Explainable Artificial Intelligence for species identification

Maya Sahraoui

Identification in biology

Definition :

- Matching a **living specimen** to its **taxonomic group**.

- Mostly based on **morphological (visual) characteristics** of the group.

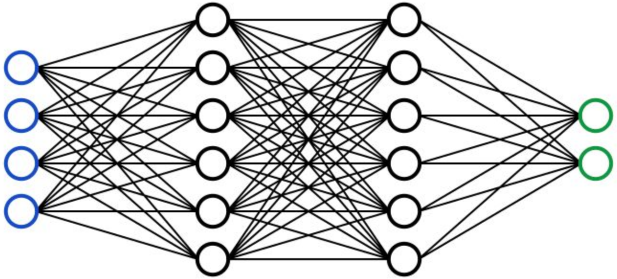
Identification in biology

Identification in biology

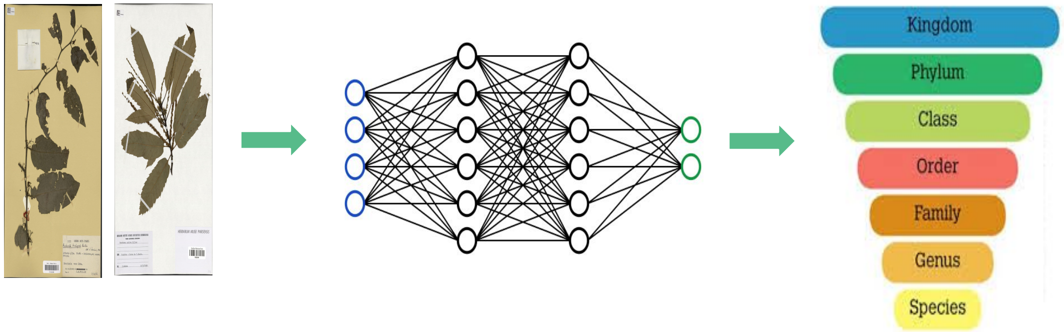
Muséum Herbarium

Flora corpora

How to automate this task ?



How to automate this task ?

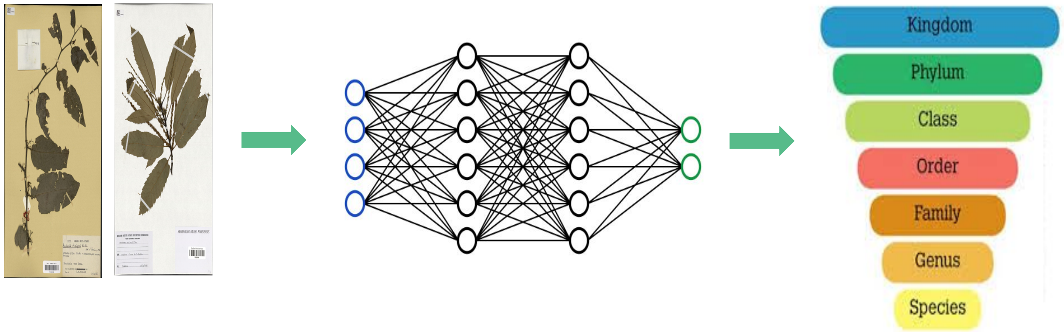


- Fully supervised

- Data is available (Herbarium of the National Museum of Natural History)

5 millions numerized specimen

How to automate this task ?

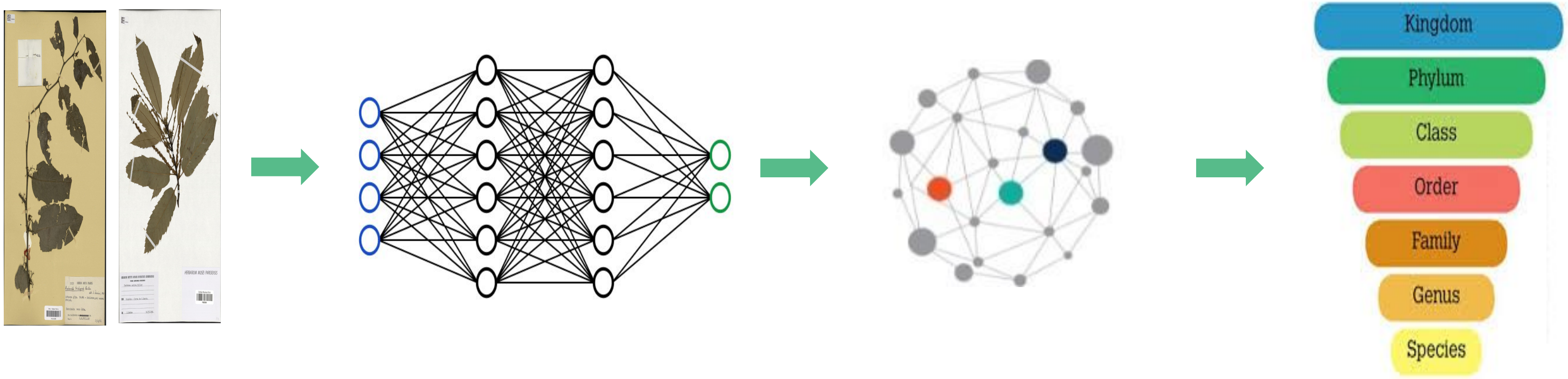


- What if the model gets it wrong ?
- What makes the model identify a specimen to a specific species ?
- Can the model explain its decision ?

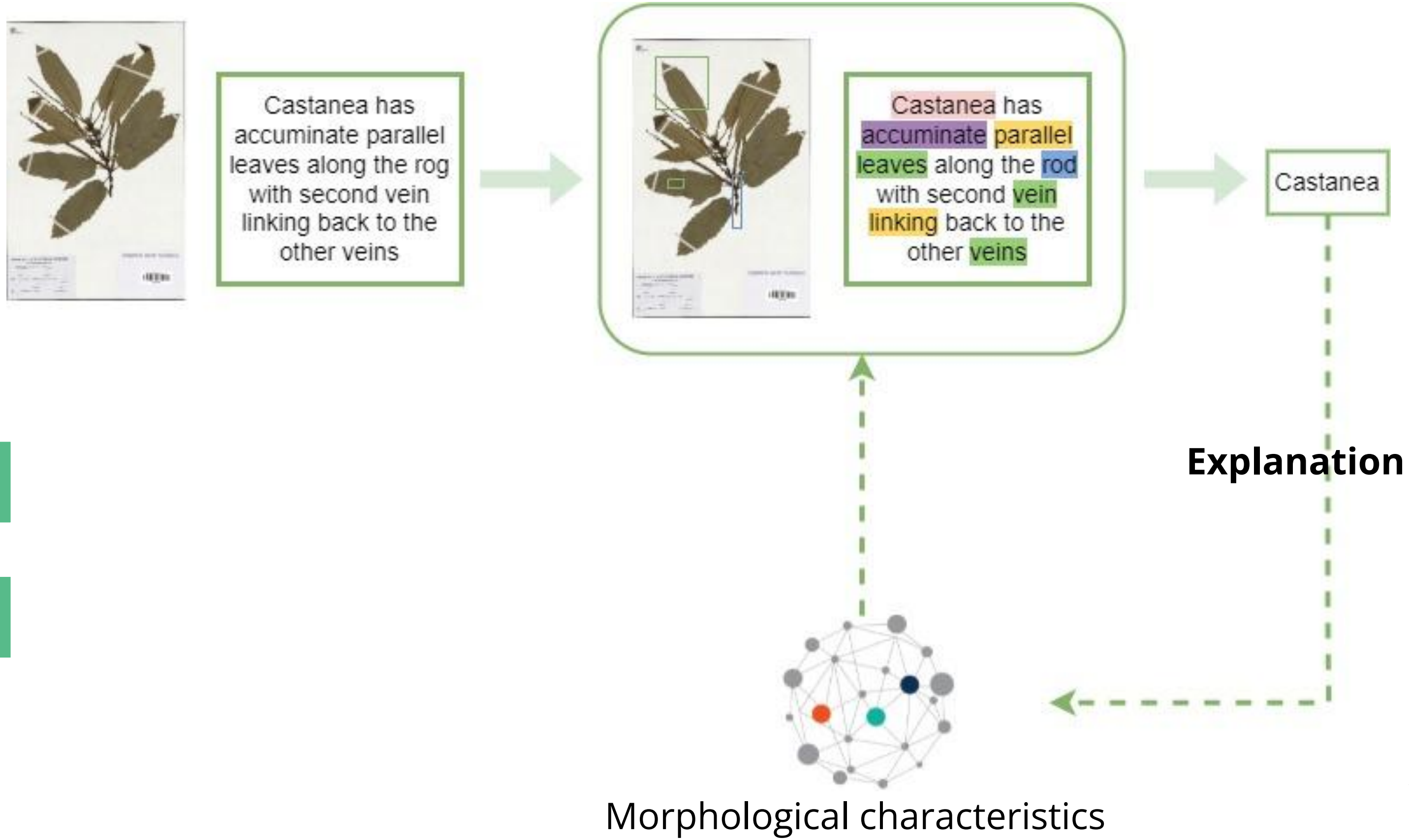
Black box

We need explainability !

What if we make use of the species descriptions ?



Proposed method



Text processing

Image processing

Proposed method

Text processing

Image processing

Extracting morphological characteristics from the texts.

bracts 7-10, chartaceous, ovate, sometimes marginally ciliolate; lamina sometimes succulent ovate or elliptic basally cruneate.

- (bracts, form, ovate)
- (bracts, form, Ciliolate)
- (lamina, surface, succulent)
- (lamina, form, ovate)
- (lamina, form, elliptic)
- (lamina, form, cruneate)

Proposed method

Text processing

Image processing

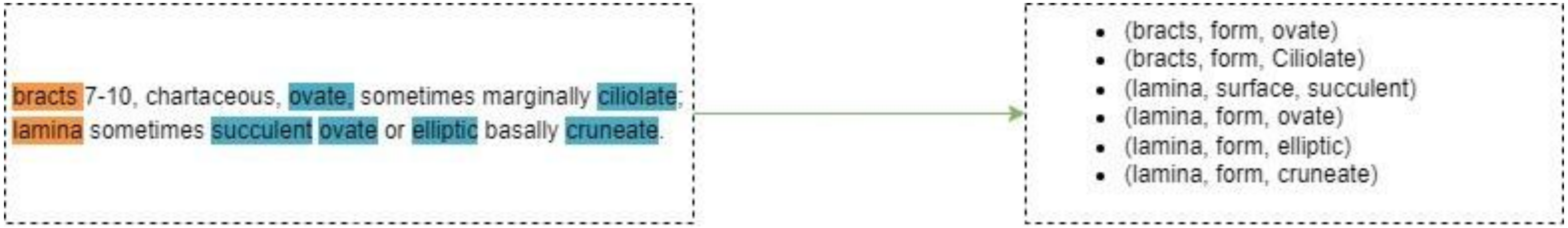
Verifying the presence of the characteristics in images

- (bracts, form, ovate) ✓
- (bracts, form, Ciliolate) ✗
- (lamina, surface, succulent) ✓
- (lamina, form, ovate) ✗
- (lamina, form, elliptic) ✓
- (lamina, form, cruneate) ✗

Text processing

Image processing

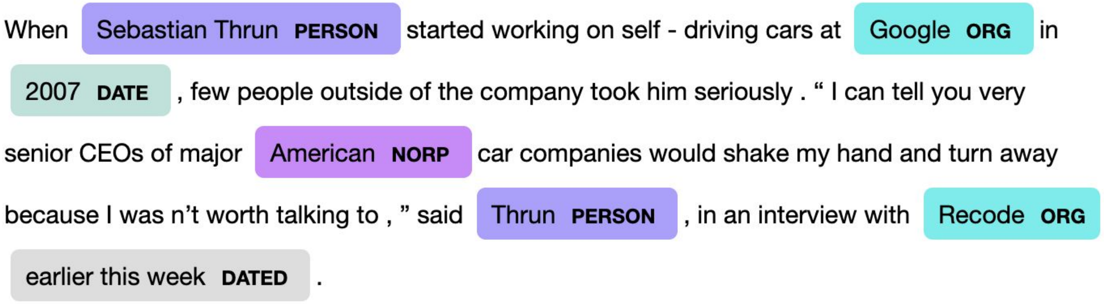
Identify the task



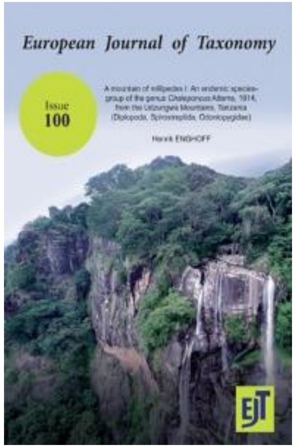
Named entity recognition
Relation extraction

Identify the task

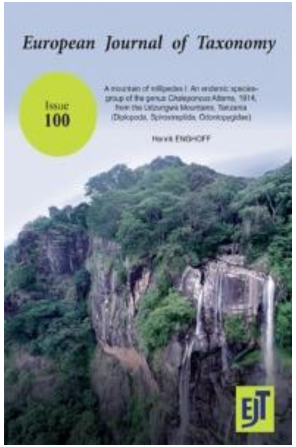
- Historical task.
- Entities detection and classification.
- Measure of performance for **detection** and **classification**.



Check the data



Check the data

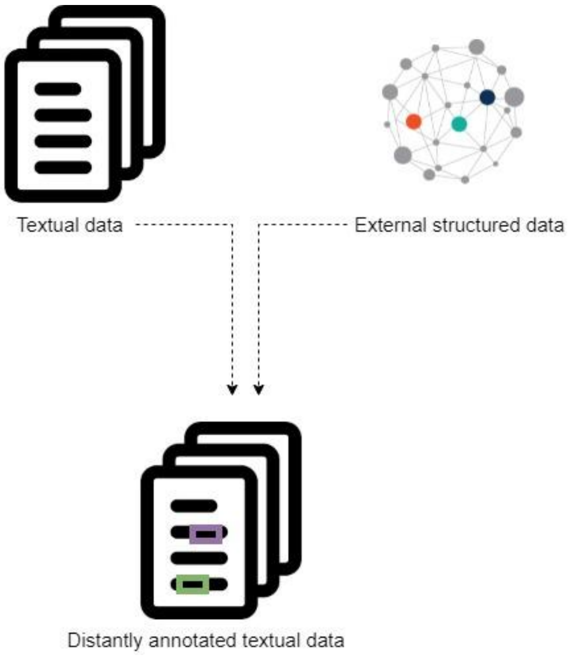


No text-to-graph direct matching

Distant supervision

Check the data

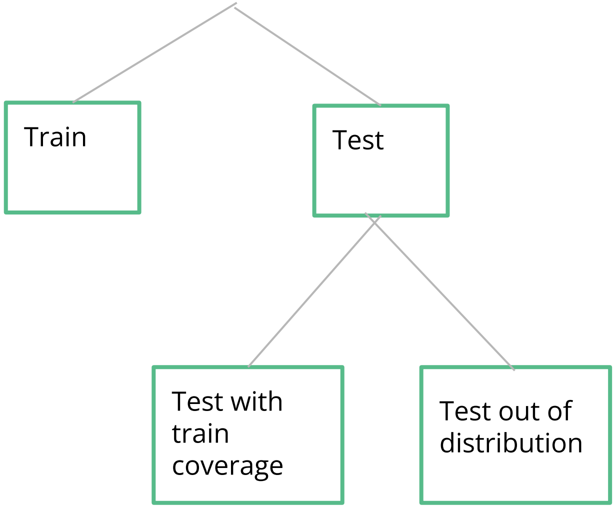
What's distant supervision ?



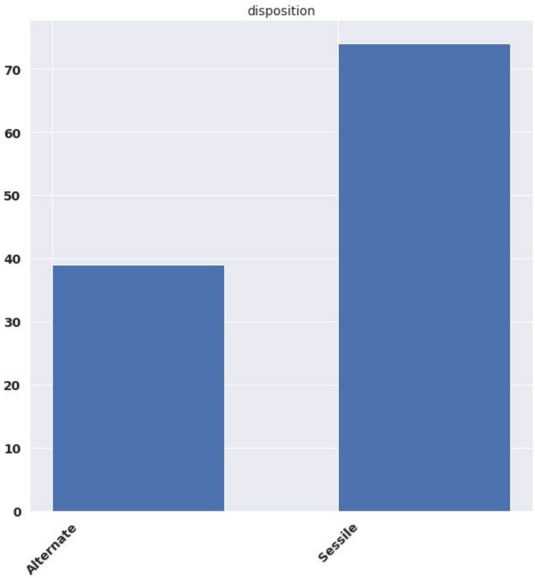
Bias in annotations

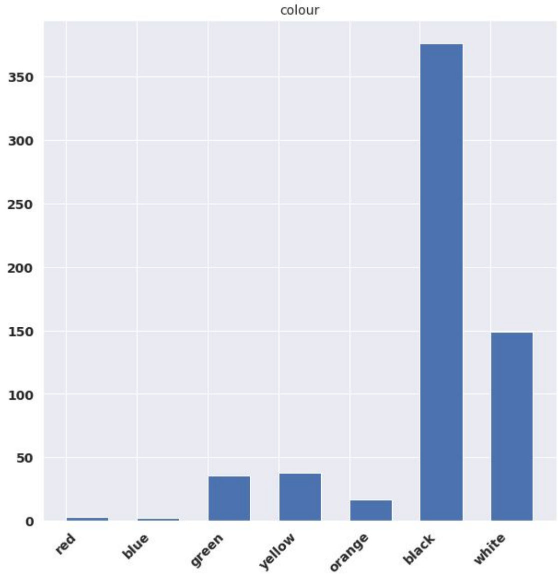
Create the dataset

Ensemble of all sentences



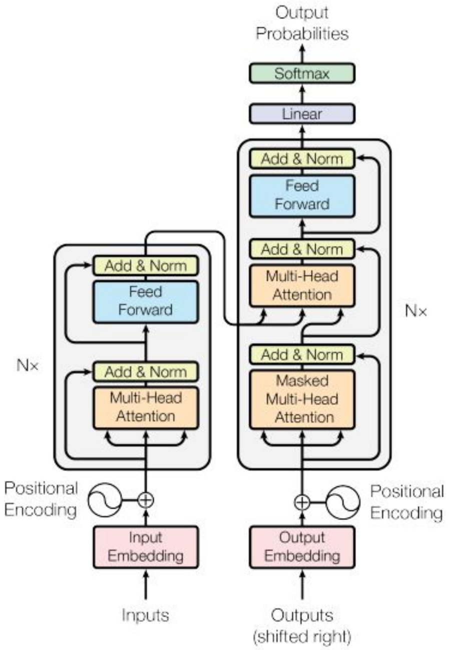
Check the histograms



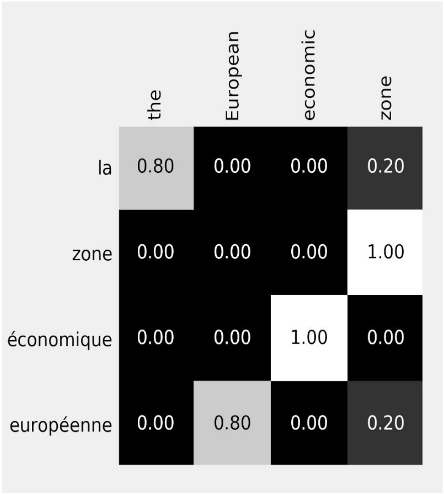


Model architecture

Transofrmers



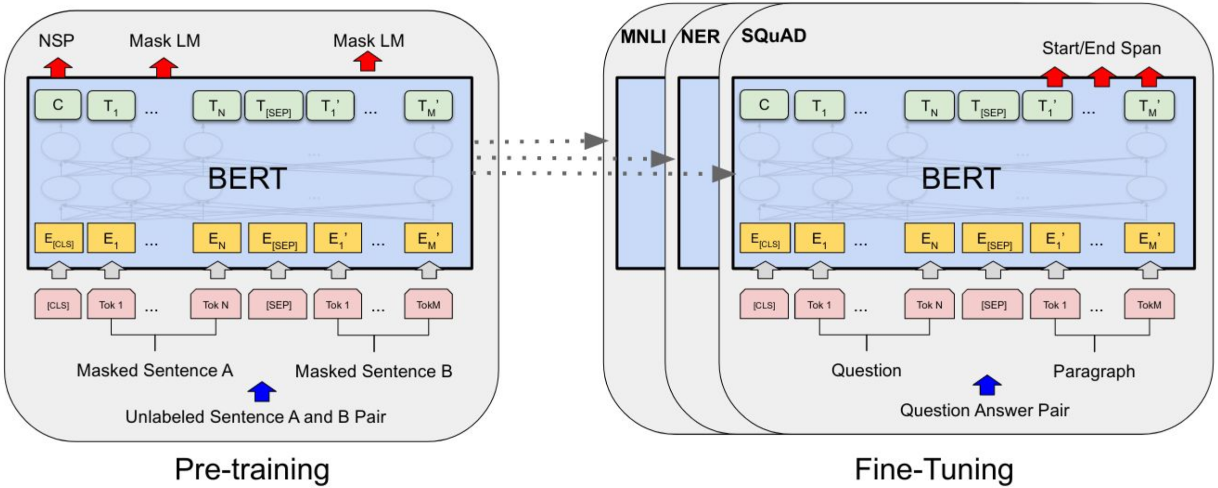
Attention mechanism



Transformers : Attention Is All You Need
(Ashish Vaswani et al, 2017)

Model architecture

Transformers



Bert : Transformer encoder.

- Next sentence prediction
- Masked word prediction

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Results

Détection

	X (New sentences)			Xc (new sentences without “O”)			Xt (New sentences + new words)		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Baseline w/ distant supervision	95.46	95.64	95.50	100	72.36	83.97	100	60.82	75.64
Baseline w/ distant supervision + lm	**96.29**	**96.41**	**96.23**	100	72.54	84.08	100	59.02	74.23
Baseline w/ distant supervision + knowledge dist.	91.64	85.36	87.40	100	**82.16**	**90.21**	100	78.46	87.93
Baseline w/ co-occurence prior	94.93	94.84	94.88	100	77.09	87.06	92.35	**92.61**	**92.19**

Results

Classification

	X (New sentences)			Xc (new sentences without “O”)			Xt (New sentences + new terms)		
	Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Baseline w/ distant supervision	94.4	94.8	94.5	61.5	**72.4**	65.3	87.4	56.9	68.4
Baseline w/ distant supervision + lm	**95.8**	**96.0**	**95.6**	**85.2**	64.9	**75.05**	**93.5**	58.3	71.0
Baseline w/ distant supervision + knowledge dist.	91.9	85.1	87.5	41.2	67.1	54.15	86.9	**68.4**	**74.6**
Baseline w/ prior on co-occurences	94.1	93.9	93.8	68.9	62.5	63.8	88.34	61.9	71.9

Results

Baseline

Inflorescences: terminal, few- or much-branched, few- or many-flowered[Auteur in1] , elongated, paniculate cymes with spirally arranged paracladia, peduncles of main florescences (4—)8.4—12 cm long and those of coflorescencse (1.5—)7—8.5 cm, bracts elliptic, 5-8 mm long, margins with knob shaped, stipitate. glands; glabrous or pubescent on both surfaces; pedicels 1.8-2 mm long. Capsules: ellipsoid, 1-1.1 χ 0.8-1 cm, surface glabrous, pubescent, or pilose, often with some red coloration. Seeds: creamy white with red-brown mottling, oval, 7-8 χ 4-4.5 mm, caruncle large and conspicuously lobed.

Knowledge distillation

Inflorescences: terminal, few- or much-branched, few- or many-flowered[Auteur in1] , elongated, paniculate cymes with spirally arranged paracladia, peduncles of main florescences (4—)8.4—12 12 cm long and those of coflorescencse (1.5—)7—8.5 cm, bracts elliptic, 5-8 mm long, margins with knob shaped, stipitate glands; glabrous or pubescent on both surfaces; pedicels 1.8-2 mm long. Capsules: ellipsoid, 1-1.1 χ 0.8-1 cm, surface glabrous, pubescent, or pilose, often with some red coloration. Seeds: creamy white with red-brown mottling, oval, 7-8 χ 4-4.5 mm, caruncle large and conspicuously lobed.

- colour
- development
- disposition
- ecology-geography
- flower
- form
- fruit
- leaf
- margin leaf
- Taxon
- part of
- plant
- position
- structure
- surface

Qualitative results over two texts of two variants of the model, the underlined tokens are the named entities present in the test set as a ground truth