

UK-DICTIONARY v0.1

ДОКУМЕНТАЦІЯ • 2025.02.12

ЗМІСТ

```
{  
  "_id": "",  
  "lemma": "",  
  "pos": "",  
  "paradigm_id": "",  
  "endings": ["", ...],  
  "forms": [{...}, ...],  
  "resources": ["", ...]  
}
```

ВИЯВЛЕНІ ПРОБЛЕМИ

- **Дублювання вхідних даних.** Наприклад: "афіш" у _ та "афіш" в «іменники 2 відміна тверда група.json».
- **Невідповідне відмінювання у вхідних даних**
 - іменники 2 відміна тверда група.json:
 - "м" стоїть для давального відмінка множини; можливо має бути "ам"; приклад помилково згенерованої форми: "єнотм".
- **Інші проблеми**
 - для леми "єдиноріг" утворюються помилкові форми "єдиноріга", "єдинорігові" та інші; потребується виправлення вхідних даних або поліпшення процесу створення форм.

_id

Тип: string

Опис: унікальний ідентифікатор запису

Примітки: як правило, _id = lemma, але можна змінити на інший необхідний формат; використовується для швидкого доступу в базі даних

lemma

Тип: string

Опис: базова словникова форма слова

Примітки: може містити наголос (напр. "бу́хка"); є ключем для групування словоформ

pos

Тип: string

Опис: частина мови

Значення	Опис
NOUN	іменник
VERB	дієслово
ADJ	прикметник
PARTICIPLE	дієприкметник

paradigm_id

Тип: string

Опис: назва морфологічної парадигми, яка визначає набір можливих словоформ для леми

ID парадигми	Частина мови	Опис морфологічного класу
uk-verb-1st-perf-refl	VERB	Дієслова 1-ї дієвідміни, доконаний вид, рефлексивні (із -ся / -сь).
uk-verb-1st-perf-nonrefl	VERB	Дієслова 1-ї дієвідміни, доконаний вид, нерефлексивні.
uk-verb-1st-imperf-nonrefl	VERB	Дієслова 1-ї дієвідміни, недоконаний вид, нерефлексивні.
uk-verb-2nd-perf-refl	VERB	Дієслова 2-ї дієвідміни, доконаний вид, рефлексивні.
uk-verb-2nd-perf-nonrefl	VERB	Дієслова 2-ї дієвідміни, доконаний вид, нерефлексивні.
uk-verb-2nd-imperf-nonrefl	VERB	Дієслова 2-ї дієвідміни, недоконаний вид, нерефлексивні.
uk-verb-2nd-imperf-refl	VERB	Дієслова 2-ї дієвідміни, недоконаний вид, рефлексивні.
uk-noun-1st-soft	NOUN	Іменники 1-ї відміни, м'яка група (основа на -я/-я'/-ья).
uk-noun-1st-hard	NOUN	Іменники 1-ї відміни, тверда група (основа на твердий приголосний).

uk-noun-1st-mixed	NOUN	Іменники 1-ї відміни, мішана група (змішана система чергувань).
uk-noun-2nd-soft	NOUN	Іменники 2-ї відміни, м'яка група (основа на м'який приголосний).
uk-noun-2nd-mixed	NOUN	Іменники 2-ї відміни, мішана група.
uk-noun-2nd-hard	NOUN	Іменники 2-ї відміни, тверда група.
uk-noun-3rd	NOUN	Іменники 3-ї відміни (завжди жіночий рід, основа на м'який приголосний, нульове
uk-noun-4th	NOUN	Іменники 4-ї відміни (іменники середнього роду на -ат/-ят).
uk-noun-plurale	NOUN	Іменники, що існують лише у множині.
uk-adj-soft	ADJ	Прикметники м'якої групи (закінчення -ій, -я, -є).
uk-adj-hard	ADJ	Прикметники твердої групи (закінчення -ий, -а, -е).
uk-participle-long	PARTICIPLE	Дієприкметники повної форми (прикметникова парадигма, 28 форм: рід x число x відмінок).

Примітки: значення парадигми визначає набір відмінків, часів, осіб, родів, чисел, які потрібно згенерувати

endings

Тип: array of strings

Опис: список закінчень, з яких будується кожна словоформа. Кожен елемент відповідає конкретному слоту з парадигми.

Примітки: основа (lemma) + закінчення = словоформа

forms

Тип: array of objects

Опис: список усіх згенерованих словоформ, кожна з яких відображає певні граматичні властивості.

Структура одного елемента:

```
{
  "form": "єнота",
  "slot": "GEN_SG",
  "features": {
    "pos": "NOUN",
    "mood": "gen",
    "number": "sg"
  }
}
```

Поле	Тип	Опис
form	string	готова словоформа
slot	string	технічне ім'я позиції у парадигмі (напр. PRES_3PL)
features	object	граматичні категорії цієї форми

Категорія	Приклади значень	Опис
tense	"past", "pres", "fut"	Час дієслова. Вказує, коли

		відбувається дія: у минулому (past), теперішньому (pres) чи майбутньому часах (fut)
mood	"ind", "imp", "inf"	Спосіб дієслова: дійсний (ind), наказовий (imp), інфінітив (inf). Визначає ставлення дії до реальності.
gender	"m", "f", "n"	Рід словоформи: чоловічий (m), жіночий (f), середній (n). Використовується для іменників, прикметників, дієприкметників та дієслів минулого часу.
number	"sg", "pl"	Число: однина (sg) або множина (pl).
person	"1", "2", "3"	Особа дієслів у теперішньому та майбутньому часі, а також наказовому способі.
case	"nom", "gen", "dat", "acc", "ins", "loc", "voc"	Відмінок: називний (nom), родовий (gen), давальний (dat), знахідний (acc), орудний (ins), місцевий (loc), кличний (voc).
aspect	"perf", "imperf"	Вид дієслова: доконаний (perf) або недоконаний (imperf). Визначає завершеність дії.
type	"synthetic"	Тип форми: у складеному словнику позначає синтетичне утворення майбутнього часу для дієслів недоконаного виду.
alt	true	Позначає альтернативну форму тієї ж словоформи (наприклад, історичні або стилістичні дублети).

resources

Тип: array of strings

Опис: назви файлів, з яких узято цю лему.

Примітки: дозволяє відстежувати походження даних; не є обов'язковим складником словника.
