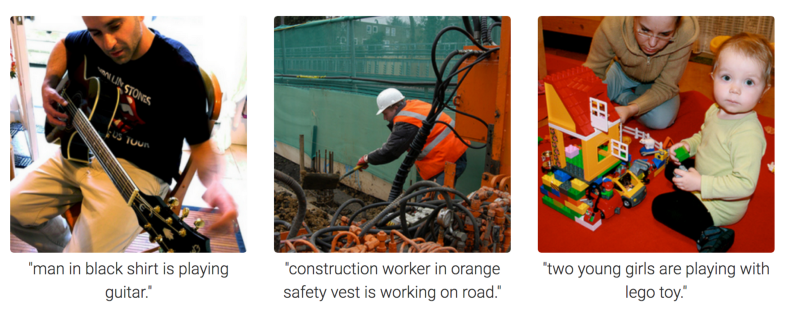
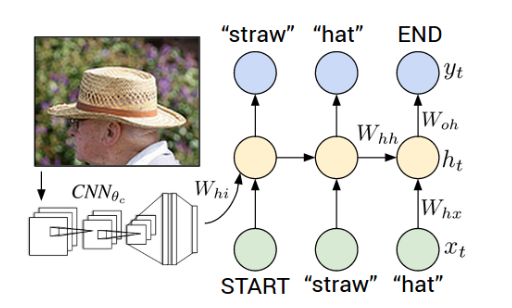
**Práctica final del módulo de Deep Learning**

**Objetivo**

El objetivo final de esta práctica es el de construir un programa que tenga como entrada una imagen, y de como salida una frase que la describa:



Para ello, la arquitectura más común es la de emplear una CNN para extraer las características, que luego se introducirán a una RNN que componga una frase que describa lo que sucede en la escena.



Al final, se puede entender como una arquitectura **Encoder-Decoder**, en la que el encoder es la CNN, y el decoder es la RNN.

Se valorará, más que la precisión, la explicación de por qué se ha optado por utilizar una arquitectura u otra y las decisiones tomadas.