

Deep Learning for search project

Images2Recipes

Mariia Shmakova

Anna Gromova

Aleksandra Kuzmich

The general idea in
short

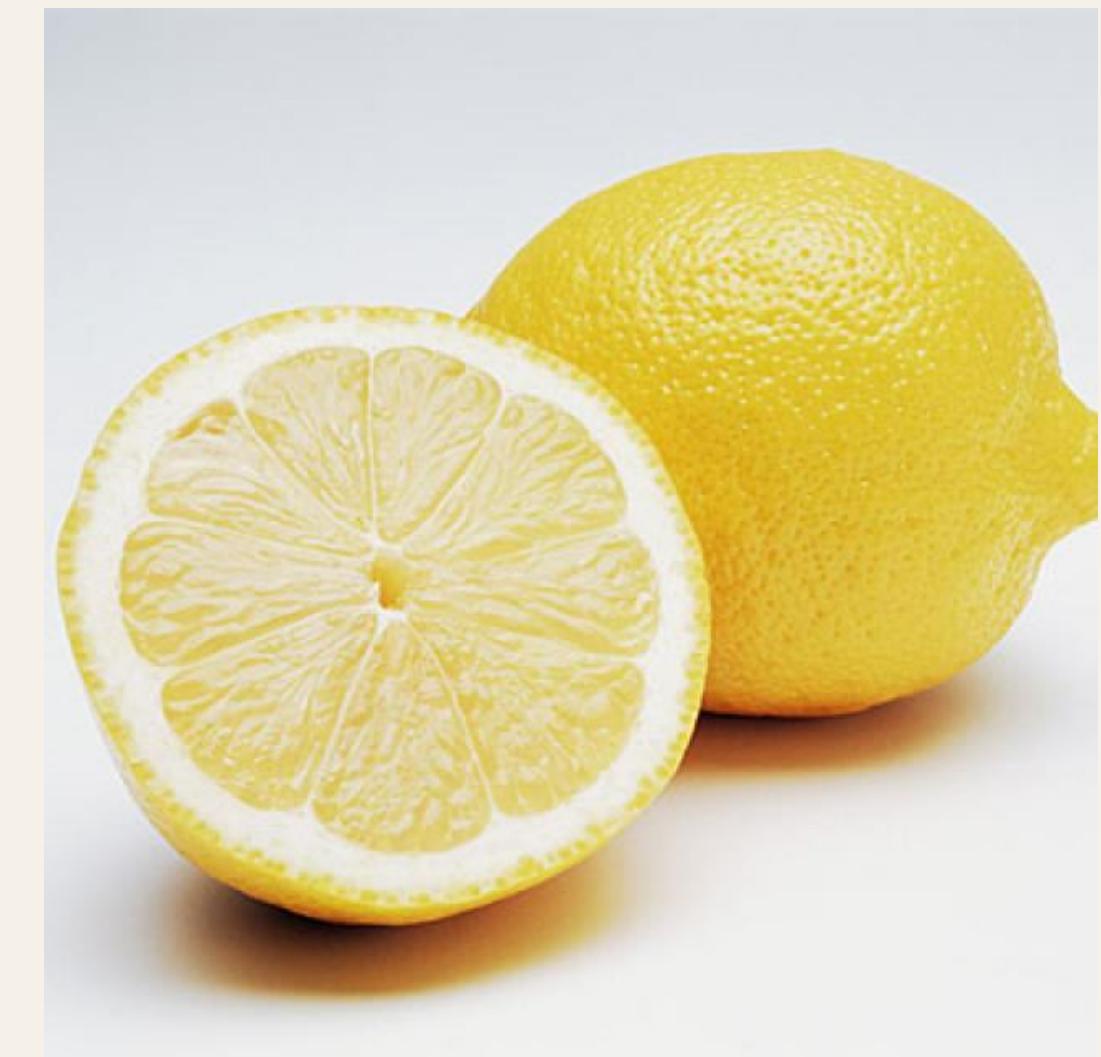
We have few ingredients in the fridge and we don't know what to cook, for example:



A. pineapple



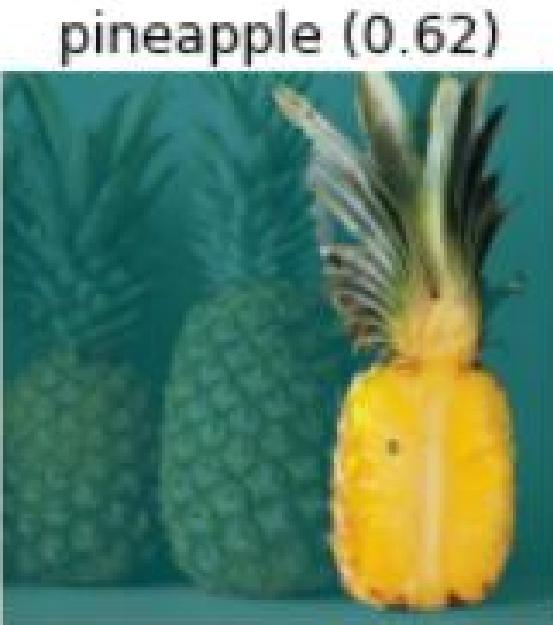
B. orange



C. lemon

- Verification and preprocessing of recognized ingredients

```
▶ display_batch_of_images(images, predictions)
```



1. Upload photos
2. Wait for a while until the ingredients in the photo are recognized

> Correcting incorrectly recognized ingredients

[Show code](#)

Photo 0

Photo 1

Photo 2



If it's not pineapple, write the correct name below

Right name

3. Check the correctness of recognition and in the opposite case write the name of the ingredient

4. Wait until synonyms for ingredients are found and get the best matching recipes

```
INITIAL ['лимон', 'апельсин', 'ананас']
```

Search results for INITIAL :

```
{"ingredient_names": "[['морковь', 'яблоко', 'апельсин', 'банан']]", "name": "Витаминный салатик", "url": "https://www.povarenok.ru/recipes/show/14613/"}  
{"ingredient_names": "[['лимон', 'апельсин', 'морковь']]", "name": "Цимес", "url": "https://www.povarenok.ru/recipes/show/73523/"}  
{"ingredient_names": "[['чай зеленый', 'яблоко', 'мед', 'вермут']]", "name": "Коктейль \"Осенний джаз\"", "url": "https://www.povarenok.ru/recipes/show/68373/"}  
{"ingredient_names": "[['яблоко', 'апельсин', 'лайм']]", "name": "Украшения для торта", "url": "https://www.povarenok.ru/recipes/show/62379/"}  
{"ingredient_names": "[['дайкон', 'морковь', 'гвоздика']]", "name": "Журавлик из дайкона", "url": "https://www.povarenok.ru/recipes/show/50636/"}
```

```
SYNONYMS ['апельсин', 'банан', 'банан']
```

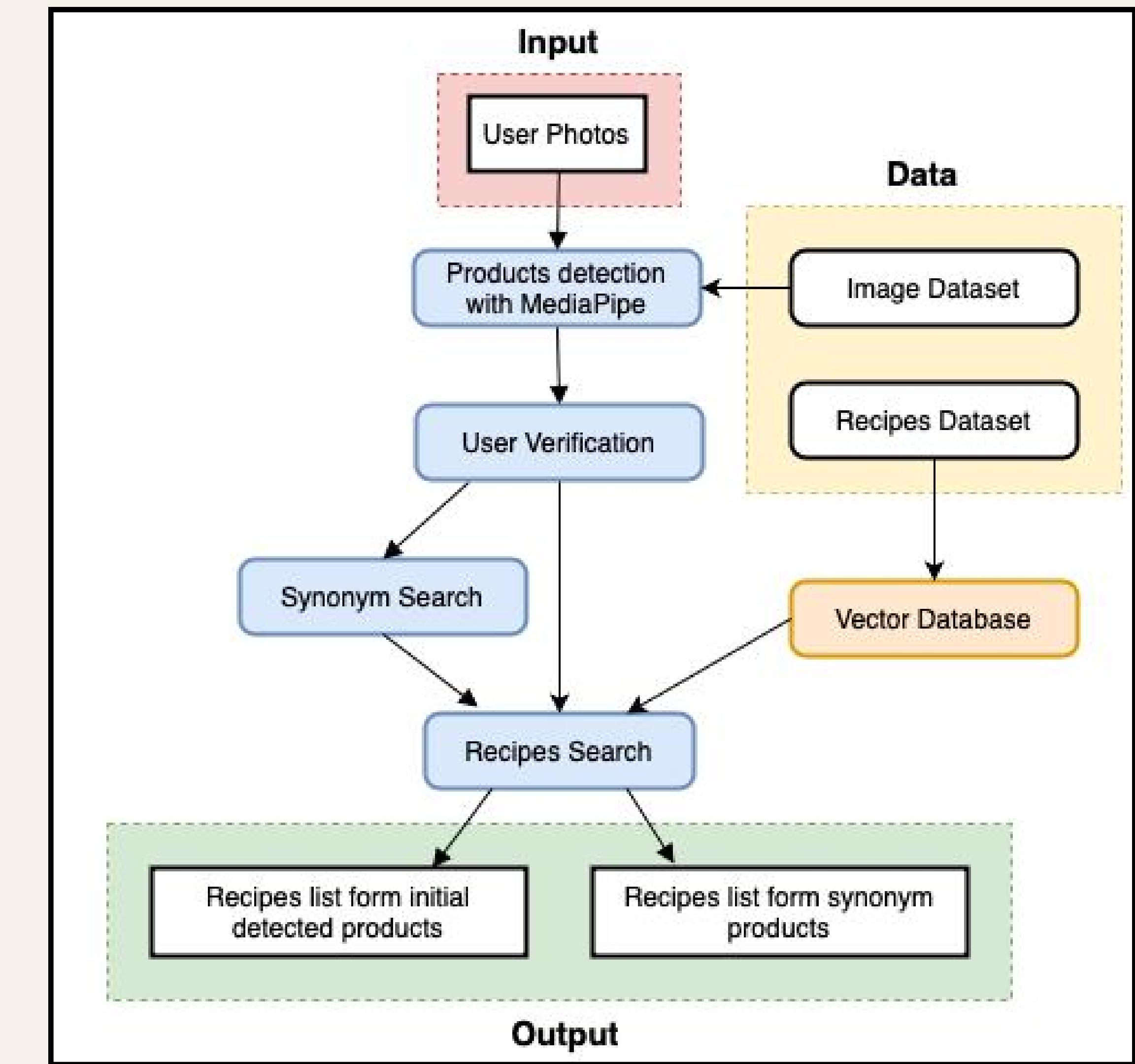
Search results for SYNONYMS :

```
{"ingredient_names": "[['морковь', 'яблоко', 'апельсин', 'банан']]", "name": "Витаминный салатик", "url": "https://www.povarenok.ru/recipes/show/14613/"}  
{"ingredient_names": "[['груша', 'апельсин', 'бадьян']]", "name": "ДискоТекА \"БГ\" или Бешеная Груша", "url": "https://www.povarenok.ru/recipes/show/49755/"}  
{"ingredient_names": "[['яблоко', 'апельсин', 'лайм']]", "name": "Украшения для торта", "url": "https://www.povarenok.ru/recipes/show/62379/"}  
{"ingredient_names": "[['тыква', 'сахар', 'апельсин']]", "name": "Варенье из тыквы с апельсином", "url": "https://www.povarenok.ru/recipes/show/168013/"}  
{"ingredient_names": "[['черника', 'банан', 'сахар']]", "name": "Чернично-банановое варенье", "url": "https://www.povarenok.ru/recipes/show/111624/"}
```

Our project ***simplifies the cooking*** process by making it easy to identify ingredients from a photo and search for the perfect recipe.

Say goodbye to endless browsing of recipe websites - with our technology, you can just ***take a pictures*** and discover a whole world of ***culinary possibilities***.

Architecture of our system



Key technologies

- **MediaPipe**
It provides a suite of libraries and tools for you to quickly apply machine learning techniques in your applications. You can plug these solutions into your applications immediately, customize them to your needs, and use them across multiple development platforms.
- **RusVectōrēs**
A service for exploring semantic relationships between words using distributional models
- **Qdrant**
Vector data management system. The program allows you to create chatbots, recommendation systems and other products based on artificial intelligence.

Image preprocessing

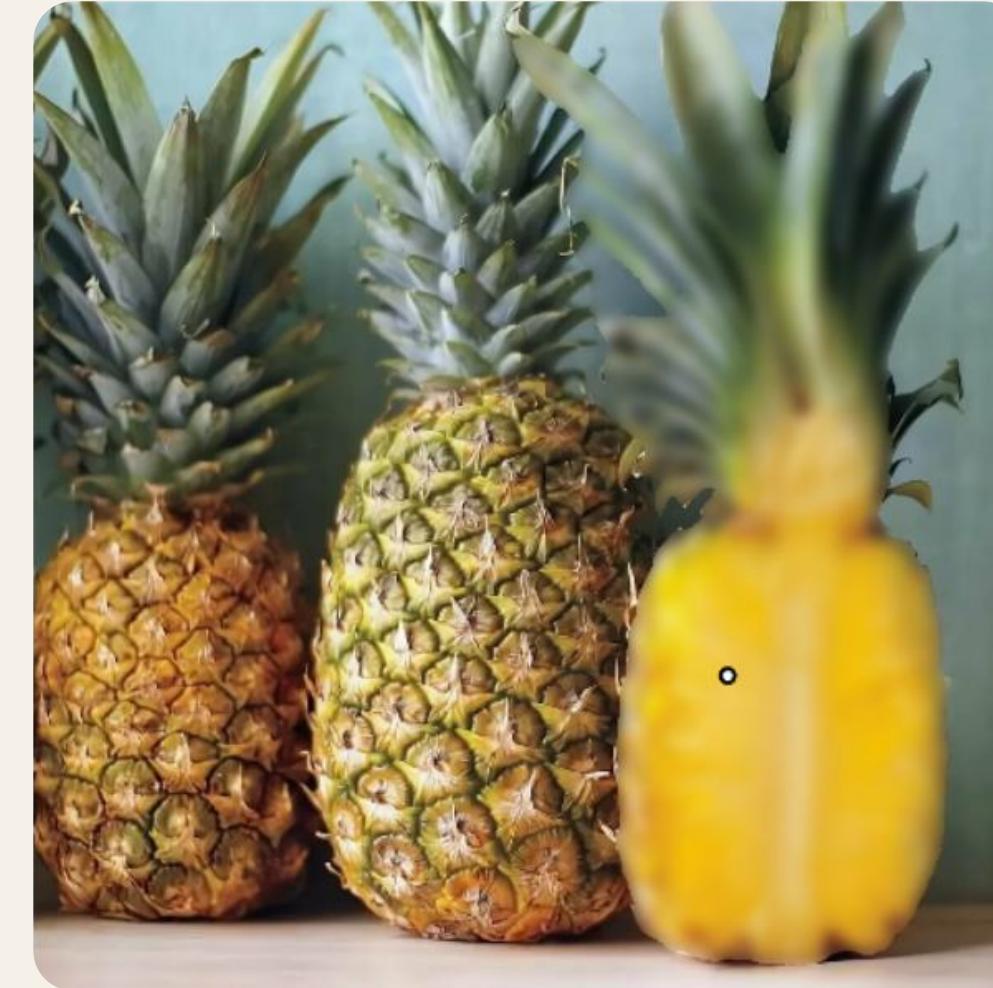


Initial image



Separating object

This example will separate the background and foreground of the image and apply separate colors for them to highlight where each distinctive area exists.



Blurring

Now we take it a step further and blur the background.



Overlaying color on background

We highlight the selected object with an overlay color (rather than a background blur).

Ingredient recognition

Number of ingredients it can
recognize

3

Custom model

12

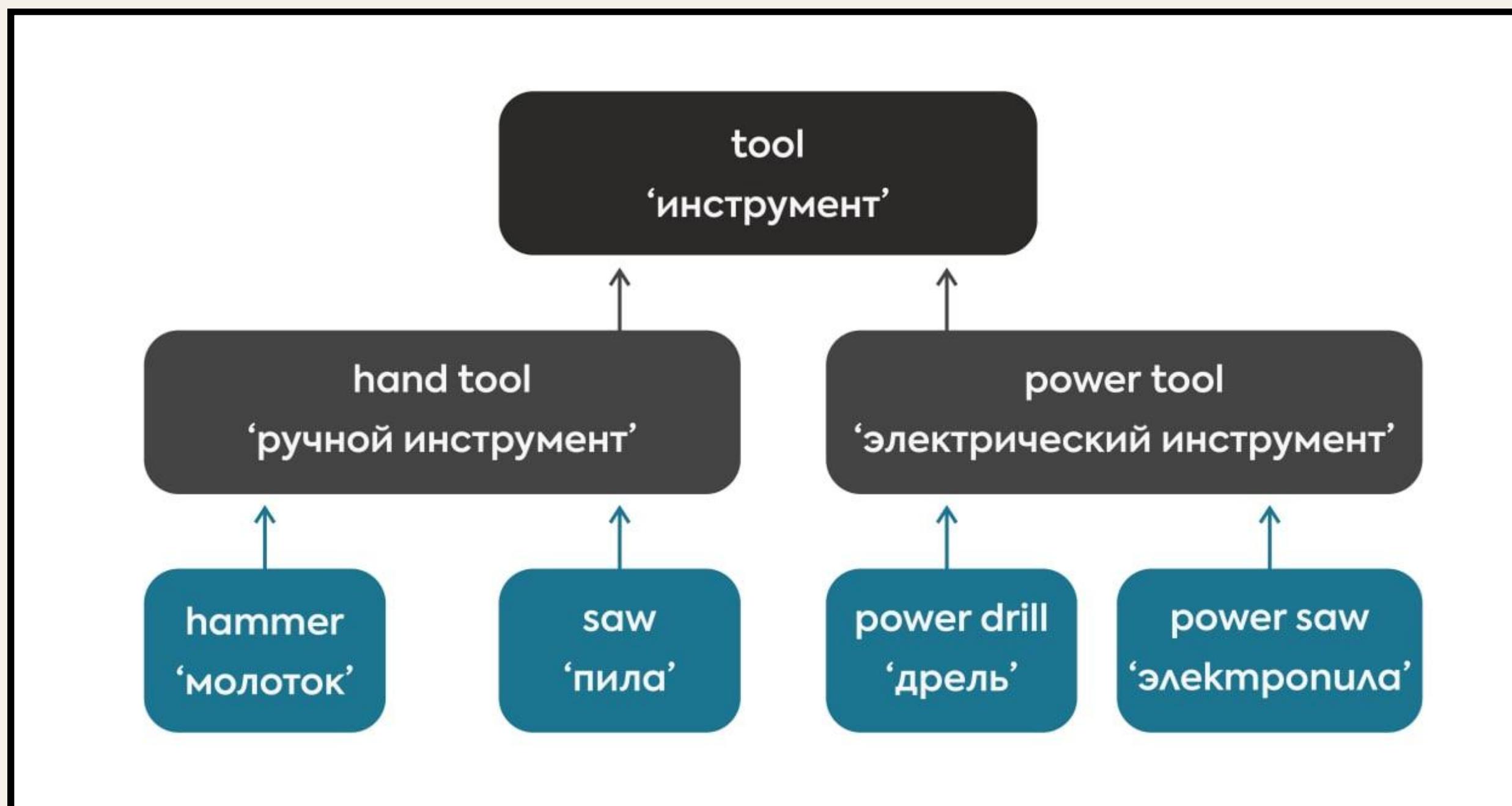
ImageAI

39

MediaPipe

Synonyms Search

RuWordNet



```

for sense in wn.get_senses('горох'):
    print(sense.synset)

Synset(id="125936-N", title="ГОРОШЕК (КРАПИНЫ)")
Synset(id="125937-N", title="ГОРОШИНА, ЗЕРНО ГОРОХА")
Synset(id="6857-N", title="ГОРОХ (РАСТЕНИЕ)")
  
```

Hypernyms for sense ГОРОХ (РАСТЕНИЕ) :

Нуоронимы для ЗЕРНОБОБОВАЯ КУЛЬТУРА: ФАСОЛЬ (РАСТЕНИЕ)
СОЯ (РАСТЕНИЕ)
ТУРЕЦКИЙ ГОРОХ
ГОРОХ (РАСТЕНИЕ)
ЧЕЧЕВИЦА (РАСТЕНИЕ)

Нуоронимы для ГОРОШИНА, ЗЕРНО ГОРОХА :

Нуоронимы для БОБ (ПЛОД) :
СОЕВЫЕ БОБЫ
ЧЕЧЕВИЦА (ПЛОДЫ)
ФАСОЛЬ (ПЛОДЫ)
ГОРОШИНА, ЗЕРНО ГОРОХА

Нуоронимы для ЗЕРНО РАСТЕНИЯ :
ЧЕЧЕВИЦА (ПЛОДЫ)
ЗЕРНО ХЛЕБНЫХ ЗЛАКОВ
ЗЕРНА МАКА
ГОРОШИНА, ЗЕРНО ГОРОХА
КОФЕЙНЫЕ ЗЕРНА

Synonyms Search

RusVectōrēs

The most similar words to 'чечевица_NOUN' are:

- фасоль_NOUN (similarity: 0.54)
- горох_NOUN (similarity: 0.51)
- ячмень_NOUN (similarity: 0.49)
- боб_NOUN (similarity: 0.48)
- шафран_NOUN (similarity: 0.47)

Initial list: ['лимон', 'цветная капуста', 'курица', 'помидор', 'молоко', 'кефир', 'чечевица', 'лук']

Synonyms list: ['апельсин', 'огурец', 'гусь', 'баклажан', 'простокваша', 'йогurt', 'чечевица', 'лук']

RuWordNet:

```
wn.get_senses('помидор')[0].synset.hypernyms[0].hyponyms[0]  
Synset(id="144181-N", title="БАКЛАЖАН (ПЛОД)")
```

Dataset

Source: <https://www.kaggle.com/datasets/rogozinushka/povarenok-recipes>

147k culinary recipes

['url', 'name', 'ingredients']

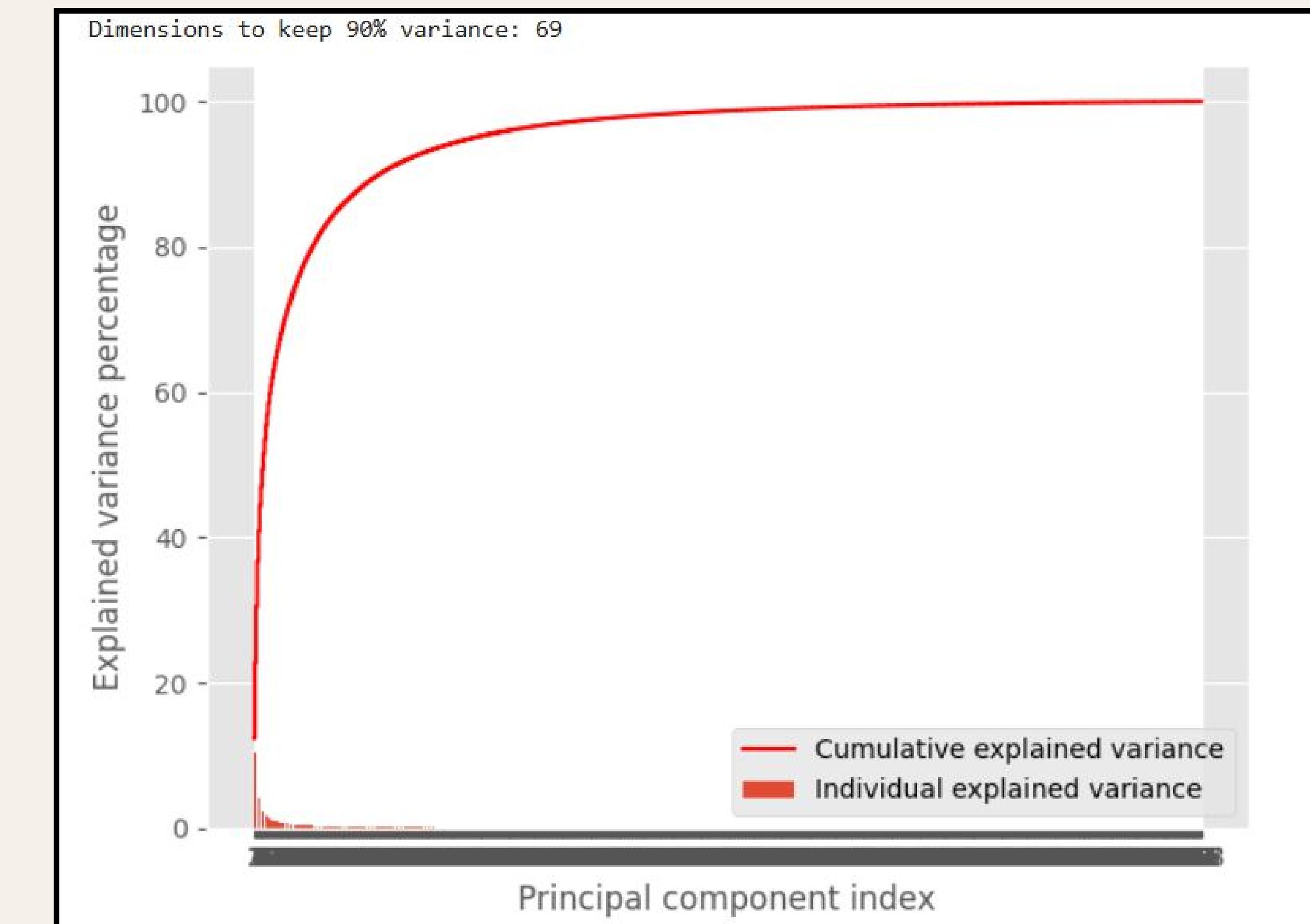
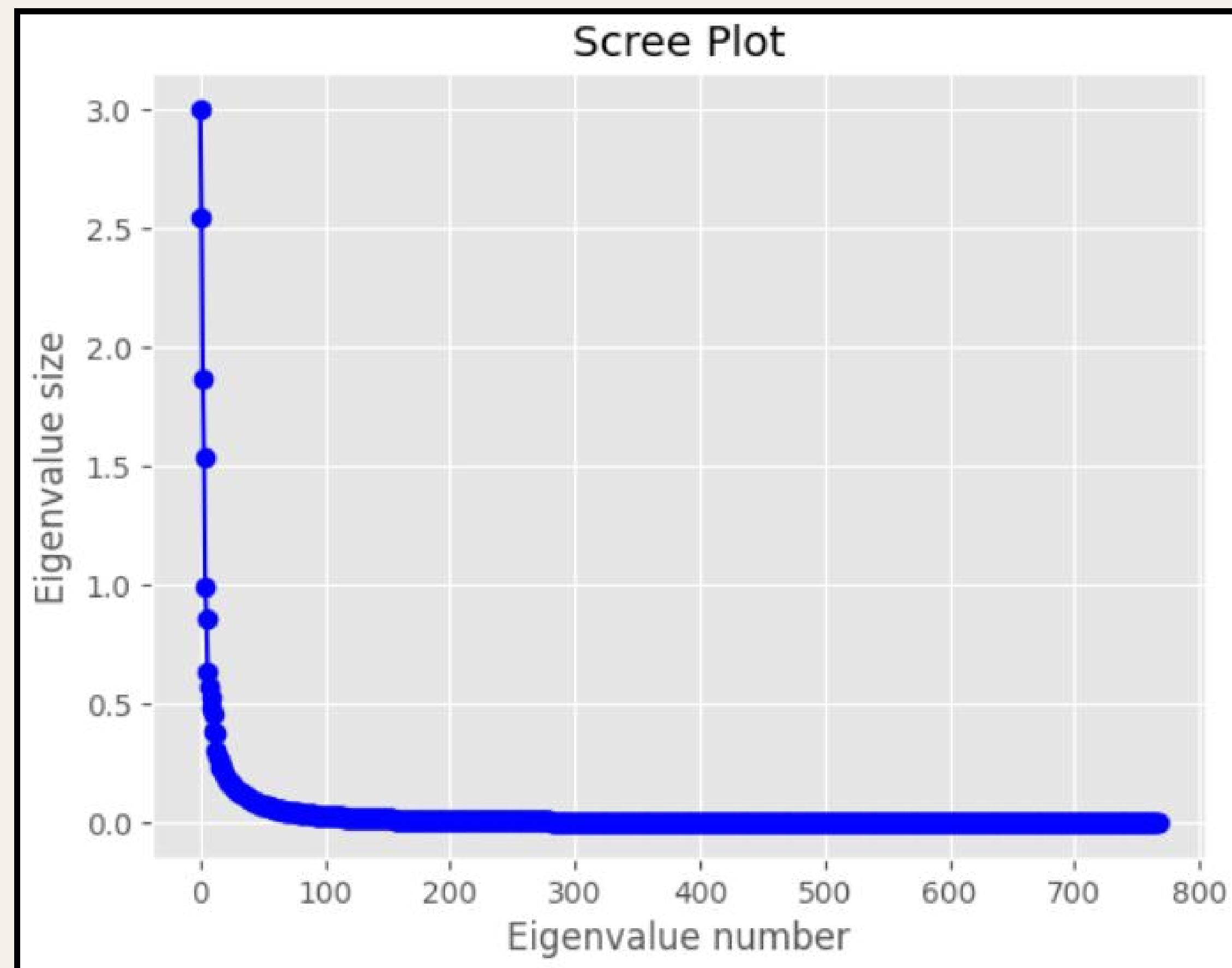
1. Generate embeddings via “DeepPavlov/rubert-base-cased”
2. Calculate embedding means
3. Bin embedding means to quantiles
4. Perform stratified sampling

Mean variance of the original dataset: 0.03166067382127971

Mean variance of the sampled dataset: 0.03167285855004068

Improving VECTOR DIMENSIONS

768 → 70



Improving

SEARCH SPEED

- string: ['фарш мясной', 'морковь', 'лук репчатый', 'сыр твердый', 'соль', 'специи']

OLD	NEW
0.064	0.062

- string: ['лимон', 'вода', 'сахар', 'лук', 'чеснок']

OLD	NEW
0.086	0.059

- string: ['курица']

OLD	NEW
0.069	0.064

Improving

SEARCH QUALITY

Initial recipies match	Synonym recipies match	Initial accuracy	Synonym accuracy
1	2	0,33	0,67
2	1	0,67	0,33
0	1	0	0,33
1	1	0,33	0,33
0	1	0	0,33

Initial accuracy	Synonym accuracy
0,27	0,4

Key metrics

for image recognition

0.67
Accuracy

1.00
Recall

0.67
Precision

0.80
F1

Final evaluation

	1 doc	2 doc	3 doc	4 doc	5 doc		
String 1	1	1	1	0	1	Mean precision:	0,62
String 2	1	0	0	1	0	Mean Kendall-Tau	0,8
String 3	1	1	0	0	1	Mean average precision:	0,928819
String 4	1	0	0	0	0		
String 5	1	0	1	1	1		
String 6	1	1	1	0	1		
String 7	1	1	0	0	0		
String 8	1	1	1	1	0		
String 9	1	0	1	0	0		
String 10	1	1	1	1	0		
String 11	1	1	1	0	0		
String 12	1	1	1	1	1		
String 13	1	1	1	0	0		
String 14	1	0	0	1	0		
String 15	1	1	0	0	0		
String 16	1	1	1	1	1		
String 17	1	1	1	1	1		
String 18	1	0	0	0	0		
String 19	1	1	0	1	0		
String 20	1	0	1	0	1		

Search algorithm

ANN - Hierarchical Navigable Small World (HNSW)

5 approaches that we found to be the most effective

- **Image preprocessing**
The ingredient is recognized more accurately if preprocessing is done first. This involves selecting an object in the image and using the resulting data to apply effects such as color overlays that highlight the object and blur the background around it.
- **Synonyms search**
Searching for related products from a set of ingredients used in the recipies
- **Products verification**
Asking user to verify the identified products
- **Dimensionality reduction**
Preserve high variance of data by using PCA analysis
- **Products preprocessing**
Bringing predicted products to their initial form before synonym search

Hardware requirements for development and inference

- **Default Colab requirements**

THE DEFAULT CPU FOR COLAB IS AN INTEL XEON CPU WITH 2 VCPUS (VIRTUAL CPUS) AND 13GB OF RAM.

- **Internet connection**

- **Default Qdrant requirements**

Thank you!