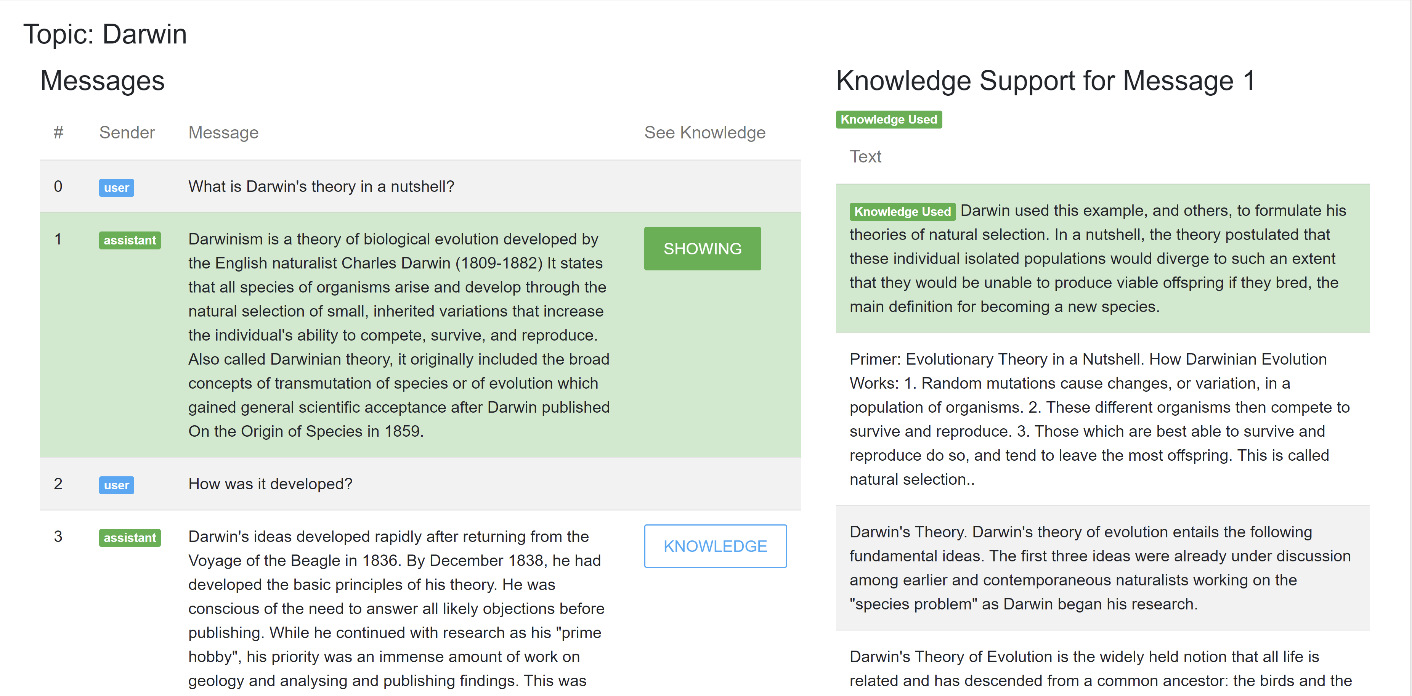


<https://www.overleaf.com/project/60b39609c650fc22d40b4e31>

Introdução



O trabalho desenvolvido no âmbito desta cadeira tem por base o tema “Open-Domain search assistant”. De facto, um dos objetivos principais é o estudo e desenvolvimento de um sistema (conversa com um chatbot) para geração de respostas sucintas e com um elevado nível de qualidade em detrimento da usual visualização de vários documentos (ex: pesquisa num search engine como o google). 

**Fig.1**: Example of such a system: “Answer Generation In Open-Domain Conversations”.

Efetivamente, foi necessário abordarmos o problema com base em 4 etapas iniciais. Inicialmente começámos pela preparação da groundtruth e do dataset, ou seja, extraímos várias perguntas de vários tópicos e as melhores 10 respostas a essas perguntas. Depois de completar a base dos nossos dados, passámos para a extração de entidades tanto das perguntas como das respostas (DBpedia Spotlight para entity linking), de seguida construção do grafo de entidades e aplicação do pageRank. Deste modo, é-nos também possível escolher quais as melhores passagens com base no pageRank. Por último, usamos o Bart para gerar a resposta única usando para tal as passagens referidas.

Assim, numa fase seguinte pretendemos proceder ao estudo das diferentes formas para gerar o grafo de entidades e comparar os resultados de cada uma delas. Para este efeito, a nossa abordagem inicial terá agora como foco a criação de um grafo dirigido com base nas ligações entre as entidades. Também, será interessante realizar comparações entre os vários modelos de criação de grafos encontrados e estudar o grafo ao longo de uma dada conversa.

Evaluation



1. **Dataset Description:**

Com efeito, o dataset sobre o qual operamos tem várias dimensões: existem 20 *conversas* sendo que cada uma destas é relativa a tópicos diferentes (ex: cancer, netflix, sports, etc.). Cada conversa é composta por 8 *perguntas* que têm associadas às mesmas 10 *respostas* de topo (adquiridas através de documentos da Wikipédia).

1. **Baselines:**

Temos de decidir onde colocar isto:

As hipóteses exploradas relativamente a quais as entidades de maior relevo a considerar foram: estudar apenas as entidades relativas à **última turn** (i.e entidades correspondentes à última query e as suas 10 passagens correspondentes). De seguida, considerar também as diferentes entidades de **todas as queries** da conversa corrente (queries de todas as turns) e por último utilizar ainda as entidades relativas às passagens da primeira turn.

Efetivamente, para além destas possibilidades também fizemos variar os “pesos” das entidades em diferentes situações como pode ser visualizado em baixo.

1. Extração dos tópicos e das perguntas e respostas
2. Extração das Entidades
3. Vetor Coleção das Entidades

Coleção das entidades

Conteúdo do vetor de posição das entidades

1. Construção da Matriz de entidades
2. Aplicação do Pagerank
3. Re-rank das Respostas
4. Summarization

------------------- Samuel ---------------

1. **Results Analysis:**

Critical discussion

Vimos que o método apenas da primeira pergunta não funcionou muito bem, pois a pergunta normalmente têm só 1 ou 2 entidades… Nos outros métodos viu-se que foi melhor...