

# Prototyping a vocabulary trainer with implicit crowdsourcing of Language Resources (LRs)



Results of Crowdfest task 3

Nikos Isaak, Anisia Katinskaia , Alexander König, Verena Lyding, Lionel Nicolas, Georgia

Matheou, Christos Rodosthenous, Federico Sangati, Umair Ul Hassan, Jolita Horbacauskiene



# Overview

- Background
- Task objective
- Task architecture and implementation
- Challenges
- Next steps





# Background

## **Aim:**

- Demonstrate the ‘implicit crowdsourcing’ paradigm
- By creating a sample implementation

## **Approach: Vocabulary training**

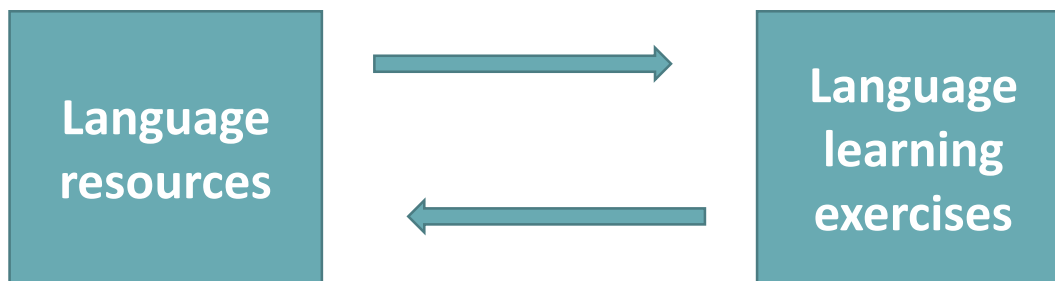
- Well-confined use case
- Still with lots of options for variability





## Background

### 'Implicit crowdsourcing'



Use LRs to create exercises, crowdsource on the learners to provide data to enhance the language resources.





## Objective of task 3

Prototype an **online vocabulary trainer** with the following characteristics:

- Exercises are automatically generated from existing language resources
- Learners answers are used to improve/extend the existing language resources

Based on the common sense ontology ConceptNet  
(<http://conceptnet.io>)

Simple vocabulary exercises on word meaning





# Language Resource ConceptNet

ConceptNet network of semantic relations between nouns, such as: types of, is located at, part of, etc.


## tree

An English term in ConceptNet 5.6

**Sources:** Open Mind Common Sense contributors, CC-CEDICT 2017-10, DBPedia 2015, JMDict 1.07, OpenCyc 2012, Unicode CLDR, Verbosity players, German Wiktionary, English Wiktionary, French Wiktionary, and Open Multilingual WordNet

### Types of tree


 b tree →

 Something you find  
outside →


 Coca →

 cherry tree →

### Things located at tree

 a bird →

 a leaf →

 fruit →

 a snake →





## Exercise design

Learning target: vocabulary training of nouns

Learner group: elementary level (A1)

The user is prompted to provide words, which are related to a given word.

- Question: “*Name one thing that is related to X*”
- The input (“X”) is requested as free text.





## Subtasks

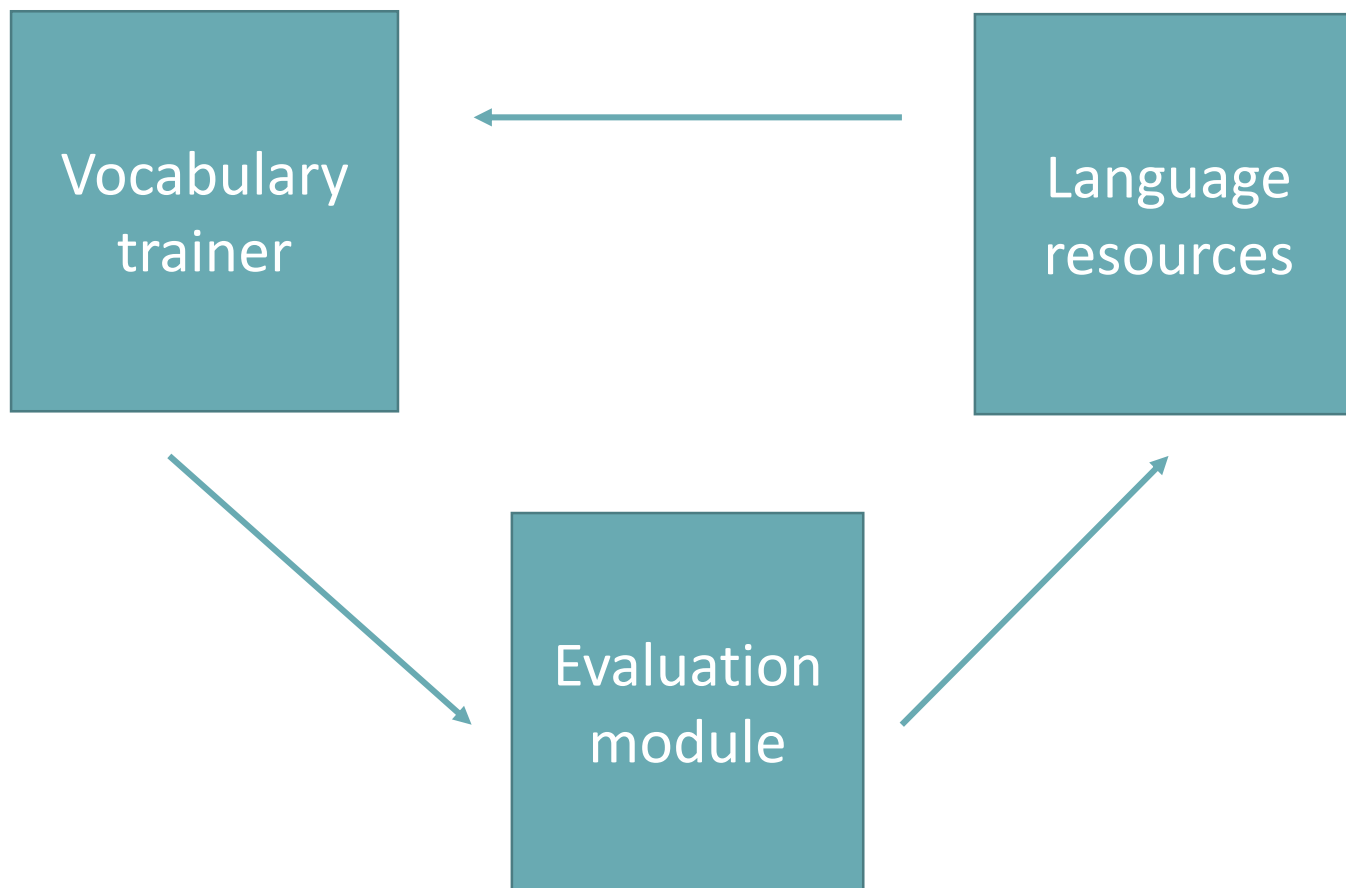
- Exercise generation from ConceptNet
- User interface design with exercises
- Validation mechanism to cross-match learner's answers







# Architecture





# Exercise generation interface

**ConceptNet Keyword**

cat,dog,bird

.....

**Difficulty Level:**

A1  A2  A3

**Category**

Animals

RelatedTo

AtLocation

PartOf

.....

.....

Add

1000





# Exercise interface

## BootStrap

### RELATION EXERCISE



Name one thing that is related to dog

Submit Answer

## Telegram





## Evaluation strategy

Candidates are collected until a threshold is met

- Aggregation mechanism: majority voting
  - Reward system: double points for new words
  - Badges for the first person suggesting a new word
- Collect data and adjust based on first insights





## Challenges encountered

- **Teachers:**  
Selection of material for different proficiency levels
- **Exercise generation:**  
Handling inflected forms etc
- **Evaluation:**  
Open-endedness of learner input
- **Learners:**  
Delayed feedback, obscure meaning of 'potential' points





## Next steps

- Consolidate the code
  - Finish some open features like leaderboards
  - Set up a set of exercises for an crowdsourcing experiment
  - Publish preliminary results
- Watch out for the call for participation.



# Any questions ?

---



Nikos Isaak, Anisia Katinskaia, Alexander König, Verena Lyding,  
Lionel Nicolas, Georgia Matheou, Christos Rodosthenous, Federico  
Sangati, Umair Ul Hassan, Jolita Horbacauskiene



# Architecture

