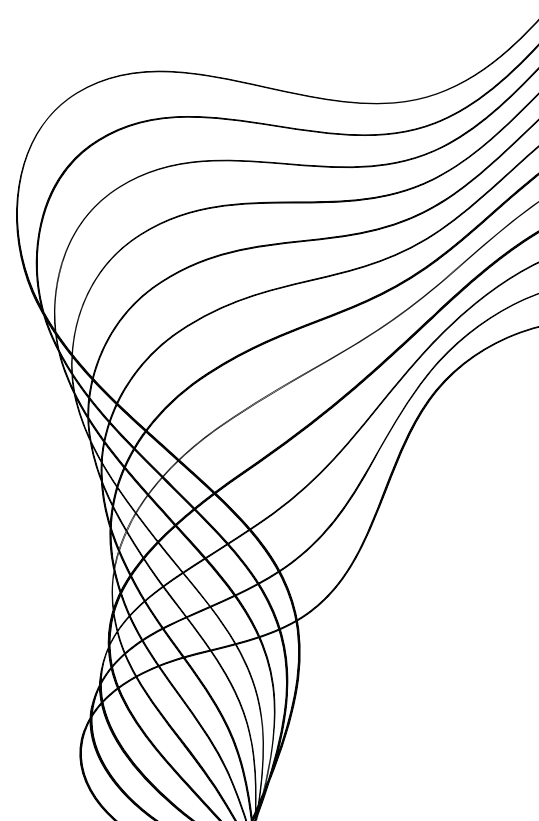


# Առանցքային բառերի հայտնաբերման համակարգի մշակումը

Առարկա՝ Բնական լեզվի մշակում  
Խումբ՝ **SS161**  
Ուսանող՝ Արևիկ Խաչատրյան



# Բնվանդակություն

01

BERT մոդել

02

Custom dataset-ների կառուցում

03

RelationshipClassifier-ի մասին

04

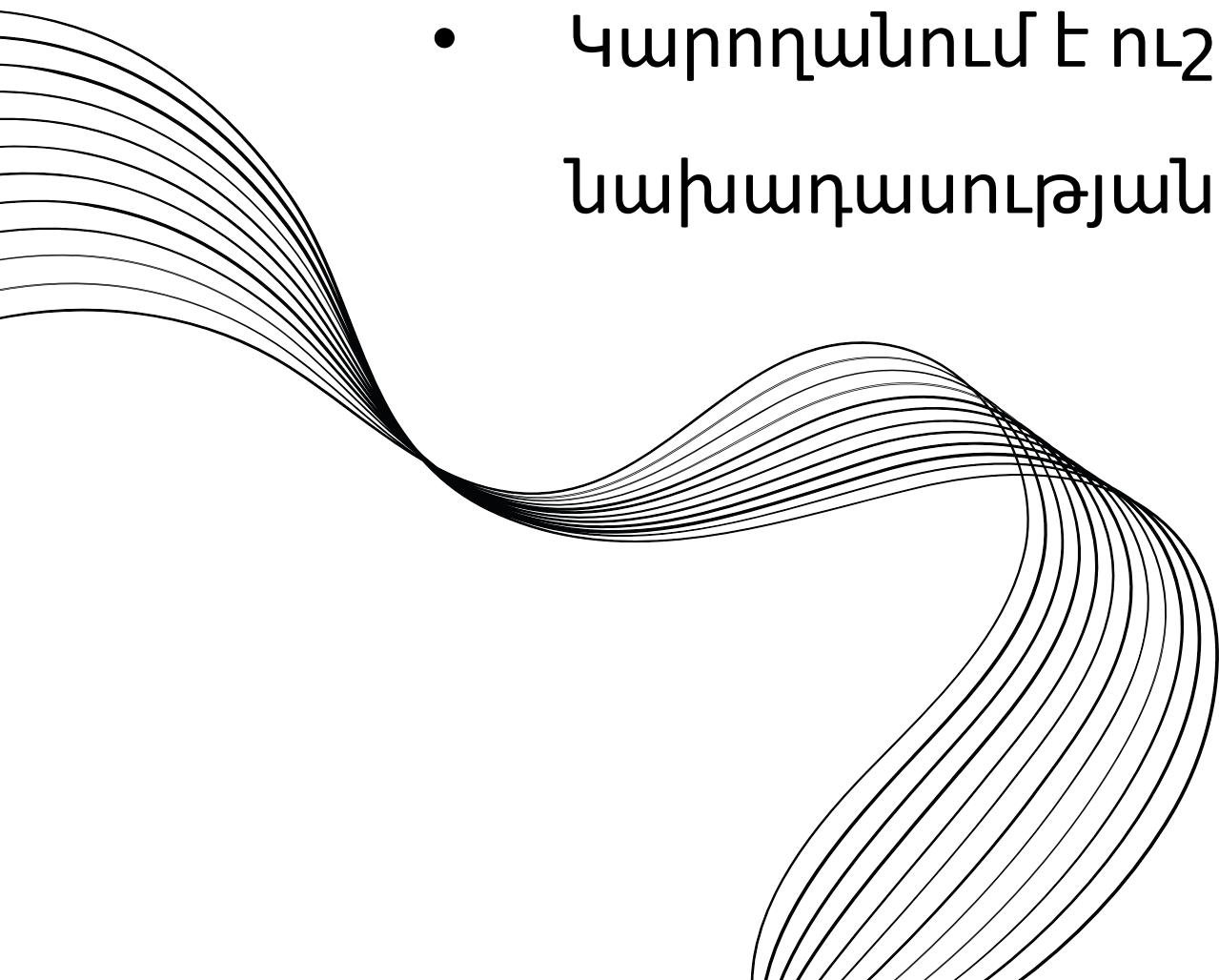
Արդյունքների մեկնաբանում

05

Եզրակացություն

# BERT մոդել

- Տարբերվում է իր երկկողմանի ուսուցմամբ, որը հնարավորություն է տալիս նախադասությունները կարդալ երկու ուղղություններով:
- Կարողանում է ուշադրություն դարձնել միաժամանակ նախադասության նախորդ և հաջորդ բառերին:



# Custom dataset-ների կառուցում

- Ավտոմատացնում է թոքենիզացիան և տեքստային տվյալների փոխարկումը BERT-ի պահանջներին համապատասխան ֆորմատի:
- Հնարավորություն է տալիս իրականացնել բարդ նախնական մշակումներ, օրինակ՝ բազմակի մուտքերի կամ առաջադրանքների վերամշակում:
- Համատեղելի է PyTorch-ի DataLoader-ի հետ՝ ապահովելով խմբաքանակային բեռնավորումն ու խառնուրդը:

# RelationshipClassifier-ի մասին

RelationshipClassifier-ը դասակարգման մոդուլ է, որը հատուկ նախագծված է հարաբերությունների հայտնաբերման և դասակարգման խնդիրների համար:

- Օգտագործում է BERT-ի pool-ային ելքը ([CLS] թոքենից) որպես հիմնական մուտքային հատկանիշ:
- Կառուցված է բազմաշերտ նեյրոնային ցանցի (MLP) վրա՝ ավելի բարդ պատկերացումներ սովորելու համար:

Մոդելի աշխատանք

# Արդյունքների մեկնաբանում

Օրինակի  
դիտարկում

Epoch 1/5, Loss: 0.9081193804740906  
Epoch 2/5, Loss: 0.8427415490150452  
Epoch 3/5, Loss: 0.2597549259662628  
Epoch 4/5, Loss: 0.30668625235557556  
Epoch 5/5, Loss: 0.09578899294137955

	precision	recall	f1-score	support
Cause-Effect(e1,e2)	0.98	0.89	0.93	61
Cause-Effect(e2,e1)	0.92	0.94	0.93	138
Component-Whole(e1,e2)	0.94	0.80	0.86	100
Component-Whole(e2,e1)	0.79	0.88	0.83	97
Content-Container(e1,e2)	0.91	0.85	0.88	88
Content-Container(e2,e1)	0.87	0.81	0.84	42
Entity-Destination(e1,e2)	0.94	0.94	0.94	158
Entity-Destination(e2,e1)	0.00	0.00	0.00	0
Entity-Origin(e1,e2)	0.82	0.81	0.82	107
Entity-Origin(e2,e1)	0.68	0.86	0.76	29
Instrument-Agency(e1,e2)	0.67	0.86	0.75	14
Instrument-Agency(e2,e1)	0.89	0.86	0.87	85
Member-Collection(e1,e2)	0.79	0.88	0.83	17
Member-Collection(e2,e1)	0.85	0.92	0.88	119
Message-Topic(e1,e2)	0.89	0.89	0.89	103
Message-Topic(e2,e1)	0.86	0.83	0.84	29
Other	0.66	0.67	0.66	260
Product-Producer(e1,e2)	0.86	0.89	0.87	66
Product-Producer(e2,e1)	0.94	0.85	0.89	87
accuracy			0.84	1600
macro avg	0.80	0.81	0.81	1600
weighted avg	0.85	0.84	0.84	1600

Example sentence:

Elon Musk is the CEO of Tesla and SpaceX, and he recently visited Berlin.

All entities:

('Elon Musk', 'PERSON')  
( 'Tesla', 'ORG')  
( 'Berlin', 'GPE')

Extracted Entities:

e1: Elon Musk  
e2: Tesla

Normalized sentence:

[E1] is the CEO of [E2] and SpaceX, and he recently visited Berlin.

Predicted Relationships:

Other: 0.77

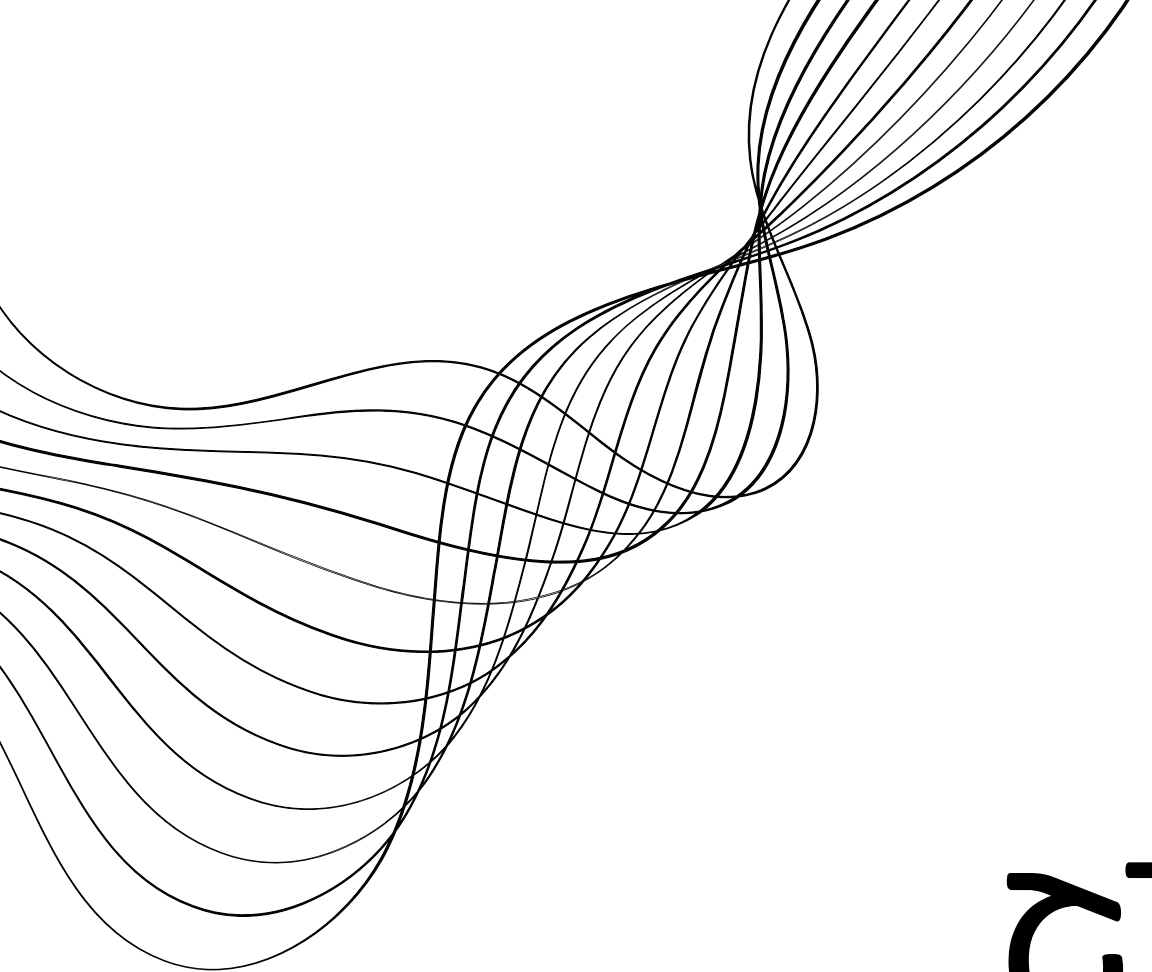
Instrument-Agency(e2,e1): 0.08

Product-Producer(e2,e1): 0.05

# Եզրակացություն

Առանցքային բառերի հայտնաբերման առաջարկվող համակարգը ցուցադրել է հաջող արդյունքներ՝ ապահովելով տեքստերի ավտոմատ վերլուծության բարձր արդյունավետություն, ճշգրտություն և հուսալիություն:

Հավելյալ հարմարեցման և ֆունկցիոնալության բարելավման շնորհիվ այն կարող է մշակել մեծածավալ տվյալներ և ապահովել ավելի լայն կիրառում՝ դարձնելով այն բազմակողմանի գործիք տեքստային տվյալների մշակման համար:



Հնորհակալություն

