Work Group 1

enet-COLLECT Bolzano meeting,
September 8, 2017

Presentations

Work group leaders



"Teachers never tell you their first names because they don't want you to *Google* them"

Presentations

Work group leaders



Elena Volodina
University of Gothenburg

WG leader



Iztok Kosem
University of Ljubljana
Institute of Trojina

WG vice-leader

Presentations

Elena Volodina

Researcher

Natural Language Processing



Email
elena.volodina@svenska.gu.se

Telephone
+46(0)31 786 6959

Adress
Språkbanken,
Institutionen för svenska språket, Göteborgs universitet

Box 200, 405 30
Göteborg

Visiting address
Lennart Torstenssonsgatan 8,
Room L411

Researcher at Language Bank (from 2017)
Research engineer at Language Bank (from 2010)
MA in Computational Linguistics from Gothenburg University (2008)
PhD in Linguistics from Moscow State Pedagogical University (1998)

Research Interests

- Intelligent Computer-Assisted Language Learning (ICALL)
- Corpus Linguistics
- Text readability and lexical profiles
- Parallel corpora

Ongoing projects

- SweLL - research infrastructure for Swedish as a Second language (RJ anslag, 2017-2019)
- Analysis av learner language
- Sentence readability - study of sentence understandability etc for L2 learners at different CEFR levels
- Lärka - an exercise generator for Swedish. Start the program here.
- COCTAILL corpus - Corpus Of Coursebook Texts As Input for Learner Levels' modelling - a corpus of course books used for teaching CEFR-based courses of L2 Swedish
- Learner essays, SweLL - a corpus of Swedish L2 essays explicitly linked to CEFR levels

Past projects:

- A System Architecture for Intelligent CALL

Presentations

Elena Volodina

Researcher

Natural Language Processing



Email
elena.volodina@svenska.gu.se

Telephone
+46(0)31 786 6959

Adress
Språkbanken,
Institutionen för svenska språket, Göteborgs universitet

Box 200, 405 30
Göteborg

Visiting address
Lennart Torstenssonsgatan 8,
Room L411

Researcher at Language Bank (from 2017)
Research engineer at Language Bank (from 2010)
MA in Computational Linguistics from Gothenburg University (2008)
PhD in Linguistics from Moscow State Pedagogical University (1998)

Research Interests

RIKSBANKENS JUBILEUMSFOND

SweLL - Research infrastructure for Swedish as a Second Language

(RJ anslag, 2017-2019)

learners at different CEFR levels

program here.

Coursebook Texts As Input for Learner Levels' modelling - a corpus of course

used for teaching CEFR-based courses of L2 Swedish

- Learner essays, SweLL - a corpus of Swedish L2 essays explicitly linked to CEFR levels

Past projects:

- A System Architecture for Intelligent CALL

Presentations

Elena Volodina

Researcher

Natural Language Processing



Email
elena.volodina@svenska.gu.se

Telephone
+46(0)31 786 6959

Researcher at Language Bank (from 2017)
Research engineer at Language Bank (from 2010)
MA in Computational Linguistics from Gothenburg University (2008)
PhD in Linguistics from Moscow State Pedagogical University (1998)

Research Interests

RIKSBANKENS JUBILEUMSFOND

SweLL -

Research infrastructure for Swedish as a Second Language

Exercise Generator | Hit-Ex | Annotation editor | Text evaluation

Lärka

Language Acquisition Reusing Korp

Lärka - "LÄR språket via KorpusAnalys" - with its English equivalent "Lark" (Language Acquisition Reusing Korp) - is a freely available online platform developed in Språkbanken for learning Swedish and Swedish linguistics.

...Coursebook Texts As Input for Learn...

...used for teaching CEFR-based courses of L2 Swedish

- Learner essays, SweLL - a corpus of Swedish L2 essays explicitly linked to CEFR levels

Past projects:

- A System Architecture for Intelligent CALL

Presentations

Elena Volodina

https://spraakbanken.gu.se/larkalabb/siwoco



David Alfter

SiWoCo - Single Word lexical Complexity

Write a lemma

matematik

Select a part-of-speech

noun

Receptive ○ **Productive** ○ **Both** ●

Go!

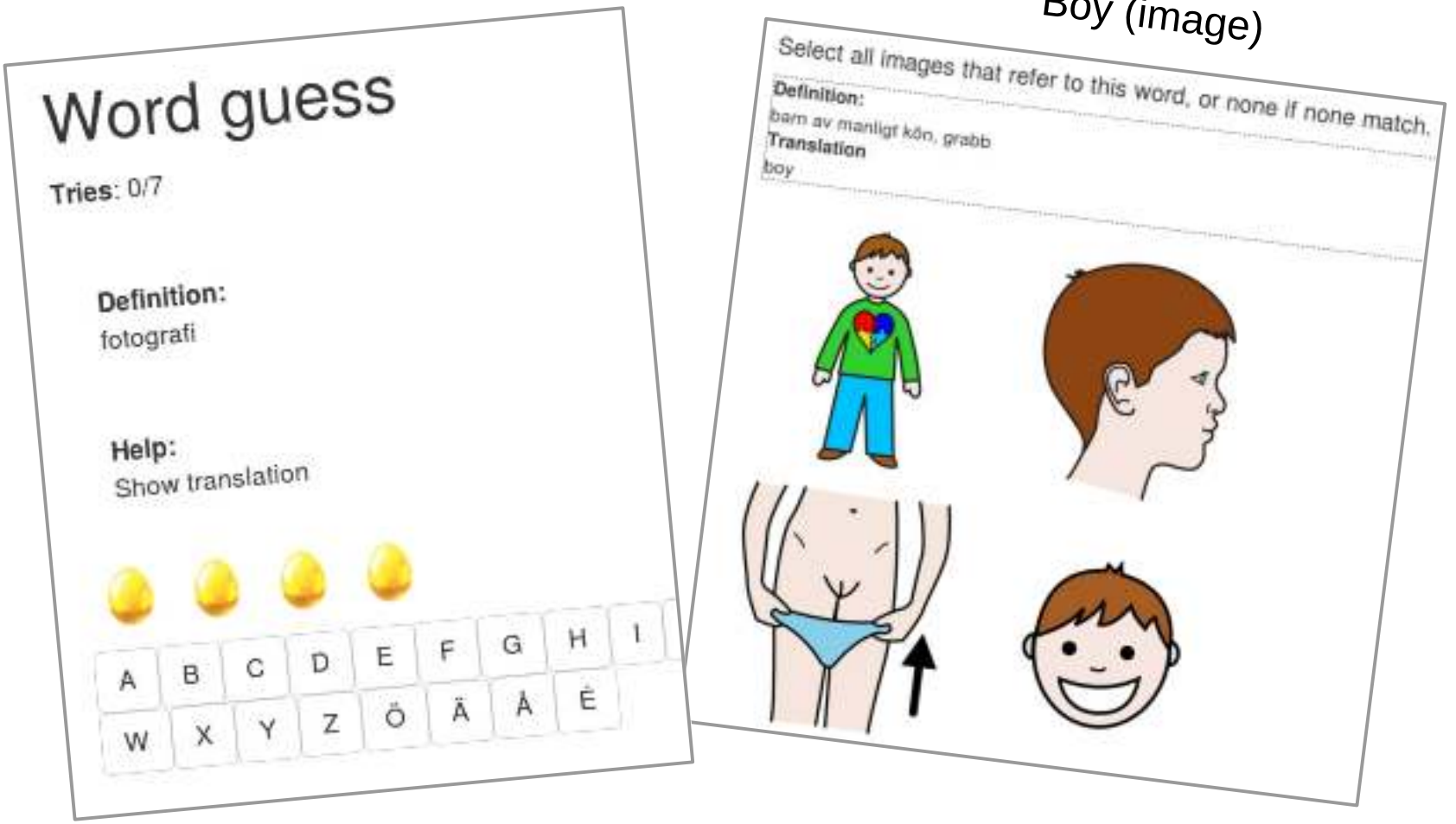
Results

Word	POS	ROP	Predicted level
matematik	NN	both	R: A2 P: B2
hotel	NN	both	R: A2 P: A2

Presentations

Elena Volodina

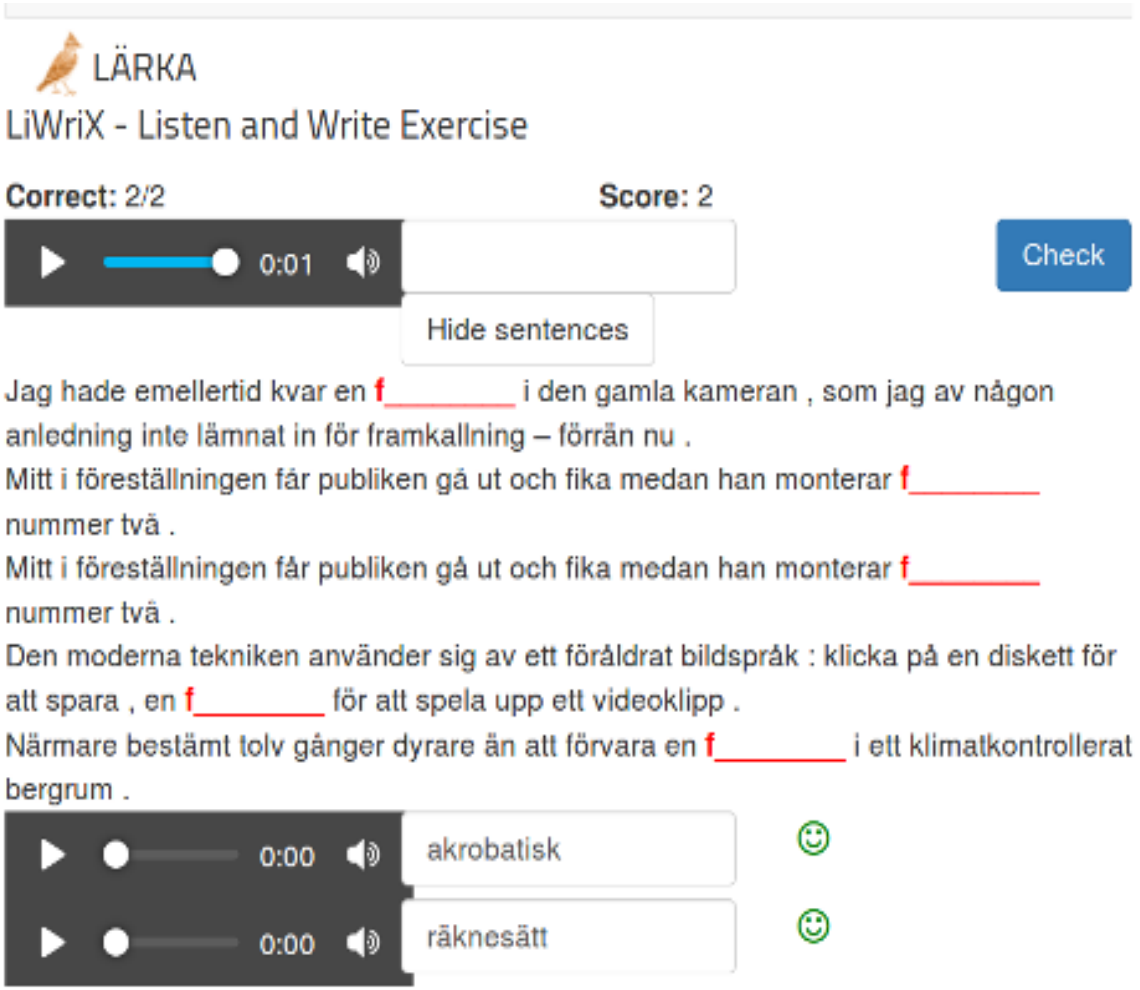
https://spraakbanken.gu.se/larkalabb/wordguess-image



Presentations

Elena Volodina

https://spraakbanken.gu.se/larkalabb/liwrix



Iztok Kosem



RESEARCH INTERESTS

- Lexicography
- Corpus linguistics
- Collocations, dictionary examples
- User needs research

PROJECTS

Large national projects

- Communication in Slovene (assistant coordinator)

European projects - COST:

- ENeL - European Network of e Lexicography (Vice Chair)
- enetCollect - European Network for Combining Language Learning with Crowdsourcing Techniques (Vice Leader of Working Group 1)

ARRS projects

- Collocations as a Resource for Language Description: Semantic and Temporal Aspects (leader)
- Devising a concept of Hungarian-Slovenian dictionary: from language resource to the user (leader)
- New grammar of contemporary standard Slovene: sources and methods

Drugo

- European Survey on Dictionary Use (member of core research team)
- Collocations Dictionary of Slovene
- Dictionary of Modern Slovene
- Enlarging the Šolar corpus - Šolar 2.0 (leader)

CONTACT & LINKS

- E-mail: iztok.kosem@ff.uni-lj.si
- SICRIS: 33796
- Mendeley
- Google Scholar

language technologies for Slovene.

COLLOCATION DICTIONARY OF SLOVENE

Currently in the pre-processing phase, the Collocation Dictionary of Slovene aims to be the first collocation dictionary of Slovene, based on the Collocation Dictionary Database of Slovene, which consists of collocation candidates automatically extracted from the Gigafida reference corpus of standard Slovene.

The dictionary aims to implement crowdsourcing in the data pre-processing phase in order to:

- filter tokenisation errors,
- categorise collocations according to headword sense,
- filter out unsuitable n-grams,
- check syntactic structures.

Several crowdsourcing tasks are currently being tested using a local installation of the PyBossa crowdsourcing platform (http://pybossa.com/).



PyBossa Task: Sorting collocations under headword senses.

About the crowd

WG1 members

enetCollect-WG1 Shared privately

☆

Topics · About

45 enlisted in the mail-group
24 primary WG1 members

Sort by name **Sorted by join date**

Showing 45 of 45 members

vidas.daudaravicius@vtex.lt
joined May 23, 2017
(Member)



Špela Arhar Holdt
joined May 24, 2017
(Member)

nihada delibegovic
joined May 24, 2017
(Member)

Marta Giralt
joined May 26, 2017
(Member)

ramune.kasperaviciene@gmail.com
joined May 26, 2017
(Member)

Karpava Sviatlana
joined May 26, 2017
(Member)

Biljana Lazić
joined May 27, 2017
(Member)

aglaznieks@eurac.edu
joined May 29, 2017
(Member)

magali.paquot@uclouvain.be
joined May 30, 2017
(Member)



Montse Martu
joined May 31, 2017
(Member)

Julie Van de Vyver
joined May 31, 2017
(Member)

Shalev Carmel
joined Jun 1, 2017
(Member)

Fanny Meunier
joined Jun 4, 2017
(Member)

Rina zviel-girshin
joined Jun 6, 2017
(Member)

ann.devitt@tcd.ie
joined Jun 7, 2017
(Member)



How to join?

https://spraakbanken.gu.se/eng/icall/enetcollectWG1



About the crowdsourcing

for language learning

https://www.youtube.com/watch?v=-Ht4qiDRZE8



Luis von Ahn

2011
Massive-scale online collaboration
DuoLingo

0:00-3:25, 8:30-13:00

WG1 Plans, year 1-4

- **Task 1.1:** Overview of existing materials and workflows for their creation Month 1-12
- Reviewing existing resources, discussion groups, and surveys with stakeholders
- D.1.1 - Summary report on existing materials and workflows for their creation

- **Task 1.2:** Review of explicit crowdsourcing approaches Month 1-12
- Literature reviews of explicit crowdsourcing approaches
- D.1.2 - Summary report on explicit crowdsourcing approaches

- **Task 1.3:** Research on producing learning material through explicit crowdsourcing Month 6-24
- Development and adaptation of approaches; description of a theoretical framework;
- D.1.3 - Theoretical framework for producing material through explicit crowdsourcing

- **Task 1.4:** Design, implementation, and testing of prototypes Month 18-45
- Design, implementation, and evaluation of prototypes
- D.1.4 - Design plans, implemented prototypes and related evaluation reports

Year 1.
Meeting 1.

Work group 1 meeting in Bolzano, Italy, September 8, 2017

The first enet-COLLECT meeting, so-called kick-off meeting, is hosted by EURAC during September, 8, 2017. All work groups (1-5) have a slot to present themselves to the action members and discuss the agenda for the coming years.



Program, Work group 1 meeting

Room: To be announced

Time	Program point
	Session 1.
08.30-08.50	**Elena Volodina and Iztok Kosem** *Opening*
08.50-09.20	**Špela Arhar Holdt** *From student writing to language tools and materials: where does crowdsourcing come in?*
09.20-10.00	Interactive round-table discussion. Part 1
10.00-10.30	*COFFEE BREAK*
	Session 2.
10.30-11.10	**Detmar Meurers***FeedBook: A Web-based Workbook for English Supporting the Interaction of Students and Teachers* Abstract
11.10-11.30	Interactive round-table discussion. Part 2
11.30-12.00	**Elena Volodina** *WG1: Planning ahead*

Year 1. Meeting 1.
Highlights

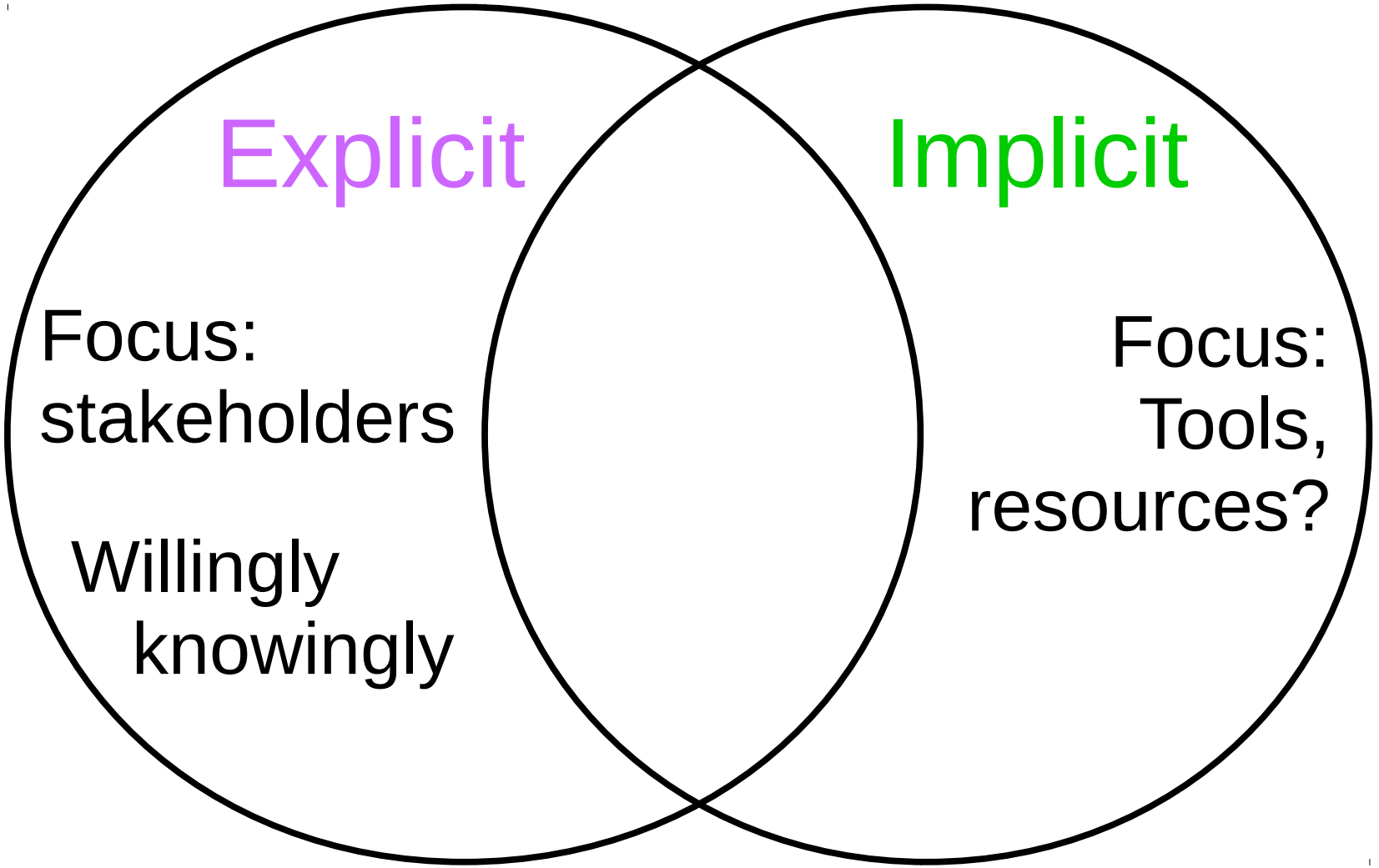
08.50-09.20	**Špela Arhar Holdt** *From student writing to language tools and materials: where does crowdsourcing come in?*
09.20-10.00	Interactive round-table discussion. Part 1

10.30-11.10	**Detmar Meurers***FeedBook: A Web-based Workbook for English Supporting the Interaction of Students and Teachers* Abstract
11.10-11.30	Interactive round-table discussion. Part 2

“Interactive” discussion (the crowdsourcing part)

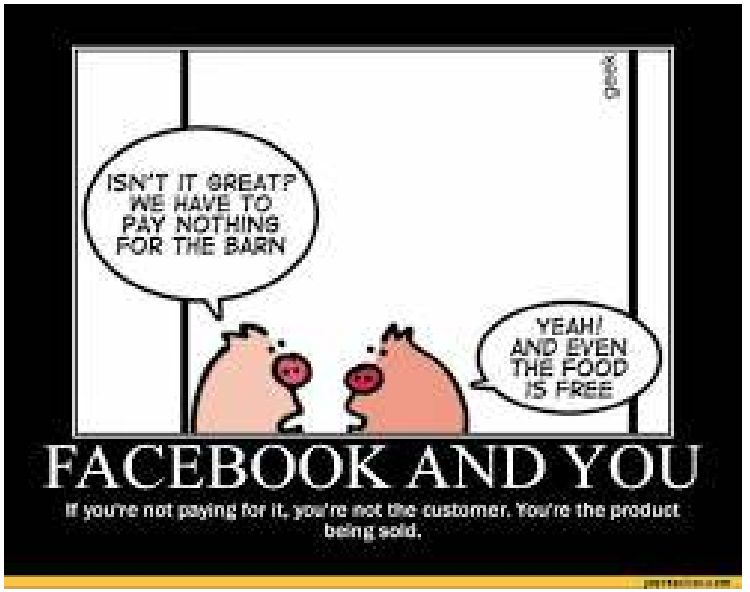
https://www.1ka.si/a/137232

Explicit versus implicit?



About the (crowd)sourcing

Isn't it great? We have to pay nothing for the barn



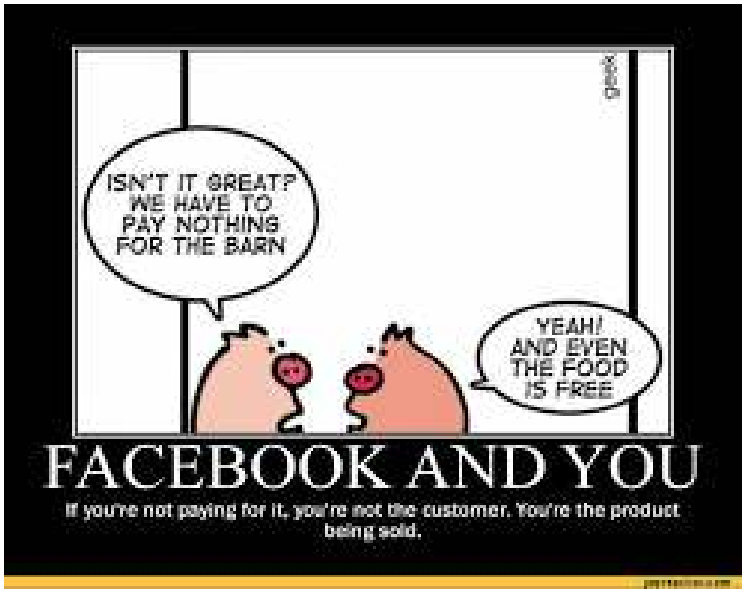
YEAH! And even the food is free

If you are not paying for it, you are not the customer.
You are the product being sold.

i.e. YOU ARE THE FOOD

About the (crowd)sourcing

Stakeholders: Teachers



Stakeholders: Learners

Motivation

If you are not paying for it, you are not the customer.
You are the product being sold.

i.e. YOU ARE THE FOOD

Attitudes

WG1 – a draft sketch of the coming years
Year 1

Meet. 1	Kick-off	Introduction, discussion, planning
Meet. 2 (2018)	Workshop Meeting	→ **Ethical & legal** workflows (+WG5 ?) → **Survey of the crowdsourcing** (STSM report) → Teacher-focused survey (draft)

- **Task 1.1:** Overview of existing materials and workflows for their creation M1-12
- **Task 1.2:** Review of explicit crowdsourcing approaches M1-12

WG1 – a draft sketch of the coming years
Year 2

Meet. 3-4 (2019)	Workshops	→ **Case studies**/workflows (+WG3 ?) → **Motivation & perspectives:** focus on users
	Meetings	→ **Teacher survey results** → Learner-focused survey (draft)

- **Task 1.3:** Research on producing learning material through explicit crowdsourcing Month 6-24
- Development and adaptation of approaches; description of a theoretical framework;
- D.1.3 - Theoretical framework for producing material through explicit crowdsourcing

WG1 – a draft sketch of the coming years
Year 3

Meet 5-6 (2020)	Workshops	→ **Gamification** of crowdsourcing (+ WG2 ?) → Proof-of-concept: **learning materials through crowdsourcing**
	Meetings	→ **Learner survey results** → **Specifications** for a robust crowdsourcing system (user perspective)

- **Task 1.4:** Design, implementation, and testing of prototypes Month 18-45
- Design, implementation, and evaluation of prototypes
- D.1.4 - Design plans, implemented prototypes and related evaluation reports

WG1 – a draft sketch of the coming years
Year 4

Meet 7-8 (2021)	Workshops	→ **Systems in use** (+WG4 ?) → Evaluation
	Meetings	→ Robust systems (+ WG4 ?) → Lessons learned

- **Task 1.4:** Design, implementation, and testing of prototypes Month 18-45
- Design, implementation, and evaluation of prototypes
- D.1.4 - Design plans, implemented prototypes and related evaluation reports

What is done

- **Task 1.1:** Overview of existing materials and workflows for their creation M1-12
- **Task 1.2:** Review of explicit crowdsourcing approaches M1-12

- STSM-1 (June 2017) Iztok + Elena planning
- STSM-2 (fall 2017, Ljubljana) targeting tasks 1.1 & 1.2

→ survey, article, presentation (Spring 2018)

WG1 member involvement

- **Task 1.1:** Overview of existing materials and workflows for their creation
- **Task 1.2:** Review of explicit crowdsourcing approaches

TODOs:

1. A special group/rotating task, 2 WG members: setting up and updating "**Reading list**"
    - ✔ gives right to get reimbursement for travelling to meetings
        - → candidate 1: David Alfter (Sweden)
        - → candidate 2:

WG1 member involvement

- **Task 1.1:** Overview of existing materials and workflows for their creation
- **Task 1.2:** Review of explicit crowdsourcing approaches

TODOs:

2. A special group/rotating task, 2 WG members: setting up and updating "**Relevant events**"
    - ✔ gives right to get reimbursement for travelling to meetings
        - → candidate 1:
        - → candidate 2:

WG1 member involvement

- **Task 1.1:** Overview of existing materials and workflows for their creation
- **Task 1.2:** Review of explicit crowdsourcing approaches

TODOs:

3. A work group 2-3 WG1 members: to prepare a **draft of a survey / interviews with teachers** (to present at Feb-18 meeting)
    - ✔ gives right to get reimbursement for travelling to meetings
        - → candidate 1:
        - → candidate 2:
        - → candidate 3:

WG1 member involvement

- **Task 1.1:** Overview of existing materials and workflows for their creation
- **Task 1.2:** Review of explicit crowdsourcing approaches

TODOs:

4. Take charge of the **organization of the next workshop**
    - ✔ topic “Case studies/workflows on ethical and legal questions within (explicit) crowdsourcing”,
    - ✔ including call for abstracts, selection of presentations, reviewing, planning, ...
    - ✔ gives right to get reimbursement for travelling to meetings
        - → candidate 1 (WG1):
        - → candidate 2 (WG5):
        - → candidate 3 :

Thank you!