Martin Benjamin
Crowdsourcing and Gamification for Multilingual Linguistic Data

EnetCollect WG3/WG5 Meeting
European Network for Combining Language Learning with Crowdsourcing Techniques
Leiden, The Netherlands – 24-25 October 2018
kamusi is Swahili for dictionary
Goal: A complete matrix of human expression across time and space

• As a knowledge resource
• As a data resource
• For myriad languages
In service since 1994 (originally at Yale Council on African Studies)
International NGO since 2009
• Registered non-profit in USA and Switzerland

Academic Home since 2013:
EPFL - Swiss Federal Institute of Technology in Lausanne
LSIR - Distributed Systems Information Laboratory
why multilingual dictionaries were impossible
why multilingual dictionaries were impossible
PWN (English Wordnet):
light x 47

WOLF (French Wordnet):
light = lumière x 44
light = léger x 37
léger = lumière x 36

why multilingual dictionaries were impossible
why multilingual dictionaries were impossible
why multilingual dictionaries were impossible
why multilingual dictionaries were impossible
### EUdict :: English-Italian dictionary

<table>
<thead>
<tr>
<th>ENGLISH</th>
<th>ITALIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>light</td>
<td>accendere; illuminare; far luce a, illuminare la strada a, agg. luminoso, illuminato; chiaro, lampada, poco pesante, rado, schiarire, acce...</td>
</tr>
<tr>
<td>light</td>
<td>leggero</td>
</tr>
<tr>
<td>light</td>
<td>lieve</td>
</tr>
<tr>
<td>light</td>
<td>luce</td>
</tr>
<tr>
<td>light</td>
<td>lume</td>
</tr>
<tr>
<td>light</td>
<td>rischiarare</td>
</tr>
<tr>
<td>flood light</td>
<td>faro, illuminazione</td>
</tr>
<tr>
<td>let there be light!</td>
<td>sia la luce</td>
</tr>
<tr>
<td>light emitting diode</td>
<td>diodo emanatore di luce (anello elettronico che qualora attraversato da corrente elettrica emana luce a frequenza regolare)</td>
</tr>
<tr>
<td>light hearted</td>
<td>sereno</td>
</tr>
<tr>
<td>light meter</td>
<td>esposimetro, esposimetro</td>
</tr>
<tr>
<td>light up</td>
<td>illuminare</td>
</tr>
<tr>
<td>light-hearted</td>
<td>giallo</td>
</tr>
<tr>
<td>to light</td>
<td>accendersi</td>
</tr>
<tr>
<td>traffic light</td>
<td>semaforo</td>
</tr>
<tr>
<td>traffic light</td>
<td>semaforo (II)</td>
</tr>
</tbody>
</table>

Search time: 0.001 seconds.
why multilingual dictionaries were impossible
why multilingual dictionaries were impossible
why multilingual dictionaries were impossible
why multilingual dictionaries were impossible
how Kamusi makes a multilingual dictionary possible
light (not dark)
light (not heavy)
light (not serious)
light (not fattening)

how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
light (not dark)

light (not heavy)

light (not serious)
fr: léger  sw: -a kuchekesha  fi: tyhjänpäiväinen  th: ซึ่งไร้สาระ

light (not fattening)

how Kamusi makes a multilingual dictionary possible
light (not dark)

light (not heavy)

light (not serious)

light (not fattening)  
fr: allégé  
sw: pungufu  
fi: kaloriton  
th: ที่แคลอรี่ต่ำ

how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
light (not dark)
light (not heavy)
light (not serious)
light (not fattening)

how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
light (not dark)

light (not heavy)

light (not serious)

light (not fattening)

how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
how Kamusi makes a multilingual dictionary possible
Crowdsourcing for Language Data

- (Games for Grandparents)
- Facebook Games
- WordRace
- Word Up
- DUCKS – 10,000 pending datasets from PanLex
- GOLDbox
- Green-lighting
- Visual Dictionary (tagging and vocabulary)
Play in Any Language
(of 7000 with ISO-639-3 codes)
Games for Grandparents

Older people often have:

• Rich language knowledge
• Passion
• Time
• Lower technology skills
• Poor vision

*Yiddish* will be our proving ground
Word Race

• Worldwide competition to provide parallel terms for many languages

• Selected words are sent to mobile devices

• Competition within and between languages to be fastest with a winning answer
Word Race

• Item considered valid when it reaches a consensus threshold

• Runners-up will be tested in a different game:
  does [term] mean:
  [Definition]
  Yes/ No/ Not sure
Word Race

• Users control their frequency of play
• Terms are sent on schedule or on the editor’s whim
Word Race

• In principle, any language. In practice, we’ll see where we can get a critical mass
• Threshold can be set to 1
• Difficulty is based on how many other languages have completed a term
Word Race

• Scores based on fastest time to submit an answer that achieves consensus (not first over the finish line)
Word Up

• Provide terms based on user interests
• No time pressure – but will be merged with Word Race
Word Up

• “Packs” for thematically related terms
• Will be customized for terminology domains (users can register for their expertise or passion)
Word Up

• Challenge to keep from burdening too many players with the same data – finished entries for a language surreptitiously removed when user updates

• Offline users may duplicate efforts
Word Up
Word Up

• Unresolved technical issue:

Mobile phones force uppercase
Facebook Games

- Fun ways to elicit microdata
- Familiar environment for users
- Networks of friends speaking the same language
- Achievements reported on wall
- Free access in some non-wealthy countries
Facebook Translation Game
Facebook Translation Game
Definition Game

Definition Game 🎁
Write or vote for a definition in English

go noun
Working definition: a usually brief attempt

% I can write the winning definition for this idea!

? I can’t say - skip this one...

► Keep the working definition. It’s spectacular as it is!

Wiktionary • Dictionary.com • Wordnik
Definition Game

Definition Game
Write or vote for a definition in English

go noun
Working definition: a usually brief attempt

I can write the winning definition for this idea!

? I can't say - skip this one...
Keep the working definition. It's spectacular as it is!

Wiktionary • Dictionary.com • Wordnik
Definition Game

1. Write or vote for a definition in English.
2. Worked definition: a usually brief attempt.

\(\text{go noun}\)

- I can write the winning definition for this idea!
- An attempt to achieve something with a recognized position.
- I can't say - skip this one...
- Keep the working definition. It's spectacular as it is!

[Wiki], [Dictionary.com], [Wordnik]
Example Game
Example Game
Twitter Examples:

- Awful for polysemous single words
- Fantastic for Party Terms (MWEs)
DUCKS (Data Unified Concept Knowledge Sets)
DUCKS (Data Unified Concept Knowledge Sets)
DUCKS (Data Unified Concept Knowledge Sets)
GOLDbox – submissions directly from dictionary
Green-lighting to verify WordNet connections

<table>
<thead>
<tr>
<th>Language</th>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>suss out (v)</td>
<td><strong>Definition:</strong> examine so as to determine accuracy, quality, or condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Example:</strong> check the brakes. Check out the engine.</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Similar:</strong> look into, go over, check up on, check, check over, check out, check into</td>
</tr>
<tr>
<td>Romanian</td>
<td>verifică (v)</td>
<td><strong>Definition:</strong> A supune controlului pentru a determina calitatea, acuratețea etc.</td>
</tr>
<tr>
<td></td>
<td>inspecta (v)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>controla (v)</td>
<td></td>
</tr>
</tbody>
</table>
Visual Dictionary
Kamusi Here! [http://kamusi.org/info/here](http://kamusi.org/info/here)

- Mobile & Web App
- Many languages linked with precision
- Much greater accuracy than Google Translate
- Forthcoming: GOLDbox (crowd submissions right from dictionary)
- Forthcoming: passive crowd ranking (arrange by user focus)