

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
ім. Ігоря Сікорського»  
ТЕПЛОЕНЕРГЕТИЧНИЙ ФАКУЛЬТЕТ

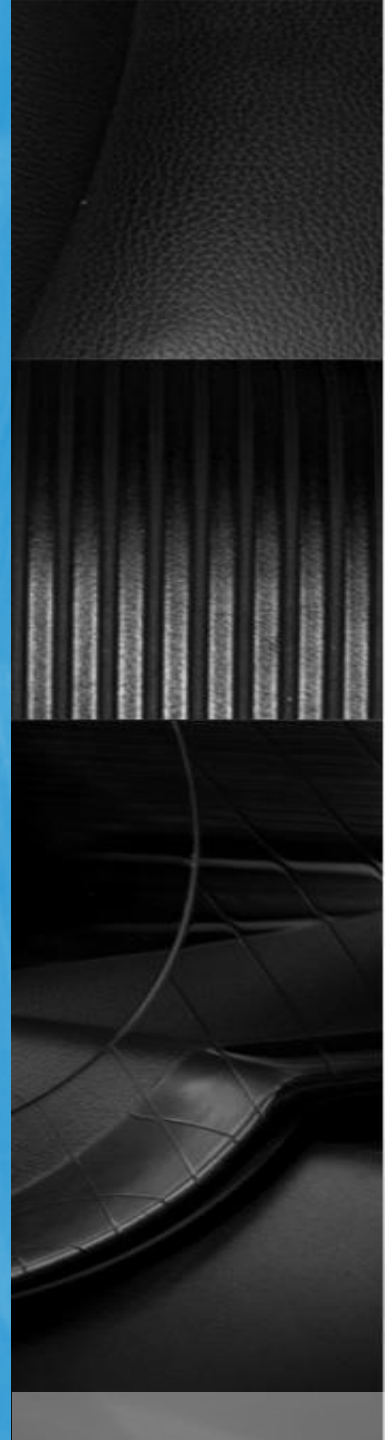
## Автоматичний морфологічний аналіз слів флексивної мови

Магістрант групи ТМ-61м

Гуровський Володимир

Науковий керівник

Доцент к.т.н. Стативка Ю.І.



## Мета дослідження

полягає в розробці програмного забезпечення для автоматизованого морфологічного аналізатора флективної мови для відкритого вільного доступу.

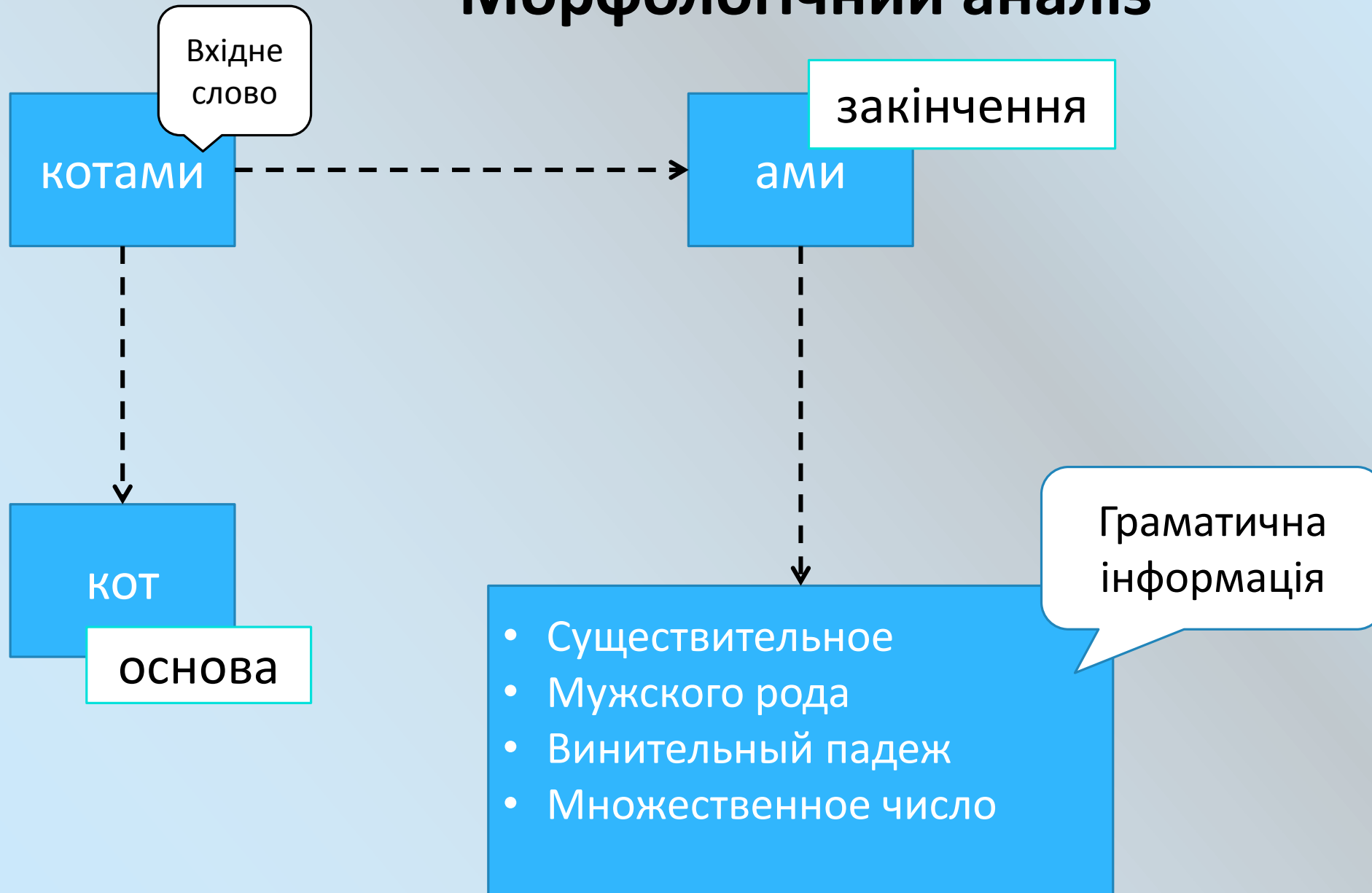
## Об'єкт дослідження

комп'ютерні інформаційні технології автоматичної обробки флективних мов.

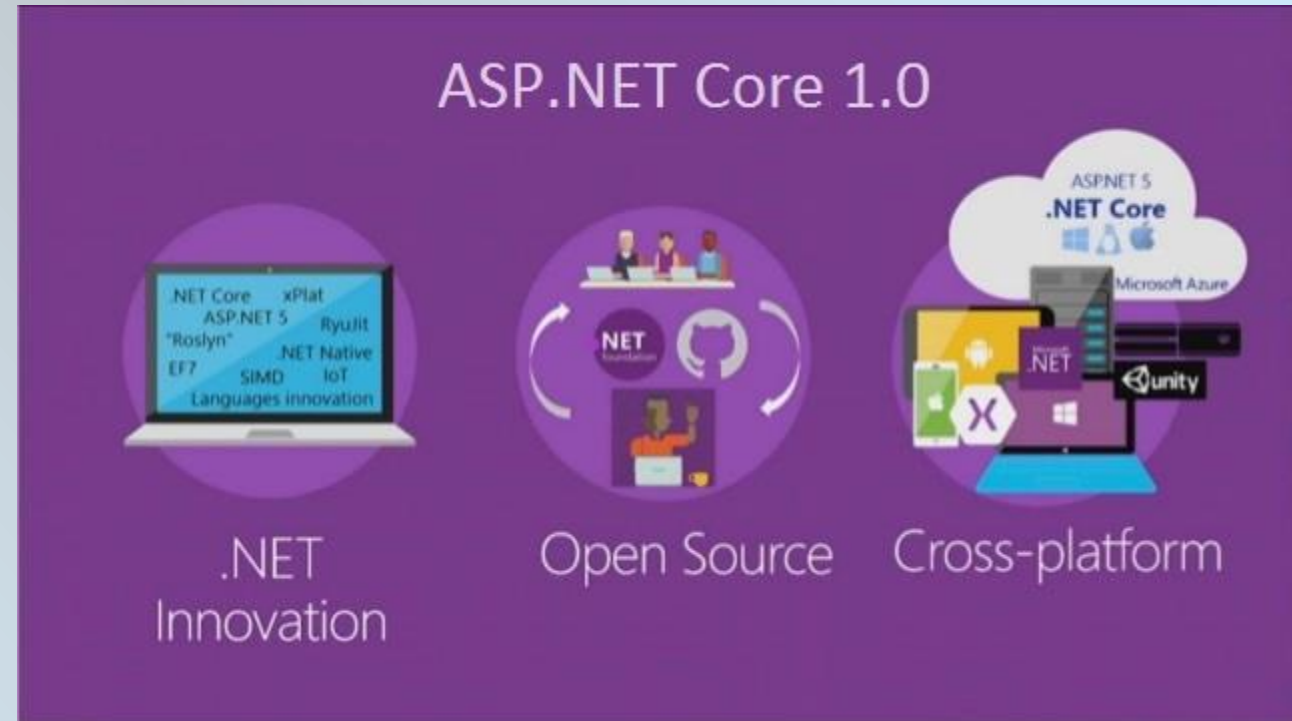
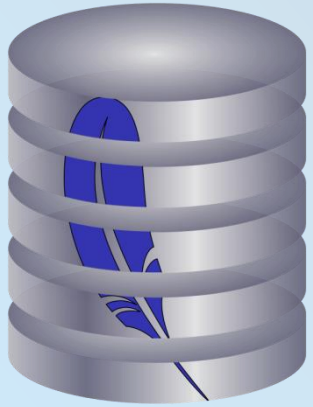
## Предмет дослідження

комп'ютерні інформаційні технології автоматичного морфологічного аналізу слів флективної мови.

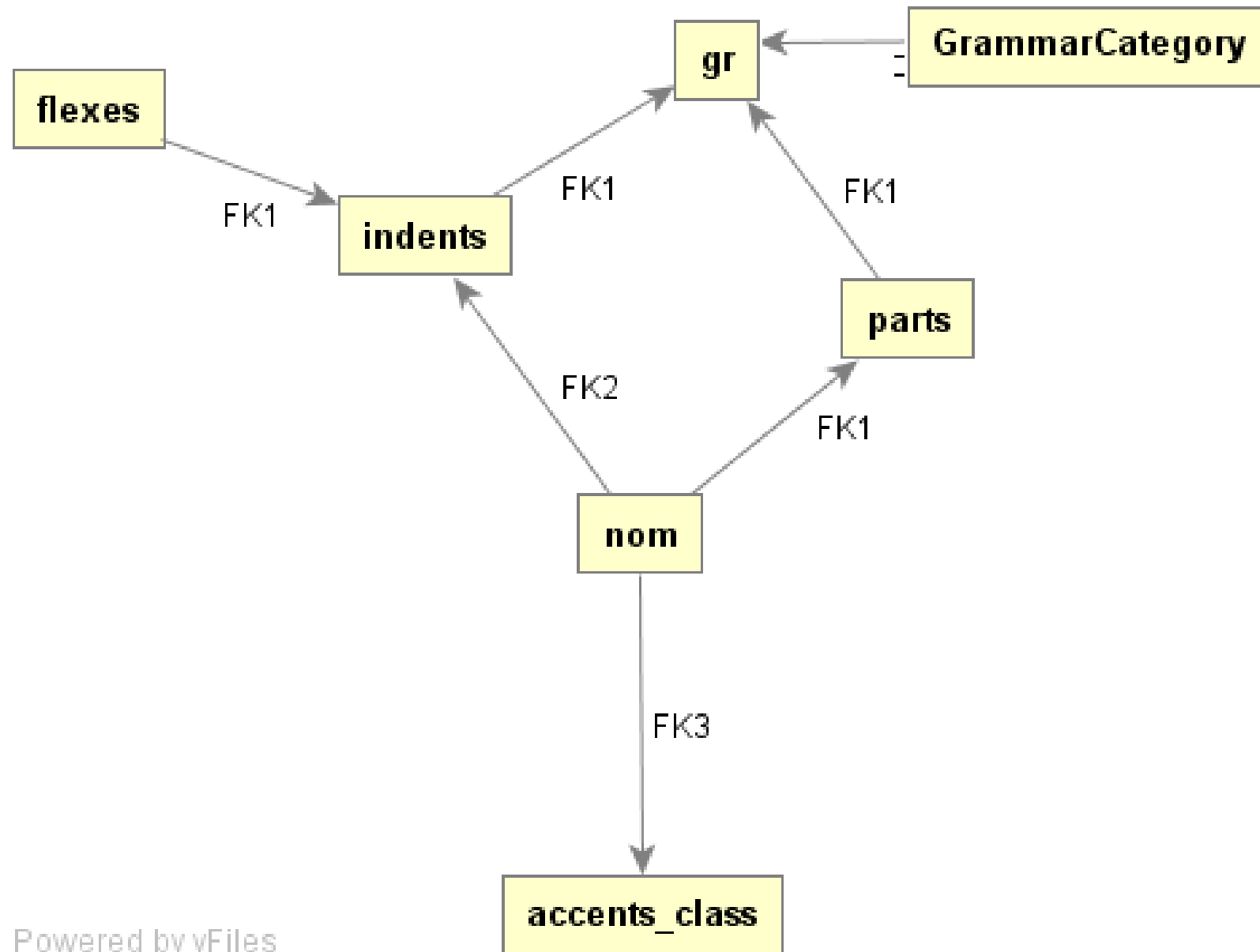
# Морфологічний аналіз



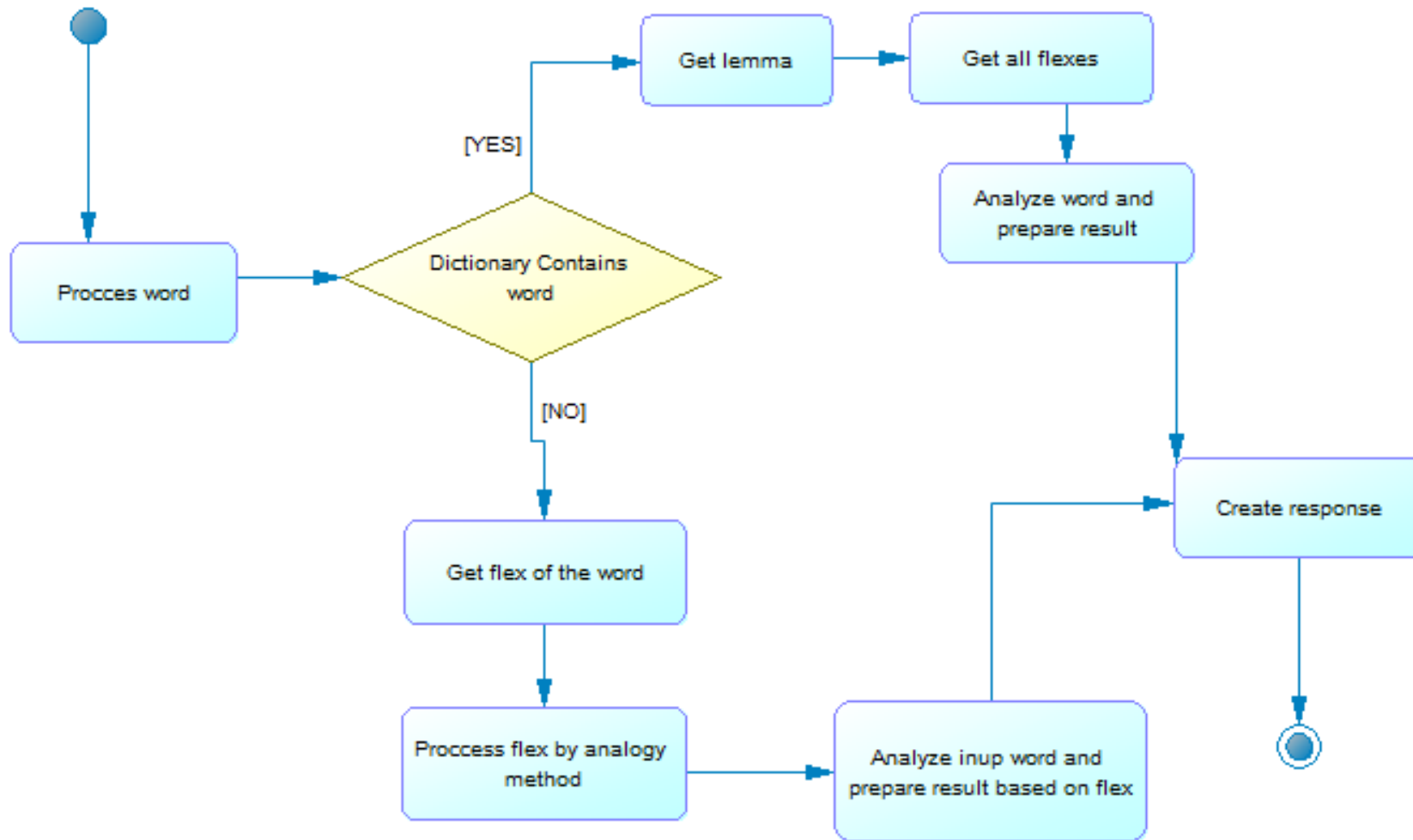
# Середовище та програмні засоби



# Схема Бази Даних



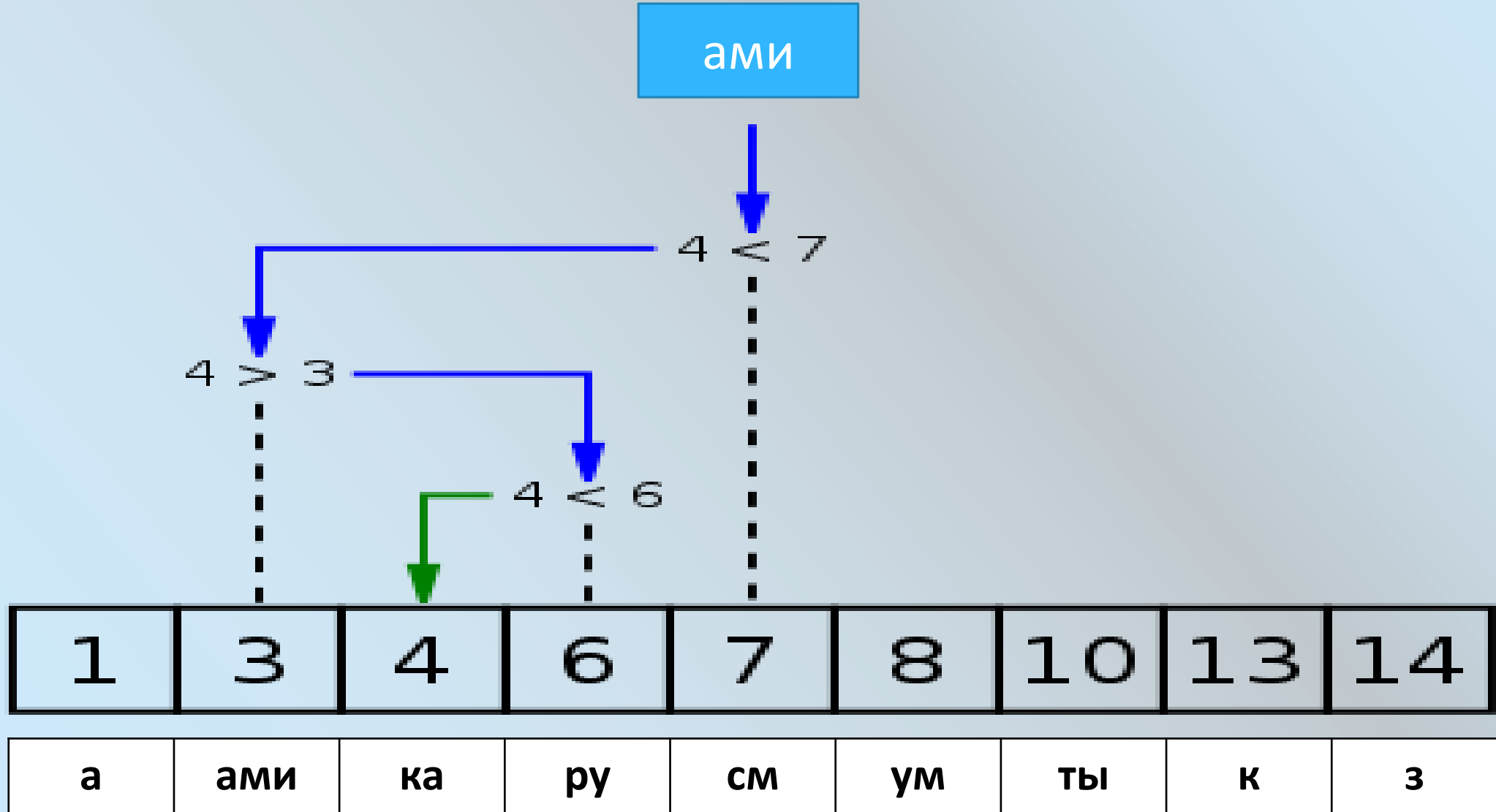
# Алгоритм пошуку слова



# Структура та принцип роботи Хеш-Таблиці



# Бінарний пошук флексів за індексом в методі аналогій





# Аналоги

**NHunspell**  
Spell Check  
Hyphenation  
Thesaurus

**py morphology 2**

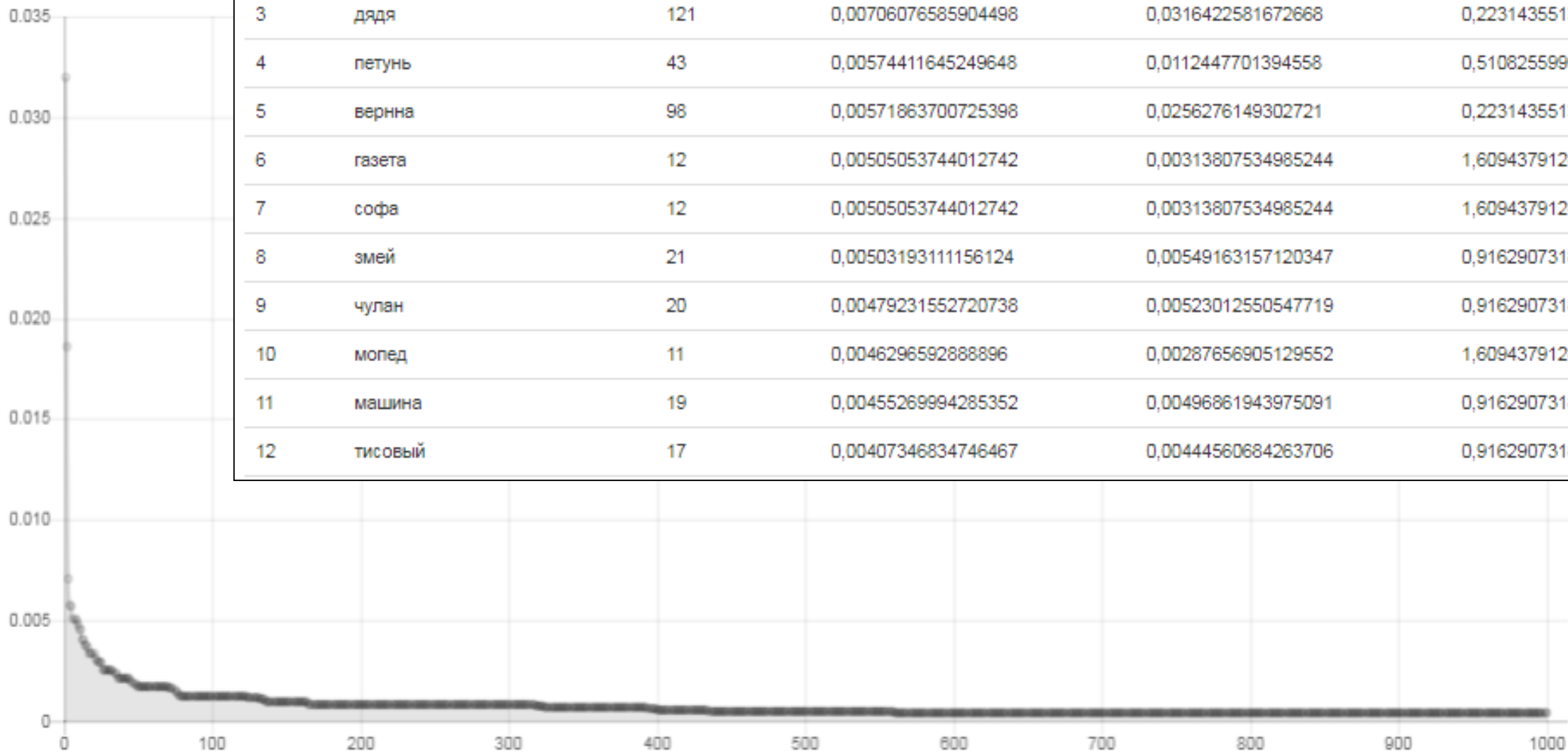


Морфологический  
анализатор Mystem 3.0

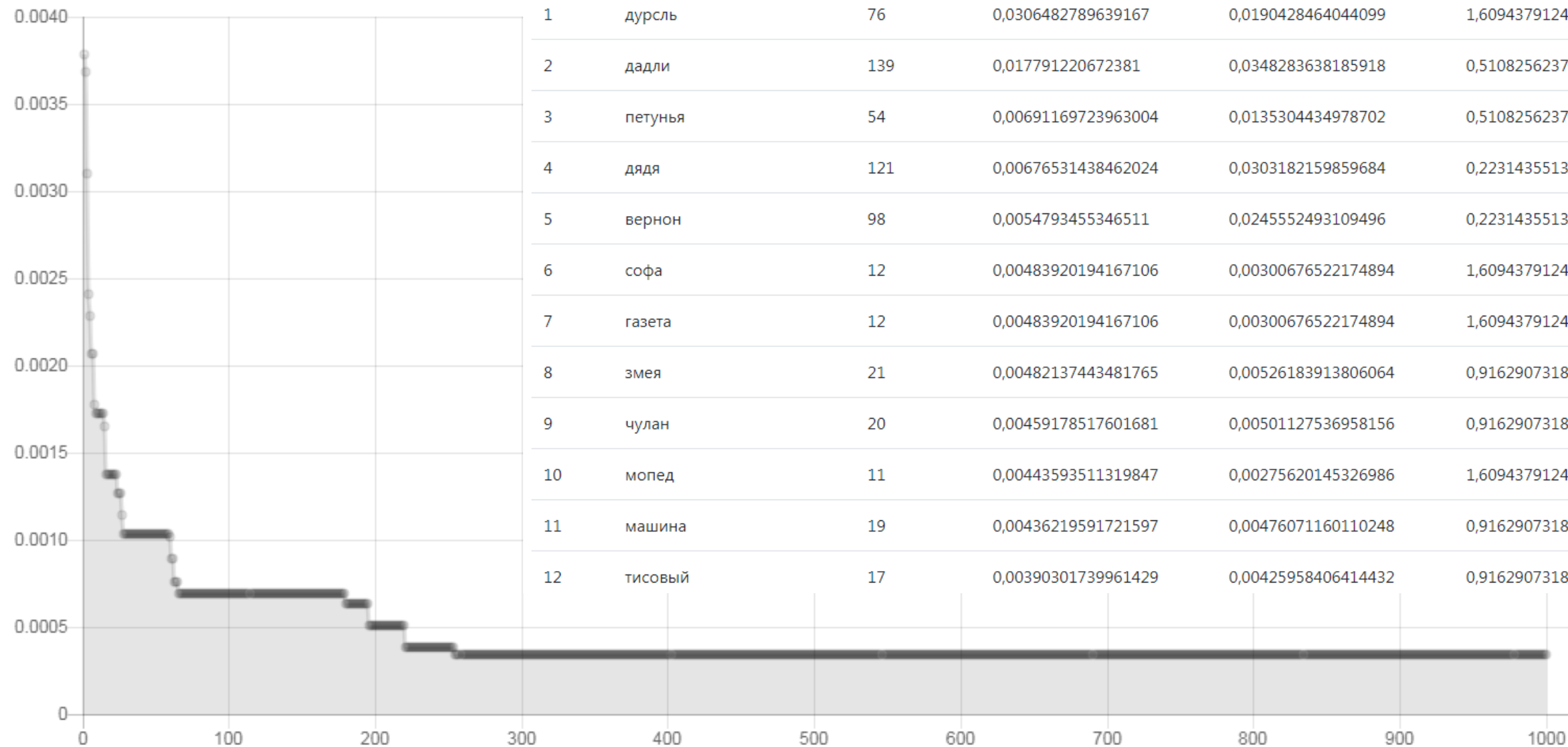
Алексей Зобнин

# Результат дослідження якості на основі методу TF-IDF

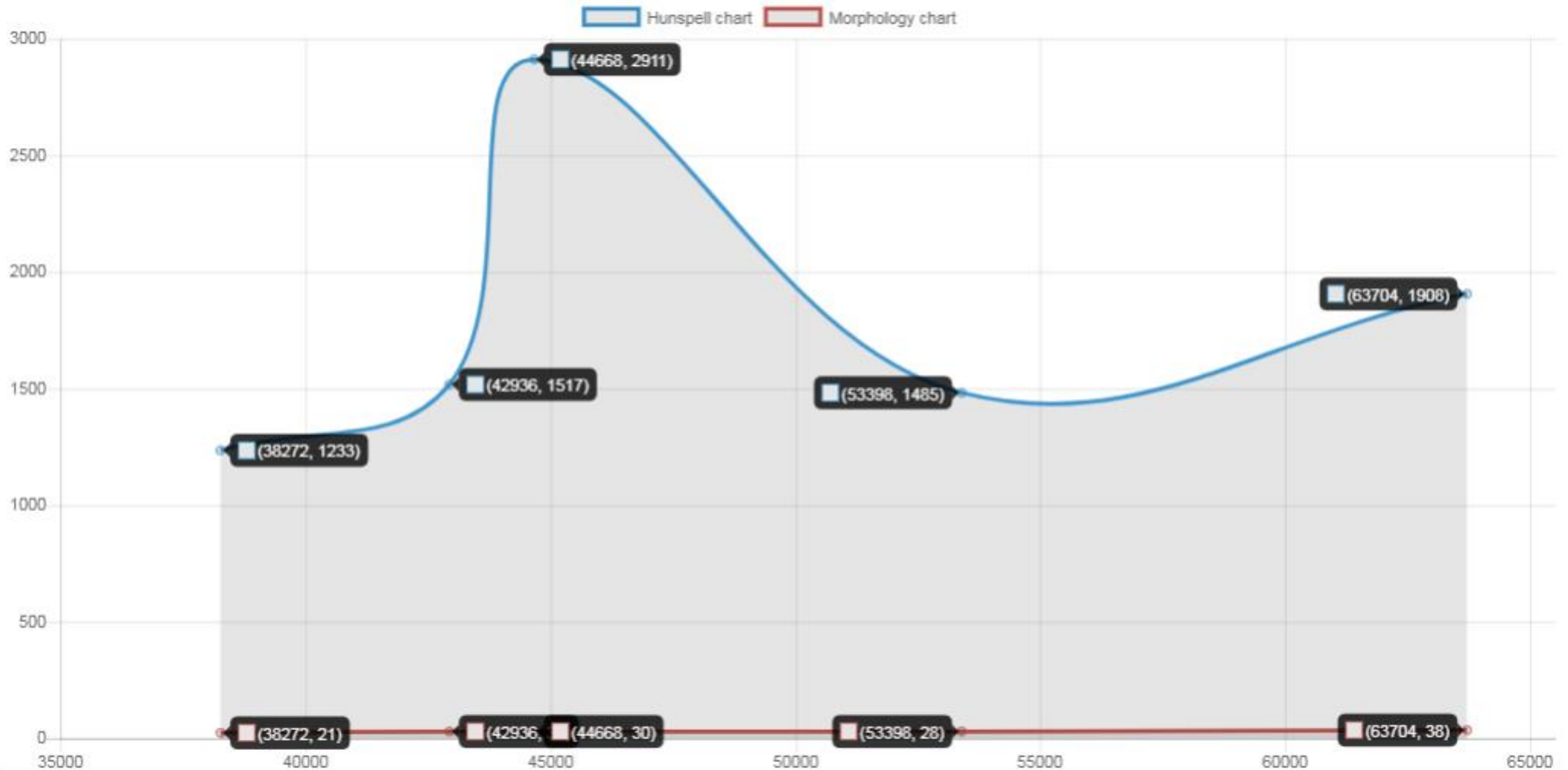
Index	Word	Frequency	TfidfValue	TfValue	IdfValue
1	дурсль	76	0,0319867379951688	0,0198744777590036	1,6094379124341
2	дадти	139	0,0185681892865737	0,0363493710756302	0,510825599924132
3	дядя	121	0,00706076585904498	0,0316422581672668	0,22314355131421
4	петуль	43	0,00574411645249648	0,0112447701394558	0,510825599924132
5	вернна	98	0,00571863700725398	0,0256276149302721	0,22314355131421
6	газета	12	0,00505053744012742	0,00313807534985244	1,6094379124341
7	софа	12	0,00505053744012742	0,00313807534985244	1,6094379124341
8	змеї	21	0,00503193111156124	0,00549163157120347	0,916290731874155
9	чулан	20	0,00479231552720738	0,00523012550547719	0,916290731874155
10	мопед	11	0,0046296592888896	0,00287656905129552	1,6094379124341
11	машина	19	0,00455269994285352	0,00496861943975091	0,916290731874155
12	тисовый	17	0,00407346834746467	0,00444560684263706	0,916290731874155



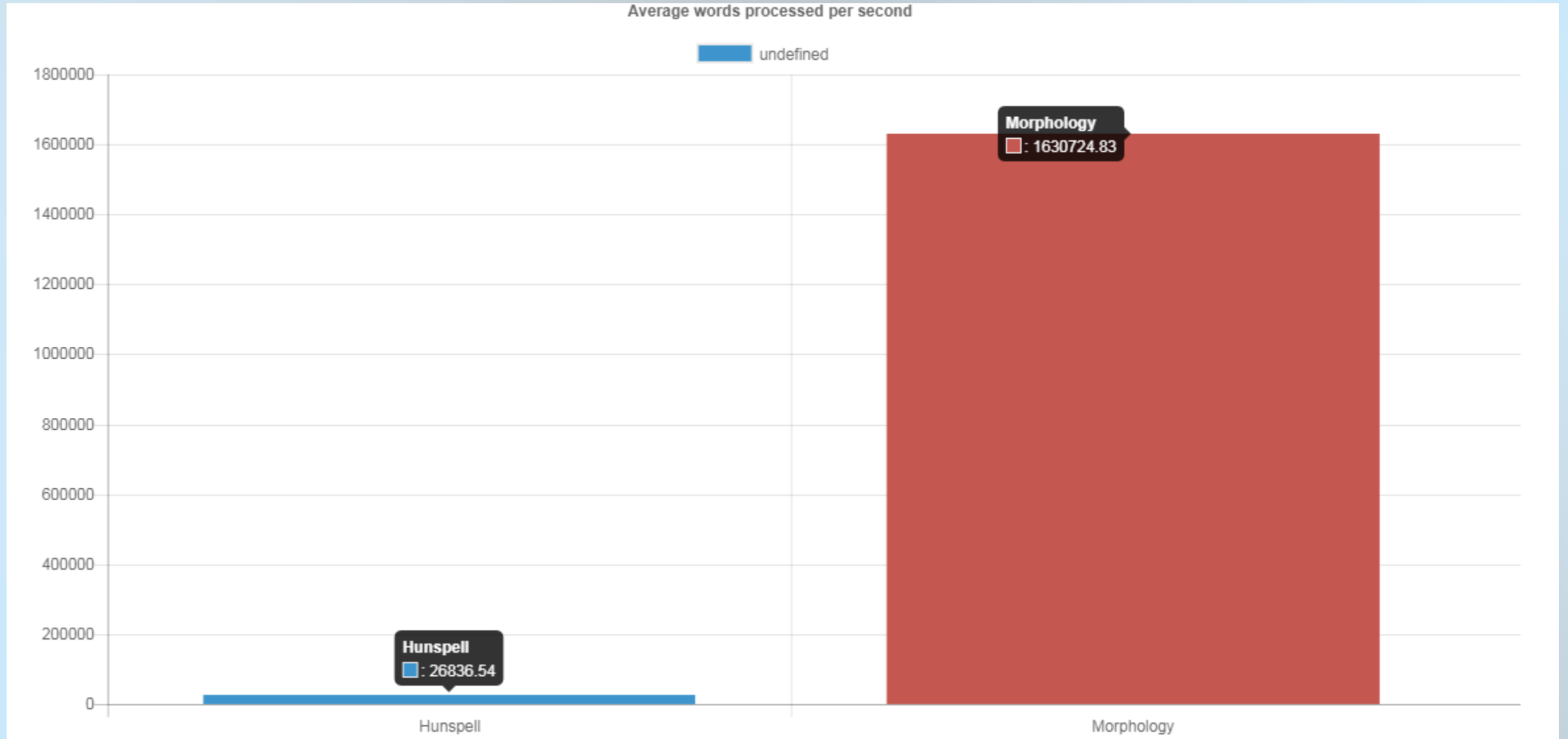
# Результат дослідження якості на основі методу TF-IDF для Hunspell



# Стабільність в порівнянні з Hunspell.



# Продуктивність системи в порівнянні з Hunspell



# Результаты работы зі словами в словнику

```
localhost:4130/api/word/головой
{
  "data": {
    "inputWord": "головой",
    "words": [
      {
        "wordBase": "голова",
        "lemma": "голов",
        "identifier": "существительное мужского рода",
        "wordForms": [
          {
            "word": "голова",
            "field2": 1,
            "grId": 0,
            "grammarCategoryDesc": "Именительный падеж, Единственное число, "
          },
          {
            "word": "головы",
            "field2": 2,
            "grId": 0,
            "grammarCategoryDesc": "Родительный падеж, Единственное число, "
          },
          {
            "word": "голове",
            "field2": 3,
            "grId": 0,
            "grammarCategoryDesc": "Дательный падеж, Единственное число, "
          },
          {
            "word": "голову",
            "field2": 4,
            "grId": 0,
            "grammarCategoryDesc": "Винительный падеж, Единственное число, "
          },
          {
            "word": "головой",
            "field2": 5,
            "grId": 0,
```

```
localhost:4130/api/word/infinitive/головой
{
  "data": "голова",
  "time": 0
}
```

```
localhost:4130/api/word/ходила
{
  "data": {
    "inputWord": "ходила",
    "words": [
      {
        "wordBase": "ходить",
        "lemma": "хо",
        "identifier": "глагол несовершенного вида",
        "wordForms": [ ... ] // 18 items
      }
    ],
    "analyzeResult": "NotFound",
    "foundByFlexLength": 0
  },
  "time": 0
}
```

# Результати роботи з новими словами(неологізмами)

```
localhost:4130/api/word/деладрить
```

```
{  
  "data": {  
    "inputWord": "деладрить",  
    "words": [  
      {  
        "wordBase": "благоустроить",  
        "lemma": "деладр",  
        "identifier": "глагол совершенного ви",  
        "wordForms": [  
          {  
            "word": "деладрить",  
            "field2": 1,  
            "grId": 0,  
            "grammarCategoryDesc": "Инфинитив, "  
          },  
          {  
            "word": "деладрю",  
            "field2": 2,  
            "grId": 0,  
            "grammarCategoryDesc": "Повелительное наклонение, 2 лицо глагола, Единственное число, "  
          },  
          {  
            "word": "деладришь",  
            "field2": 3,  
            "grId": 0,  
            "grammarCategoryDesc": "Повелительное наклонение, 1 лицо глагола, множественное число, "  
          },  
          {  
            "word": "деладрит",  
            "field2": 4,  
            "grId": 0,  
            "grammarCategoryDesc": "Повелительное наклонение, 2 лицо глагола, множественное число, "  
          },  
          {  
            "word": "деладрим",  
            "field2": 5,  
            "grId": 0,  
            "grammarCategoryDesc": "будущее время, 1 лицо глагола, Единственное число, "  
          }  
        ]  
      }  
    ]  
  }  
}
```

```
localhost:4130/api/word/infinitive/деладрила
```

```
{  
  "data": "деладрить",  
  "time": 27  
}
```

```
localhost:4130/api/word/infinitive/деладрили
```

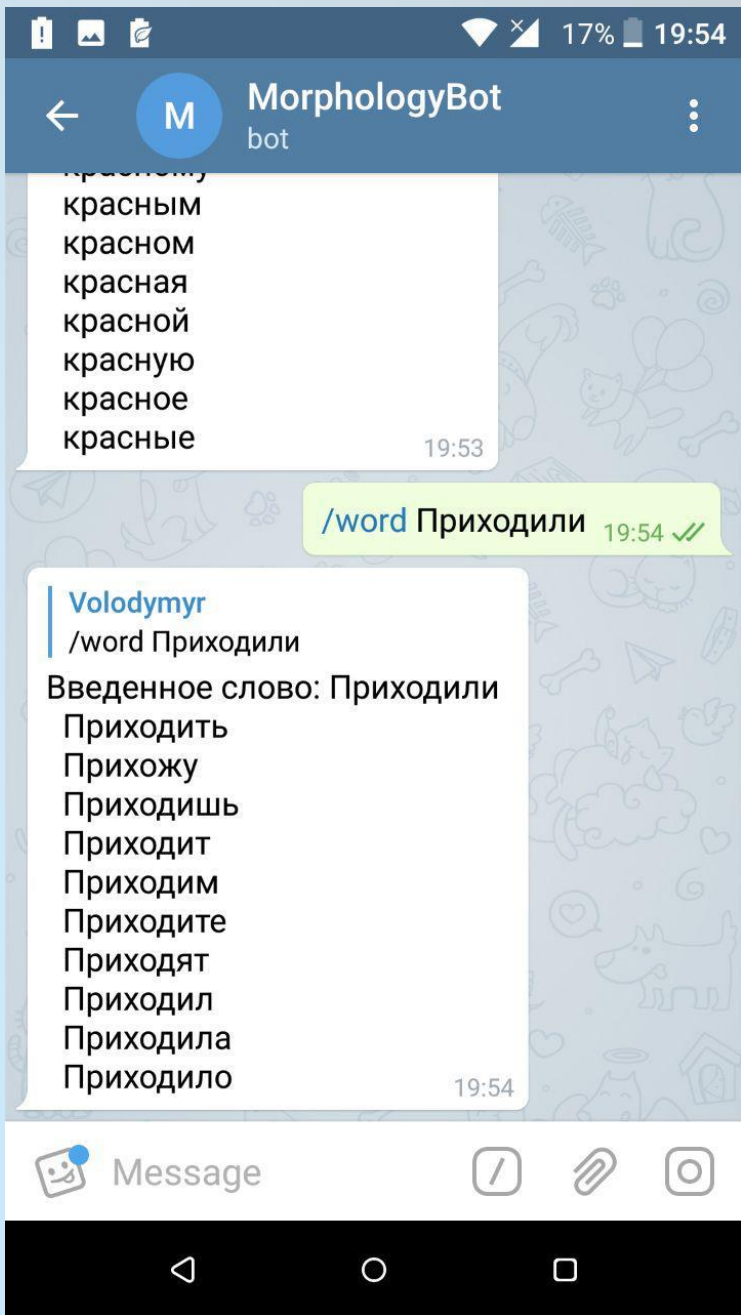
```
{  
  "data": "деладрить",  
  "time": 11  
}
```

```
localhost:4130/api/word/infinitive/притикрасные
```

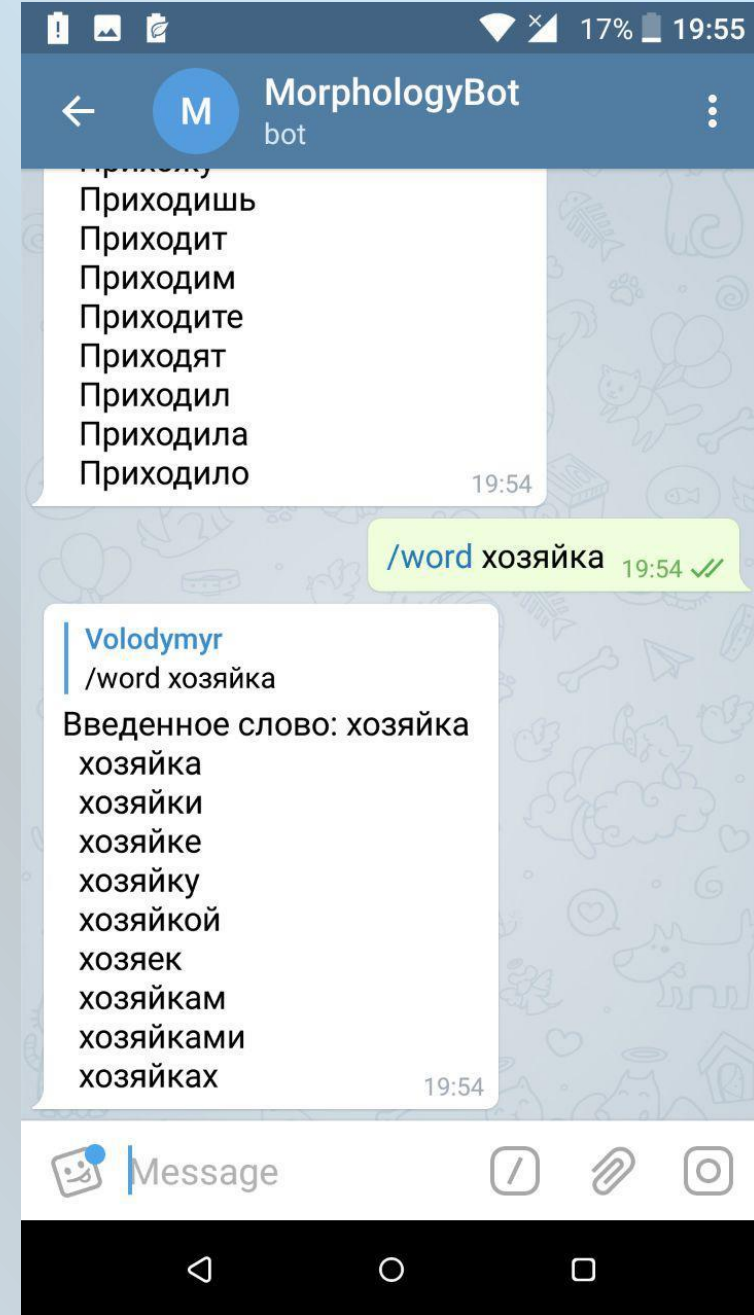
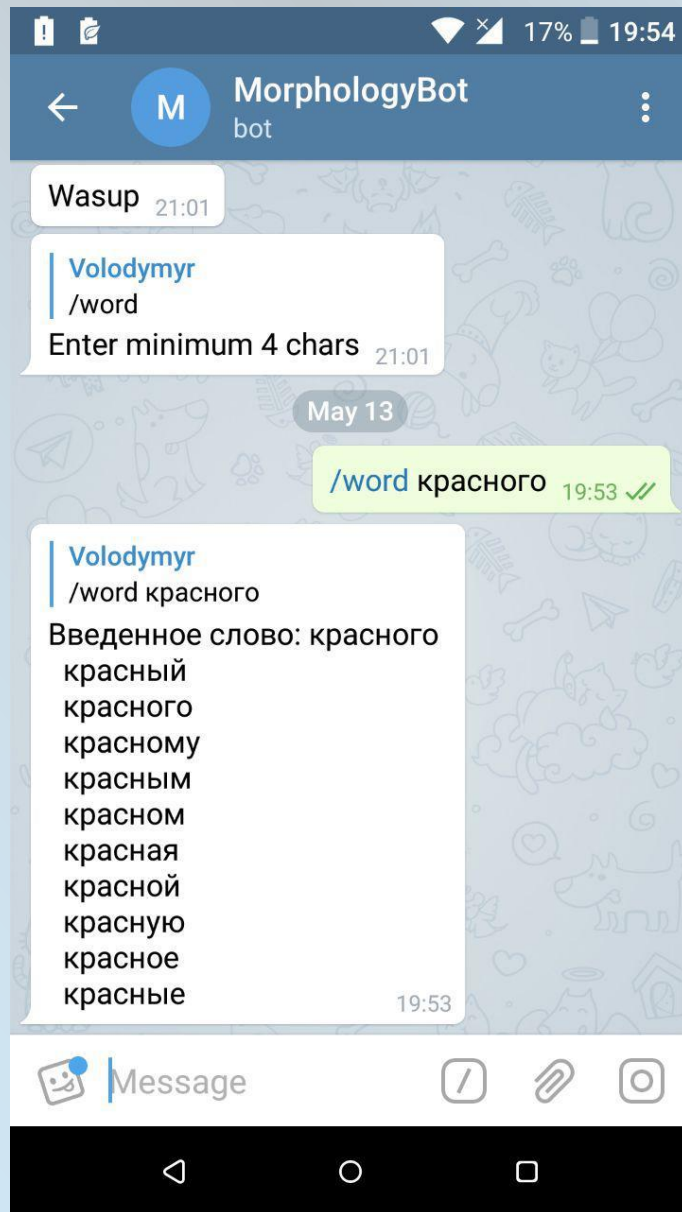
```
{  
  "data": "притикрасной",  
  "time": 19  
}
```

```
localhost:4130/api/word/infinitive/притикрасные
```

```
{  
  "data": "притикрасной",  
  "time": 14  
}
```



# Telegram бот







# ВИСНОВКИ



Розглянуто проблему морфологічного аналізу флективних мов та проаналізовано існуючі методики, їх реалізацію у комп'ютерних системах, виявлено їх головні недоліки.



Удосконалено метод морфологічного аналізу за допомогою використання словника основ слів, за рахунок організації даних до хеш-таблиці, та використанням бінарного пошуку для методу аналогій.



Розроблено універсальний морфологічний аналізатор флективних мов та покращено роботу алгоритму аналогій, який лежить в основі роботи з новими, рідкісними словами або okazionalizмами.