio, los cuales están relacionados con la base de datos en la que se almacena la información.

Podemos expresar las bases de datos de manera simples o como tripletas formadas por un predicado y dos argumentos más, que expresan una relación con el predicado. Las más sencillas son las formadas por sujeto, predicado y objeto.

**Métodos basados en reglas**

Sencillamente, en aquellas consultas que sean más frecuentes, utilizaremos reglas escritas a mano. En estas reglas tendremos relaciones entre algunos patrones. Por ejemplo, sabemos que siempre cuando preguntamos quien nos vamos a referir a una persona que realiza una acción que va a ir después de la cláusula “Quien”.

**Métodos supervisados**

Relacionado con lo anterior, tenemos que existe la posibilidad de que haya preguntas emparejadas, es decir, que estén relacionadas. Nuestra misión es crear tuplas y entrenarlas, para producir un sistema de análisis de preguntas, de manera que primero se analizan las preguntas y, posteriormente, se crea un árbol con la forma más lógica.

A través de unas preguntas, las wh questions del inglés y un sujeto, podemos establecer relaciones, como por ejemplo la que veremos a continuación: birth-year, en la que el sujeto es Ada Lovelace, aparece la cláusula wh y el nombre de dicha relación. A través de ella, creamos una base para establecer las siguientes preguntas que tengan el mismo modelo, pudiendo contestarlas todas a través de la relación que creemos en la base de datos. Esta relación nos sirve como modelo para todas aquellas preguntas que tengan un was born y una cláusula wh, haciendo un entrenamiento del modelo podremos obtener muchas más respuestas. En caso de concretar una estructura similar pero que no tenga un resultado en nuestra base de datos, se creará una nueva entrada en esta.

También, contaremos las apariciones de cada una de las preguntas y respuestas y mediante el feedback, sabremos cuales son las respuestas a las preguntas que más han sucedido y, por tanto, las que son más cercanas a lo correcto.

**Métodos semi-supervisados.**

Las preguntas que se pueden responder con los métodos anteriores son, quizá, preguntas demasiado simples que no se adaptan tanto al habla humana natural. Por tanto, muchas de las técnicas de mapeo que hemos mencionado anteriormente se sobrecargan gracias a masivas cantidades de texto por redundancia. ¿Y cual es la mayor fuente de texto del mundo? Efectivamente, Internet, la cual muestra la gran mayoría de las diferentes formas de expresar la misma cosa.

Hay softwares de extracción de información que se encargan de encontrar una enorme cantidad de tripletas a lo largo de la web, ofreciendo una base de datos robusta en la que podemos encontrar muchas maneras diferentes de expresar una pregunta o, por el contrario, muchas respuestas a preguntas que tienen la misma forma. Muchas de estas se apoyan en fuentes como Wikipedia. Estas herramientas se encargan de alinear una String, como puede ser un nombre, a una, o varias, páginas de Wikipedia. En caso de haber varias, gracias a la distancia coseno entre la palabra y la página, obtenemos cual es la más indicada.

Debemos de alinear las consultas. Primero alineamos los argumentos y después el predicado, es decir, unimos el término o frase con su relación en la base de datos. En el caso del ejemplo del punto anterior, estábamos relacionando el “was born” con la relación birth-year.

A modo de resumen veremos un esquema conceptual de las funciones que realizan las herramientas que utilizan respuestas basadas en conocimiento.

Primero, traducimos la pregunta y vemos si hace pattern matching con algún segmento o relación que ya exista en nuestra base de datos. Si no coincide con ninguna, creamos una nueva, obtenemos los documentos en los que aparece y los listamos gracias a la similitud coseno. Después, listamos los segmentos que sean similares y gracias a eso inferimos una serie de respuestas. De entre ellas, cogemos la que más posibilidades de ser correcta tenga.