[ Report on Task 4 ]

Exploration of implicit crowdsourcing scenarios

Lionel Nicolas, Eurac Research, Italy
Verena Lyding, Eurac Research, Italy
Claudia Borg, University of Malta, Malta
Anabela Barreiro, INESC-ID, Portugal
Špela Arhar Holdt, University of Ljubljana, Slovenia
Jaka Čibej, University of Ljubljana, Slovenia
Julia Ostanina-Olszewska, Pedagogical University of Cracow, Poland
Silvia Pokrivčáková, University of Trnava, Slovakia
Michal Bodorík, University of Trnava, Slovakia

3rd enetCollect Annual Meeting, WG2 meeting; 15th March 2019, Lisbon, Portugal
Summary

➢ The team

➢ Generic Implicit crowdsourcing model

➢ Approach & foreseen outputs

➢ Current results & next steps

➢ Let’s make a quick game!
Summary

➢ The team

➢ Generic Implicit crowdsourcing model

➢ Approach & foreseen outputs

➢ Current results & next steps

➢ Let’s make a quick game!
The team

Language datasets

(Lionel)

Claudia)

Anabela)

Spela)

(Jaka)

Language Learning

(Michal)

(Silvia)

(Julia)
Summary

✓ The team

➢ Generic Implicit crowdsourcing model

➢ Approach & foreseen outputs

➢ Current results & next steps

➢ Let’s make a quick game!
Implicit crowdsourcing theoretical model

[ General idea ]

Some language learning exercises can be automatically generated from a language-related (e.g. NLP) dataset (e.g. POS lexica, treebanks, wordnets).

**Theoretical model**

**IF**

a dataset can be used to generate a language learning exercise,

**THEN**

the answers to such exercise can be used to enhance the dataset.
Implicit crowdsourcing theoretical model

[ Examples of exercises ]

“Word search” exercises

“Classify words” exercises

⇒ Questions can be generated from a POS lexicon.
⇒ Answers can be used to correct or extend a POS lexicon.
Summary

✓ The team

✓ Generic Implicit crowdsourcing model

➢ Approach & foreseen outputs

➢ Current results & next steps

➢ Let’s make a quick game!
Approach & foreseen outputs

[Approach]

- Teamwork Language Learning ↔ Language-datasets participants.

- From the language learning side.
  ⇒ review of existing language learning exercises and platforms.

- From the language-datasets side.
  ⇒ review of crowdsourcing work relying on non-expert crowd (e.g. GWAP, AMT, Crowdflower etc.).

- Reviews done in short iterative cycles.
  ⇒ Independent review/searching/reading first.
  ⇒ Collaborative discussion and organization of the cycle’s findings.
Approach & foreseen outputs

[ Foreseen outputs ]

Scenarios that combine:

⇒ Language-related dataset with language learning exercises,
⇒ (optional) language platforms and language-datasets providers.

Creation of a prototypical galaxy-like display to allow:

⇒ Learning platforms to identify relevant language-dataset providers,
⇒ Language-dataset providers to identify relevant learning platforms.
Summary

✓ The team

✓ Generic Implicit crowdsourcing model

✓ Approach & foreseen outputs

➢ Current results & next steps

➢ Let’s make a quick game!
Current results & next steps

[ Current results ]

⇒ We have reviewed ~60 papers published at LREC 2017, 2015 & 2013.

⇒ We have identified ~13 types of language datasets that can be paired with language learning exercises.

⇒ We have gathered ~83 pictures/examples illustrating exercises than can be paired with the identified language datasets.

⇒ We have created a prototypical galaxy-like display with “Prezi”

Quite productive 1.5 days despite the unusual and difficult nature of the task!!
Current results & next steps

[ Next steps ]

- Calmly keep on adding some content to reach a critical mass of content.

- Identify and adapt tools to
  1. register papers and authors on the language-datasets side (Zotero?),
  2. catalog existing language learning exercises and platforms (Dspace?),
  3. display the collected content fostering networking (D3.js?).

- Communicate about the initiative on relevant communication channels.

- Hold a WG2 meeting and invite interested members.

- Let people arrange themselves in specialized groups and foster publications, prototyping and project proposals.
Current results & next steps

[ Next steps ]

⇒ The WG2 meeting should ideally be in a place with a high concentration of language learning schools.

⇒ The first experiments should logically target English as it is the language with the largest crowd of learners.

⇒ There is a very adequate place in Europe. Can you guess?

Malta Indeed!
- thanks Claudia! -
Summary

✓ The team

✓ Generic Implicit crowdsourcing model

✓ Approach & foreseen outputs

✓ Current results & next steps

➢ Let’s make a quick game!
Let’s play a quick game

➢ We scheduled an interactive session at the very end of this WG2 meeting.

➢ Members interested in participating or supporting the initiative are invited to join.

➢ We printed some of the papers and some of the pictures illustrating language learning exercises that we collected during Task 4.

➢ The game is to see if you can pair them together the way we did.

➢ Group papers and pictures and “nail” them together on the walls.
Thank you all for your attention

Lionel Nicolas, Verena Lyding et al.

chair.enetcollect@eurac.edu