v-trel: Vocabulary Trainer for Tracing Word Relations – An Implicit Crowdsourcing Approach

Verena Lyding, Christos T. Rodosthenous, Federico Sangati, Umair ul Hassan, Lionel Nicolas, Alexander König, Jolita Horbacauskiene, Anisia Katinskaia

enetCollect

Abstract

We present v-trel, a vocabulary trainer on word relations that uses exercises generated from the language resource ConceptNet and crowdsources the learners' answers to enrich it further. An empirical evaluation with 60 non-native speakers was carried out and evaluated with the assistance of an expert from language teaching. It shows that the approach allows to efficiently gather new entries to expand ConceptNet, and suggests that v-trel has educational potential, even though in its current state some shortcomings could be identified on both ends.

An implicit crowdsourcing approach

V-trel builds on a prototype architecture for crowdsourcing language resources, which implements the following paradigm:

IMPLICIT CROWDSOURCING PARADIGM

IF a language resource (e.g., a lexicon) can be used to generate language learning exercises, then the answers of learners to these exercises can be used to improve the resource, which in turn will improve the quality and versatility of the exercises generated.

Vocabulary trainer on word relations

V-trel delivers interactive vocabulary exercises for learning word senses. For a given word, learners have to indicate words which are related to the word by a specific semantic relation, such as RelatedTo, PartOf, AtLocation etc..

The learners receive feedback on the inputted words and can request hints for unknown words.

Crowdsourcing a language resource

The learners are used as crowdworkers to enhance the language resource that the vocabulary exercises are generated from, namely the commonsense ontology ConceptNet. While using the vocabulary trainer for learning word senses the learners are providing their knowledge of related words which is collected and evaluated in order to validate and enhance the LR.

Data

Exercise data is generated from ConceptNet.

Table 1 shows a number of terms and a selection of their RelatedTo words.

Term	RelatedTo term	#terms
cat	animal, dog, house...	94
dog	friend, puppy, bone...	135
bird	chicken, chick, canary...	219
cow	sacred, animal, steak...	96
fish	water, creature, lure...	221

Table 1: An example of the terms retrieved to generate the experiment exercises. The subject is `RelatedTo` the object. In the last column the number of filtered terms is depicted.
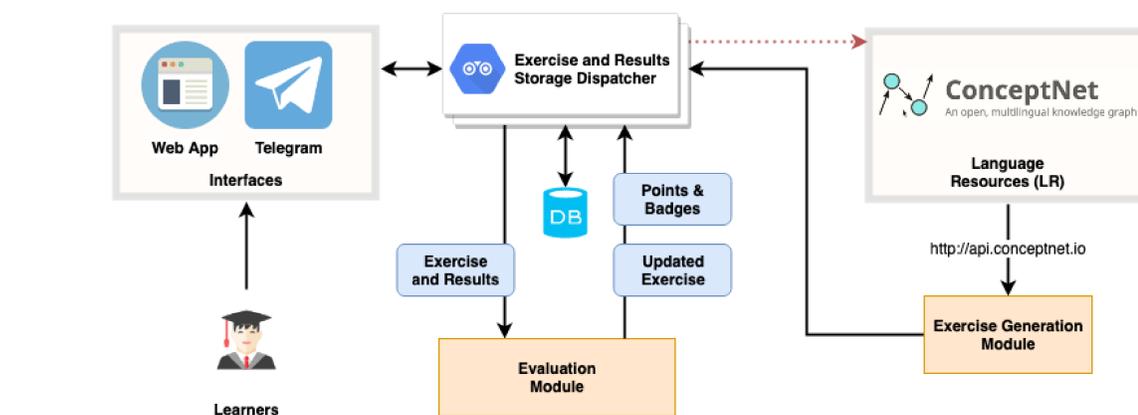
The architecture

The vocabulary trainer is composed of four modules:

1) The **exercise generation module** retrieves words from ConceptNet and generates exercise content.

2) The **exercise and result storage dispatcher** ingests the previously created exercise content, dispatches it to the interfaces and handles the responses of the learners.

3) The **evaluation module** checks if learners' contributions are fit for expanding the language resource and assigns points to learners according to the response given.

4) The *user interaction module with prototypical interfaces* presents exercises to the users and allows them to submit their responses and receive feedback.

Transactions between modules are handled using web services through API calls. Data are presented in a JSON format that can be consumed by any programmatically created interface. See project repository: https://gitlab.com/crowdfest_task3

User Interfaces – try it!

Evaluation experiment

Setup

26 exercises were generated from ConceptNet using the RelatedTo relation. Each exercise had a minimum of 20 related terms. Evaluation thresholds were set to 5 candidates and a minimum of 2 matches to include the new term. Also, answers following the "I don't know" hint button were recorded.

60 non-native speakers with high English proficiency were taking part in the experiment for 2 days. Participants received a link to both the Web interface and the Telegram Chatbot. After testing, participants compiled a feedback questionnaire.

Results

4533 contributions were gathered for the 26 exercises, they were collected in 44 distinct user sessions (17 on the Telegram chatbot and 27 on the Web interface).

449 new words were crowdsourced, 119 words on animals, 168 words on clothing, and 162 words on buildings.

New word	Frequency	Level
grass	15	basic
calf	6	advanced
meat	5	basic
cowboy	4	basic
farmer	4	basic
herd	4	advanced
horn	3	moderate
pasture	3	advanced

Table 2: Top new words for the term "cow", their frequency of mentions and proficiency levels.

A manual expert analysis of the proficiency level of all new words showed that 65% of all new words are basic vocabulary, 32% are of moderate level and 3% are advanced vocabulary.
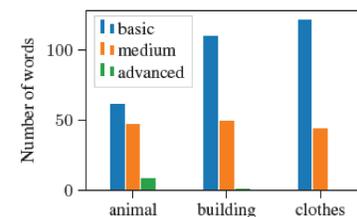
Querying ConceptNet for relations with the newly added words, showed: on average 14.15 words (per exercise) have no other direct relation, 100% have any bidirectional relation, of which 4.54 words have new relations (other than direct RelatedTo).

On average, learners named <= 40% of the words present in ConceptNet, while they can still learn >= 60%, e.g. through hints.

12 users approved the interactivity and simplicity of exercises, and the hints to learn new words, while 5 users suggested they would be difficult for beginners and feedback was inconsistent.

Conclusions

With v-trel, as a first instantiation of a prototypical architecture implementing the implicit crowdsourcing paradigm, a relevant and successful first step towards the crowdsourcing of language resources through language learning has been done!!

Acknowledgements

Questions? Contact us at: **verena.lyding@eurac.edu**