A close-up of a sign

AI-generated content may be incorrect.

**ΕΠΕΞΕΡΓΑΣΙΑ ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ**

**ΤΕΛΙΚΗ ΕΡΓΑΣΙΑ**

**ΟΜΑΔΑ**

**ΓΙΩΡΓΟΣ ΠΑΝΗΣ Π22217**

**ΛΟΡΕΝΖΟ ΔΗΜΗΤΡΙΟΥ Π22209**

**ΕΥΡΟΣ ΝΙΚΟΛΑΟΥ Π22213**

**📄 Αναλυτική Αναφορά Εργασίας NLP: Ανακατασκευή Προτάσεων & Σημασιολογική Ανάλυση**

**Πίνακας Περιεχομένων**

1. [**Εισαγωγή**](https://chatgpt.com/c/6856c6e0-ca88-8012-83b7-026d19206d35#1-%CE%95%CE%B9%CF%83%CE%B1%CE%B3%CF%89%CE%B3%CE%AE)
2. [**Μεθοδολογία**](https://chatgpt.com/c/6856c6e0-ca88-8012-83b7-026d19206d35#2-%CE%9C%CE%B5%CE%B8%CE%BF%CE%B4%CE%BF%CE%BB%CE%BF%CE%B3%CE%AF%CE%B1)
   * **2.1** [**Ανακατασκευή Προτάσεων**](https://chatgpt.com/c/6856c6e0-ca88-8012-83b7-026d19206d35#21-%CE%91%CE%BD%CE%B1%CE%BA%CE%B1%CF%84%CE%B1%CF%83%CE%BA%CE%B5%CF%85%CE%AE-%CE%A0%CF%81%CE%BF%CF%84%CE%AC%CF%83%CE%B5%CF%89%CE%BD)
   * **2.2** [**Ροές NLP**](https://chatgpt.com/c/6856c6e0-ca88-8012-83b7-026d19206d35#22-%CE%A1%CE%BF%CE%AD%CF%82-NLP)
   * **2.3** [**Σημασιολογική Ανάλυση**](https://chatgpt.com/c/6856c6e0-ca88-8012-83b7-026d19206d35#23-%CE%A3%CE%B7%CE%BC%CE%B1%CF%83%CE%B9%CE%BF%CE%BB%CE%BF%CE%B3%CE%B9%CE%BA%CE%AE-%CE%91%CE%BD%CE%AC%CE%BB%CF%85%CF%83%CE%B7)
3. [**Πειράματα & Αποτελέσματα**](https://chatgpt.com/c/6856c6e0-ca88-8012-83b7-026d19206d35#3-%CE%A0%CE%B5%CE%B9%CF%81%CE%AC%CE%BC%CE%B1%CF%84%CE%B1--%CE%91%CF%80%CE%BF%CF%84%CE%B5%CE%BB%CE%AD%CF%83%CE%BC%CE%B1%CF%84%CE%B1)
4. [**Αποτέλεσμα**](https://chatgpt.com/c/6856c6e0-ca88-8012-83b7-026d19206d35#4-%CE%A3%CF%85%CE%B6%CE%AE%CF%84%CE%B7%CF%83%CE%B7)
5. [**Συμπεράσματα**](https://chatgpt.com/c/6856c6e0-ca88-8012-83b7-026d19206d35#5-%CE%A3%CF%85%CE%BC%CF%80%CE%B5%CF%81%CE%AC%CF%83%CE%BC%CE%B1%CF%84%CE%B1)

**1 Εισαγωγή**

Αντικείμενη η στρατηγική επεξεργασία φυσικής γλώσσας μέσω επαναδιατυπό NLP pipelines και εργαλεία cosine similarity.

**2 Μεθοδολογία**

**2.1 Ανακατασκευή Προτάσεων**

Χειρόποινη ανακατασκευή για εναρθρά γνωστό τονο περιεχόμενο κείμενο.

**2.2 Ροές NLP**

* custom\_reconstruct: static mapping
* T5: pretrained T5 ("Vamsi/T5\_Paraphrase\_Paws")
* Gramformer: grammatical correction via transformer
* LanguageTool: κανονικοποιημενος ελέγκτηκας

**2.3 Σημασιολογική Ανάλυση**

* Χρήση cosine similarity ανάμεσα σε Sentence-BERT (MiniLM)
* Προβολή PCA για 2D οπτικοίζεις

**2.4 Δομή Αρχείων**

📦project**/**

**│**

**├── main.py # Κύριο script εκτέλεσης**

**├── reconstruction/**

**│ ├── custom.py # Χειροκίνητη ανακατασκευή**

**│ ├── t5\_pipeline.py # T5 paraphrasing με Transformers**

**│ └── grammar\_tools.py # Γραμματική διόρθωση με Gramformer & LanguageTool**

**├── README.md # Τεκμηρίωση έργου**

**├── requirements.txt # Λίστα εξαρτήσεων**

**└── .gitignore**

**3 Πειράματα & Αποτελέσματα**

A black background with white and blue lines

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A graph with lines in it

AI-generated content may be incorrect.

**3.1 Ανακατασκευές Εξόδου**

Original: "Thank your message to show our words to the doctor, as his next contract checking..."

Custom: Χρήση συγκεκριμένη βάση mapping

T5: Αποκωδικό απόσπασμα paraphrasing

Gramformer / LanguageTool: Γραμματική διόρθωση

**3.2 Cosine Similarity (1o Κείμενο)**

| Μέθοδος | Similarity |
| --- | --- |
| Custom | 1.0 |
| T5 | 0.891 |
| Gramformer | 0.67 |
| LanguageTool | 0.99 |

**3.3 PCA**

Οπτική σε 2D, δείχνεται ομοιότητες ανάμεσα στον σημασιολογικό χώρο.

**4 Αποτελέσμα**

* Η T5 ανταποκρίνεται ικανοποιητικά
* Η Custom είναι επιλεκτικά για fixed περιπτώσεις
* Ο Συνδυασμός cosine + PCA ειναι σχετικό εργαλείο

**5 Συμπεράσματα**

* Η NLP ανακατασκευή βοήθησε το χάος
* Οι πιοστικές ροές T5/Γραμματικής είναι χρήσιμες για ανακατασκευές ποιοτικά
* Τα embeddings BERT είναι αξιόπιστα για ανάλυση νοήματος