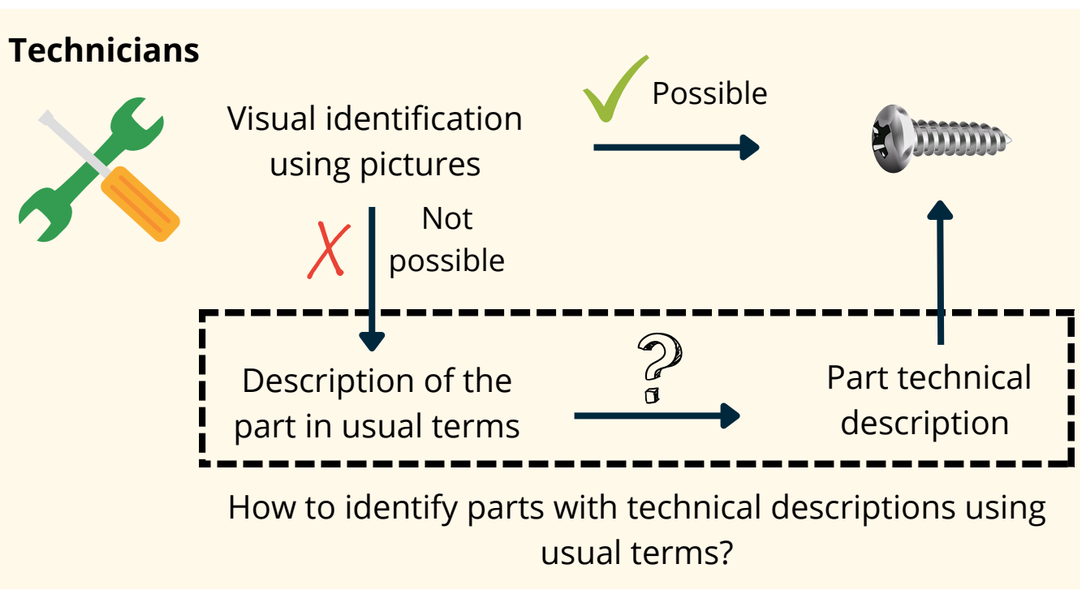
P8 Semantic Search Engine For Industrial Parts

Hiba Bentaleb (1), Wiam Adnan (1), Mohamed Ait Hajjoub (1), Aymane Hanine (1), Jeremie Brabet-Adonajlo (2), Redha Moulla (3)
(1) Final year data science student, Ecole Centrale Casablanca. (2) CEO of Pzartech . (3) Data science & AI consultant and professor
Tel. Wiam: +212 6 52 06 94 66 / Aymane : +212 6 37 92 54 46

1 - Introduction

Pzartech, a digital solution startup for industrial production and maintenance, aims to develop a smart search engine for it's client, a multi-national manufacturing concern that develops, produces and markets fertilizers, metals and other special-purpose chemical products. This project intends to help its maintenance teams identify parts successfully and efficiently and improve the fluidity of maintenance operations.

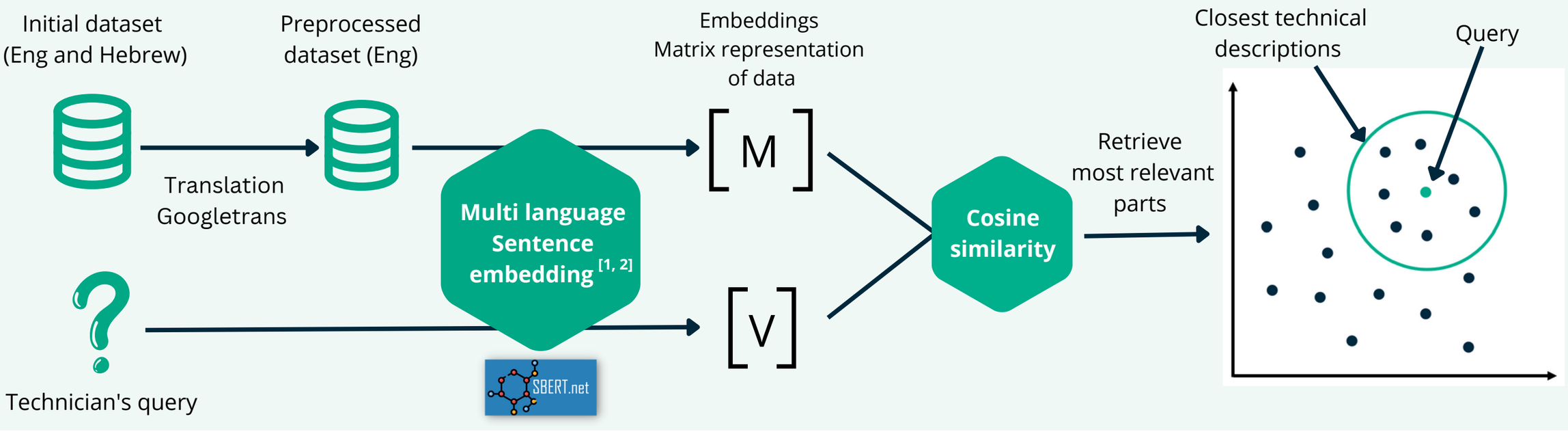
2 - Context



3 - Objective

- Search engine allowing technicians to find technical descriptions of maintenance piece from their usual terms
- Minimum Recall@10 of 70% to be achieved
- Easy deployement and intuitive interface

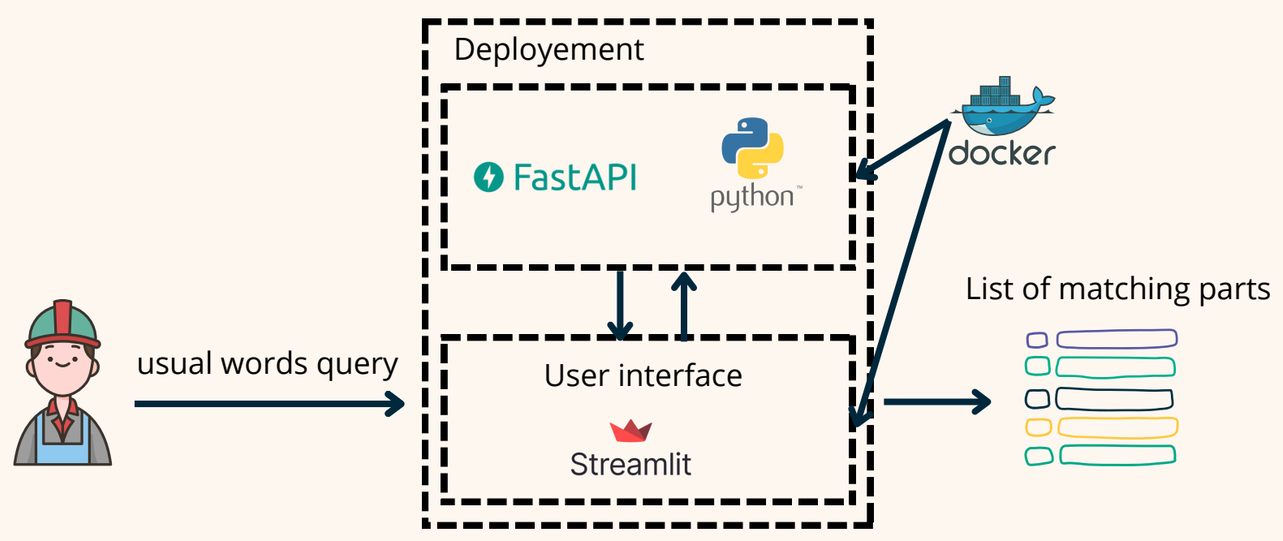
4 - Solution steps



5 - Test and results

- Create test dataset with technical descriptions and usual terms equivalents via google
- Choose **Recall@10** as the performance metric, i.e. the percentage of queries with corresponding description within 10 best results
- Obtain a Recall@10 of 83%

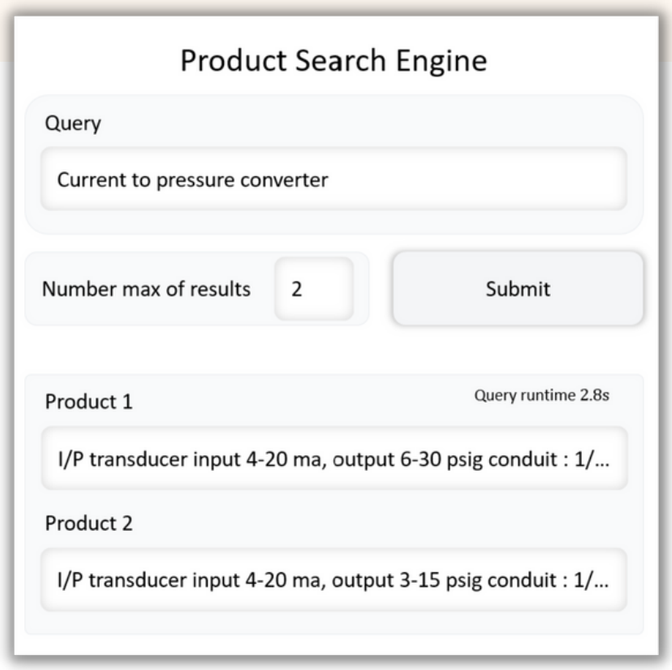
6 - PoC phase (Done)



7 - Pilot phase (Next step)

This step consists of a real time test of the engine to add some improvements:

- Collect query data from the technicians
- Collect the matching descriptions for each query
- Fine-tune the sentence transformer model that generates embeddings [1, 2]



8 - References

[1] Muennighoff, N. (2022, February 17). SGPT: GPT Sentence Embeddings for Semantic Search. arXiv.org. https://arxiv.org/abs/2202.08904
[2] Reimers, N. (2019, August 27). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. arXiv.org. https://arxiv.org/abs/1908.10084