Play the Language: an alternative way of annotation

Barbora Hladká, Jiří Mírovský

in cooperation with
Jirka Hana, Jan Kohout, Vladimír Rovenský
Human-machine communication

- applications comfortable and friendly to users
- communication in the user’s native language
- knowledge of the language
  - From where? → From (annotated) corpora.
Annotated corpora

- source of information on natural language structure (e.g. Prague dependency treebanks)
  - theoretical linguistics
  - natural language processing
    - supervised learning methods
  - ...
    - the more data, the better
Data, more data, even more data

- 1969
  - Brown corpus of 1M tokens
- 1994
  - British National Corpus of 100M tokens
- 2006
  - Web 1T 5gram of 1,024,908,267,229 tokens
An alternative way of annotation
Human computation

- The idea: there is a lot of things
  - that humans have no problem with, but
  - computers cannot yet solve.
  - I.e. tasks trivial for humans, but very challenging for computers.

- How to address these things? Human computation.
  - Use massive groups of networked human minds to solve the problem that computers cannot.

- Hot issue over the last decade.
In other word, crowd-sourcing
Human computation - examples

- Voluntary
  - Wikipedia, Open source software
- Motivation by money
  - Amazon Mechanical Turk
- Motivation by fun
  - Games with a purpose
- No other choice
  - CAPTCHA/reCAPTCHA
Crowdsourcing

Crowdsourcing is the act of outsourcing tasks, traditionally performed by an employee or contractor, to an undefined, large group of people or community (a crowd), through an open call.

Jeff Howe, one of first authors to employ the term, established that the concept of crowdsourcing depends essentially on the fact that because it is an open call to an undefined group of people, it gathers those who are most fit to perform tasks, complex problems and contribute with the most relevant and fresh ideas to benefits from their inputs.

For example, the public may be invited to develop a new technology, carry out a design task (also known as community-based design[1] and distributed participatory design), refine or carry out the steps of an algorithm (see human-based computation), or help capture, systematize or analyze large amounts of data (see also citizen science).

The term has become popular with businesses, authors, and journalists as shorthand for the trend of leveraging the mass collaboration enabled by Web 2.0 technologies to achieve business goals. However, both the term and its underlying business models have attracted controversy and criticisms.
Amazon Mechanical Turk

### All HITs

#### All HITs

<1-10 of 2080 Results

Sort by: HITs Available (most first)

<table>
<thead>
<tr>
<th>Requester</th>
<th>HIT Expiration Date</th>
<th>Time Allotted</th>
<th>Reward</th>
<th>HITs Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chris Callison-Burch</td>
<td>Dec 31, 2010 (2 weeks 3 days)</td>
<td>15 minutes</td>
<td>$0.05</td>
<td>13917</td>
</tr>
<tr>
<td>James Gunter</td>
<td>Dec 27, 2010 (1 week 6 days)</td>
<td>60 minutes</td>
<td>$0.04</td>
<td>5489</td>
</tr>
<tr>
<td>Dolores Labs</td>
<td>Dec 21, 2010 (6 days 23 hours)</td>
<td>60 minutes</td>
<td>$0.08</td>
<td>4638</td>
</tr>
<tr>
<td>Jaime arguello</td>
<td>Dec 23, 2010 (1 week 2 days)</td>
<td>5 minutes</td>
<td>$0.04</td>
<td>2736</td>
</tr>
</tbody>
</table>

Sign In
A **CAPTCHA** or Captcha (pronounced /ˈkæptʃə/) is a type of challenge-response test used in computing to ensure that the response is not generated by a computer. The process usually involves one computer (a server) asking a user to complete a simple test which the computer is able to generate and grade. Because other computers are unable to solve the CAPTCHA, any user entering a correct solution is presumed to be human. Thus, it is sometimes described as a reverse Turing test, because it is administered by a machine and targeted at a human, in contrast to the standard Turing test that is typically administered by a human and targeted at a machine. A common type of CAPTCHA requires the user to type letters or digits from a distorted image that appears on the screen.

The term "CAPTCHA" (based upon the word capture) was coined in 2000 by Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford (all of Carnegie Mellon University). It is a contrived acronym for "Completely Automated Public Turing test to tell Computers and Humans Apart." Carnegie Mellon University attempted to trademark the term [2] but the trademark application was abandoned on 31 April.
Games with a Purpose (1)

- By Lenore Blum and Luis von Ahn.
- The users enjoy playing the games.
- The players perform basic tasks that cannot be automated.
- The players work simply by playing the game.
- The data are generated as a by-product of the game.
- If the game is enjoyable, it brings human resources and saves financial resources.
- The game popularity brings more game sessions and thus more annotated data.
Games with a Purpose (2)

Key issues:
- mechanisms that make the game enjoyable,
- mechanisms that guarantee the game output is correct, i.e. quantity vs. quality
  - correctness
  - fidelity
  - usability
  - applicability

If we get the players’ agreement on the output then we can expect more reliable results.
Remarkable gwap.com statistics

“A total of 13,630 people played the ESP Game during this time [August 9 - December, 2003], generating 1,271,451 labels for 293,760 different images. Over 80% of the people played on more than one occasion. Furthermore, 33 people played more than 1,000 games (this is over 50 hours playing the game). “

Luis von Ahn, 2004
Let’s turn text annotation into games
### Annotation vs. Game

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Goal, Team</th>
<th>Game</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguists</td>
<td>Input Knowledge</td>
<td>Crowd of Internet users, motivation needed</td>
</tr>
<tr>
<td>Theoretical Linguistic Framework</td>
<td>Strategy</td>
<td>No linguistic knowledge</td>
</tr>
<tr>
<td>E.g. Single, Parallel</td>
<td>Instructions</td>
<td>[Single</td>
</tr>
<tr>
<td>Guidelines, Annotation Schemes</td>
<td>Environment</td>
<td>What can (not) the players do</td>
</tr>
<tr>
<td>Editor</td>
<td>Evaluation</td>
<td>Game Implementation</td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
<td>Evaluation</td>
</tr>
</tbody>
</table>

Cheaper and Faster
... In the year 1878 I took my degree of Doctor of Medicine of the University of London, and proceeded to Netley to go through the course prescribed for surgeons in the army. Having completed my studies there, I was duly attached to the Fifth Northumberland Fusiliers as Assistant Surgeon. The regiment was stationed in India at the time, and before I could join it, the second Afghan war had broken out. On landing at Bombay, I learned that my corps had advanced through the passes, and was already deep in the enemy's country. I followed, however, with many other officers who were in the same situation as myself, and succeeded in reaching Candahar in safety, where I found my regiment, and at once entered upon my new duties...
Text GWAPs so far

- **PhraseDetectives** – anaphora resolution
- **Jinx** – word sense disambiguation
- **Verbosity** – collecting common-sense facts
- **OntoGames** – semantic content annotation
Acoustic Kitty (Wikipedia)

Acoustic Kitty was a CIA project launched by the Directorate of Science & Technology in the 1960s attempting to use cats in spy missions.
**Acoustic Kitty (Wikipedia)**

Acoustic Kitty was a CIA project launched by the Directorate of Science & Technology in the 1960s attempting to use cats in spy missions. A battery and a microphone were implanted into a cat and an antenna into its tail.

The phrase in blue is the closest phrase that refers to the phrase in orange.
Our contribution

- **Play the Language**: LGame portal
  - text GWAPs
    - coreference, language modelling

- **Elysium**
  - an annotation editor for school kids
    - morphology, syntax
  - educational aspect for children’s learning activities
LGame - basic features

1. Playing the games requires no extra knowledge of the language in question.
2. The game rules are language independent.
3. Parallel data uploaded into the sessions.
Registered: 296
Logged-in: tester2

Shannon game - a game with words in the sentence?
Place the Space - a game with words in the sentence?
PlayCoref - a game with words in texts?

A competition before Christmas 2010!
Shannon Game

- A game of intentionally hidden words in the sentence; players guess them.

- For two players only; could be for one, too.
Shannon Game

Time 00:01  Guesses 3  Opponent ○○

I felt rather indignant at having two characters I had admired treated in this cavalier .
I felt rather indignant at having two characters whom I had admired treated in this cavalier way.
I felt rather indignant at having two characters whom I had admired treated in this cavalier manner.
Shannon Game

Results

Player  ●● -20 pts
Opponent ●● 30 pts

You lost this game.

PLAY AGAIN
Shannon Game – Scoring Func.

For each guessed word:

- 40 pts – if 1\textsuperscript{st} guess correct
- 20 pts – if 2\textsuperscript{nd} guess correct
- 10 pts – if 3\textsuperscript{rd} guess correct
- -10 pts – if not guessed correctly

For 2 words: -20 … 80 pts
Shannon Game – Data Prep.

Arthur Conan Doyle: A Study in Scarlet

- free to use
- many translations
- free audio book

Only sentences without punctuation and of certain length selected.

Sentence segmentation, tokenization, (proper name recognition).
Place the Space

- A game of intentionally hidden spaces in the sentence; players restore them.

- For a single player.
I have a mule and two horses waiting in the Eagle Ravine.
Place the Space

I have a mule and two horses waiting in the Eagle Ravine.
Place the Space

Results

Correct solution:
I have a mule and two horses waiting in the Eagle Ravine.

Player's solution:
I have amule and two horses waiting in the Eagle Ravine.

Score: 95

Play again
Place the Space – Scoring Func.

\[ \text{Score}(\text{Player}) = 100 \times F_1(\text{Player, Correct}) \]

Positions in text are given by the sequence of characters without spaces.

\[ F_1 = 2 \times P \times R / (P + R) \]

*Precision* = \# correctly guessed / \# guesses

*Recall* = \# correctly guessed / \# correct
Place the Space – Data Prep.

Arthur Conan Doyle: A Study in Scarlet

- only sentences without punctuation and of certain length selected
- sentence segmentation, tokenization
PlayCoref

- A game of marking coreferential pairs in the document.

- For a single player or or two players.
PlayCoref

Select number of players, please ...

Players waiting: 0
On the very day that I had come to this conclusion, I was standing at the Criterion Bar, when some one tapped me on the shoulder, and turning round I recognized young Stamford, who had been a dresser under me at Barts. The sight of a friendly face in the great wilderness of London is a pleasant thing indeed to a lonely man.
On the very day that I had come to this conclusion, I was standing at the Criterion Bar, when someone tapped me on the shoulder, and turning round I recognized young Stamford, who had been a dresser under me at Barts. The sight of a friendly face in the great wilderness of London is a pleasant thing indeed to a lonely man.
On the very day that I had come to this conclusion, I was standing at the Criterion Bar, when some one tapped me on the shoulder, and turning round I recognized young Stamford, who had been a dresser under me at Barts. The sight of a friendly face in the great wilderness of London is a pleasant thing indeed to a lonely man in old days. Stamford had never been a particular crony of mine, but now, I hailed him with enthusiasm, and he in his turn appeared to be delighted to see me.
PlayCoref

Match results

Player
Pairs created: 13  Score: 30
PlayCoref – Scoring Function

\[ \text{Score}(\text{Player A}) = \lambda_1 \times F_1(\text{Player A, Player B}) + \lambda_2 \times F_1(\text{Player A, ACR or Manual}) + \lambda_3 \times \min(12, \text{sentences})/12 \]

(and multiplied by 100; \( \sum(\lambda_i) = 1, \lambda_i > 0 \))
PlayCoref – Scoring Function

OUR morning's exertions had been too much for my weak health, and I was tired out in the afternoon. After Holmes' departure for the concert, I lay down upon the sofa and endeavoured to get a couple of hours' sleep. It was a useless attempt. My mind had been too much excited by all that had occurred, and the strangest fancies and surmises crowded into it.
PlayCoref – Scoring Function

OUR morning’s exertions had been too much for my weak health, and I was tired out in the afternoon. After Holmes’ departure for the concert, I lay down upon the sofa and endeavoured to get a couple of hours’ sleep. It was a useless attempt. My mind had been too much excited by all that had occurred, and the strangest fancies and surmises crowded into it.
PlayCoref – Data Preparation

Arthur Conan Doyle: A Study in Scarlet

- analysis up to the tectogrammatical layer using TectoMT modules; it includes simple named entity recognition

- manual annotation of coreference (automatic procedures – too low accuracy)
Elysium

- to involve school kids in annotation
- offer them a language independent annotation editor
  - grammar practicing
  - not as enthusiastically as playing games
  - they do not have other choice
Elysium as a part of the STYX system (1)
V loňském prosinci přijel k ostravské firmě Daklar
Rozdíl do regulované ceny byl hrazen z dotací.
Elysium as a part of the STYX system (2)

- Prague Dependency Treebank
- More annotated data
- Transformation
- Sentence annotation by school kids
- Elysium: practicing
- Sentences by school kids
- Charon
- Exercises
- Styx: practicing with key answers
- Exercise book
Elysium (1)
### Elysium - Tools->Options->Tagsets

#### Tagsets Tab

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>ufal.mff.cuni.cz...</td>
<td>descr - todo</td>
<td>positional</td>
<td>cze</td>
</tr>
<tr>
<td><a href="http://www.cis.upenn">www.cis.upenn</a>...</td>
<td>descr - todo</td>
<td>atomic</td>
<td>eng</td>
</tr>
<tr>
<td>test-morph</td>
<td>descr - todo</td>
<td>atomic</td>
<td>??</td>
</tr>
</tbody>
</table>

### Syntactic Tagsets

- **CC**
- **CD**
- **DT**
- **EX**
- **FW**
- **IN**
- **JJ**
- **JJR**
- **JJZ**
- **LS**
- **MD**
- **NN**
- **NNS**
- **NNP**
- **NNPS**
- **PDT**
- **POS**
- **PRP**
- **PP**
- **RB**
- **RBR**
- **RBS**
- **RP**
Kids are looking forward to a visit to Prague in December.
Elysium - analyses (4)
A competition before Christmas 2010

December 6-16, 2010
“Although interesting as a topic, I would question the originality of the proposition and the possibility of the team to complete the research within the project.”

An anonymous reviewer, summer 2008
Text GWAPs

- open-ended stories
- hard to predict success or failure in advance
- simply, they have to be tried ...
- Are we able to attract the Internet users to play our games? ... We are looking forward to your feedback 😊
Links & References

- British National Corpus
  - http://www.natcorp.ox.ac.uk/
- Brown Corpus
- Elysium
- GWAP
  - http://gwap.com
- 100K ESP game sample
  - http://server251.theory.cs.cmu.edu/ESPGame100k.tar.gz
- Jinx
- LGame
  - http://www.lgame.cz
- OntoGames
  - http://ontogame.sti2.at/
- PageHunt
  - Page
- Phrase Detectives
  - http://anawiki.essex.ac.uk/phrasedetectives/
- STYX system
  - http://ufal.mff.cuni.cz/styx
- Verbosity game
  - http://gwap.com
- Web 1T 5-gram
  - http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13