

Hesaplamalı Anlambilim Dersi 1.Ödevi

Konusu: Sorular Cevaplar

https://huggingface.co/datasets/ytu-ce-cosmos/gsm8k_tr adresinde bulunan dosyadaki question ları soru, answer ları cevap olarak kullanınız.

Bu dosyadan rasgele 1000 soru ve o sorunun cevabı 2'lisi seçiniz.

Bir soruya en benzeyen 5 cevabı (1000 cevap içinden) bulup gerçek cevaba göre top1 ve top5 başarılarını hesaplayınız.

Bir cevaba en benzeyen 5 soruyu (1000 soru içinden) bulup gerçek soruya göre top1 ve top5 başarılarını hesaplayınız.

2 metin arasındaki benzerlik ölçümü için

<https://huggingface.co/spaces/mteb/leaderboard> adresinde ilk 100'e girmiş, multilingual olan ve 1 milyar parametreden küçük olan 5 model ve <https://huggingface.co/ytu-ce-cosmos/turkish-colbert> modeli kullanarak elde edeceğiniz temsiller arasındaki açı kullanılacaktır. 6 modelin başarıları top1 ve top5 e göre karşılaştırılacaktır. (7/10 puan)

Soru ve cevabı, bu temsiller üzerine tsne uygulanarak 2 boyutta türlerine (soru/cevap) göre renklendirilerek görselleştirilecektir (6 modelin her biri için 1 grafik). (3/10 puan)

Derste sunmak üzere bir sunum hazırlayınız. Yorumlarınız önemli.

Sınıfta sunum yapmayanlar bu ödevden en fazla %50 alabilirler.

Bu ödevde en çok 2 kişilik gruplar halinde çalışabilirsiniz.

Teslim edilecekler: Kod, sunum → 5 dk'lık sunumlar

Teslim şekli: online.yildiz.edu.tr

Son teslim tarihi: 24 Mart 2025 saat: 09:30

Tüm sorular için tsne geliştirilerek

cos distance!