Crowdsourcing for NLP

Chris Callison-Burch, Ellie Pavlick and Lyle Ungar
Computer and Information Science
University of Pennsylvania

What is crowdsourcing?

- **How many of you have done crowdsourcing?**
  - Mechanical Turk
  - Other?

Many uses of crowdsourcing

- **Outsourcing**
  - Research, design, writing, coding, …
    - Kaggle, Innocentive, Guru, oDesk, eLance
    - Wikepedia, Foldit …
- **Find crowd wisdom**
  - Prediction markets, product ratings, news selection
- **Social science research**
  - **Online experiments:** e.g. sensitivity to pay
  - **Surveys**: personality, politics, preferences,
- **Obtain labels**
  - Images, speech, videos …
  - Text

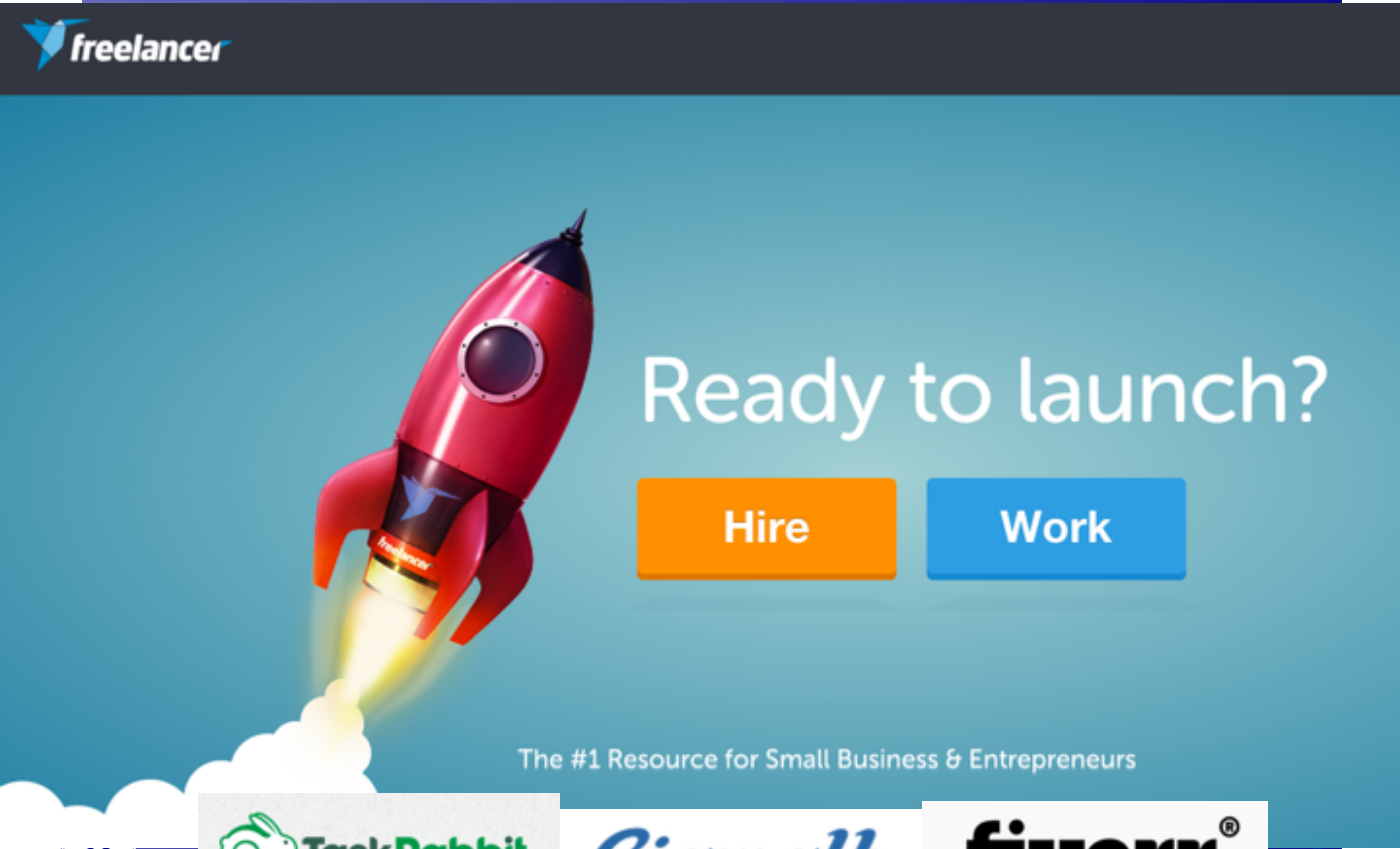
Today: NLP microtasks

Elance work differently

guru

zombal Scientific Outsourcing

99designs



TaskRabbit

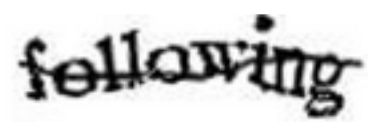
Gigwalk

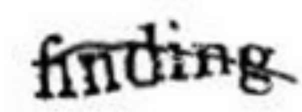
fiverr®


Unbabel
HOW IT WORKS
PRICING & LANGUAGES
FEATURES
LOG IN
TRANSLATION AS A SERVICE
Human corrected machine translation service that enables businesses to communicate globally.
TRY IT FOR FREE
SIGN UP
*no credit card needed
6,951 editors
2,849,184 words translated
6 languages

Stop spam, read books

- **Digitizing NYTimes archives, books**
  - The parts where OCR fails.
- **750,000,000 “volunteers”**
  - 100 million CAPTCHAs per day





Microtasks: Mechanical Turk

Task type	Estimated Proportion
Web scouring	42%
Images related	22%
Text related & OCR	17%
Audio related	14%
Video related	3%
Testing / Quality Assurance	3%

Ipeirotis

Webscouring

- Verify a restaurant listing
- Match my products to Amazon products
- Find official websites for places, companies, …
- Find the email addresses for wedding venues
- Find the school website and its school supply list
- Find Yelp reviews for businesses
- Categorize a Twitter search query
- Find a company name from an email domain
- Find main title, subtitle, and authors for a book
- Categorize web pages

First big NLP use: 2008

Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks

Rion Snow[†] Brendan O'Connor[‡] Daniel Jurafsky[§] Andrew Y. Ng[†]

[†]Computer Science Dept.
Stanford University
Stanford, CA 94305
{rion,ang}@cs.stanford.edu

[‡]Dolores Labs, Inc.
832 Capp St.
San Francisco, CA 94110
brendano@doloreslabs.com

[§]Linguistics Dept.
Stanford University
Stanford, CA 94305
jurafsky@stanford.edu

Abstract

Human linguistic annotation is crucial for many natural language processing tasks but can be expensive and time-consuming. We explore the use of Amazon's Mechanical Turk system, a significantly cheaper and faster method for collecting annotations from a broad base of paid non-expert contributors over the Web. We investigate five tasks: affect recognition, word similarity, recognizing textual entailment, event temporal ordering

and financial cost. Since the performance of many natural language processing tasks is limited by the amount and quality of data available to them (Banko and Brill, 2001), one promising alternative for some tasks is the collection of non-expert annotations.

In this work we explore the use of Amazon Mechanical Turk[1] (AMT) to determine whether non-expert labelers can provide reliable natural language

EMNLP

Snow, O'Connor, Jurafsky & Ng

- **Affect Recognition**

fear(“Tropical storm threatens NYC”) >
fear(“Awesome goal for Beckham”)

- **Word Similarity**

sim(man, boy) > sim(man, rooster)

- **Textual Entailment**

if “Microsoft was established in Italy in 1985” then
“Microsoft was established in 1985”?

- **Word Sense**

“the West Bank” v. “the Bank of America”

- **Temporal Annotation**

denoted happens before collapsed in:
“The condemned building collapsed when the

Image annotations: 2010

Collecting Image Annotations Using Amazon's Mechanical Turk

Cyrus Rashtchian Peter Young Micah Hodosh Julia Hockenmaier
Department of Computer Science
University of Illinois at Urbana-Champaign
201 North Goodwin Ave, Urbana, IL 61801-2302
{crashtc2, pyoung2, mhodosh2, juliahmr}@illinois.edu

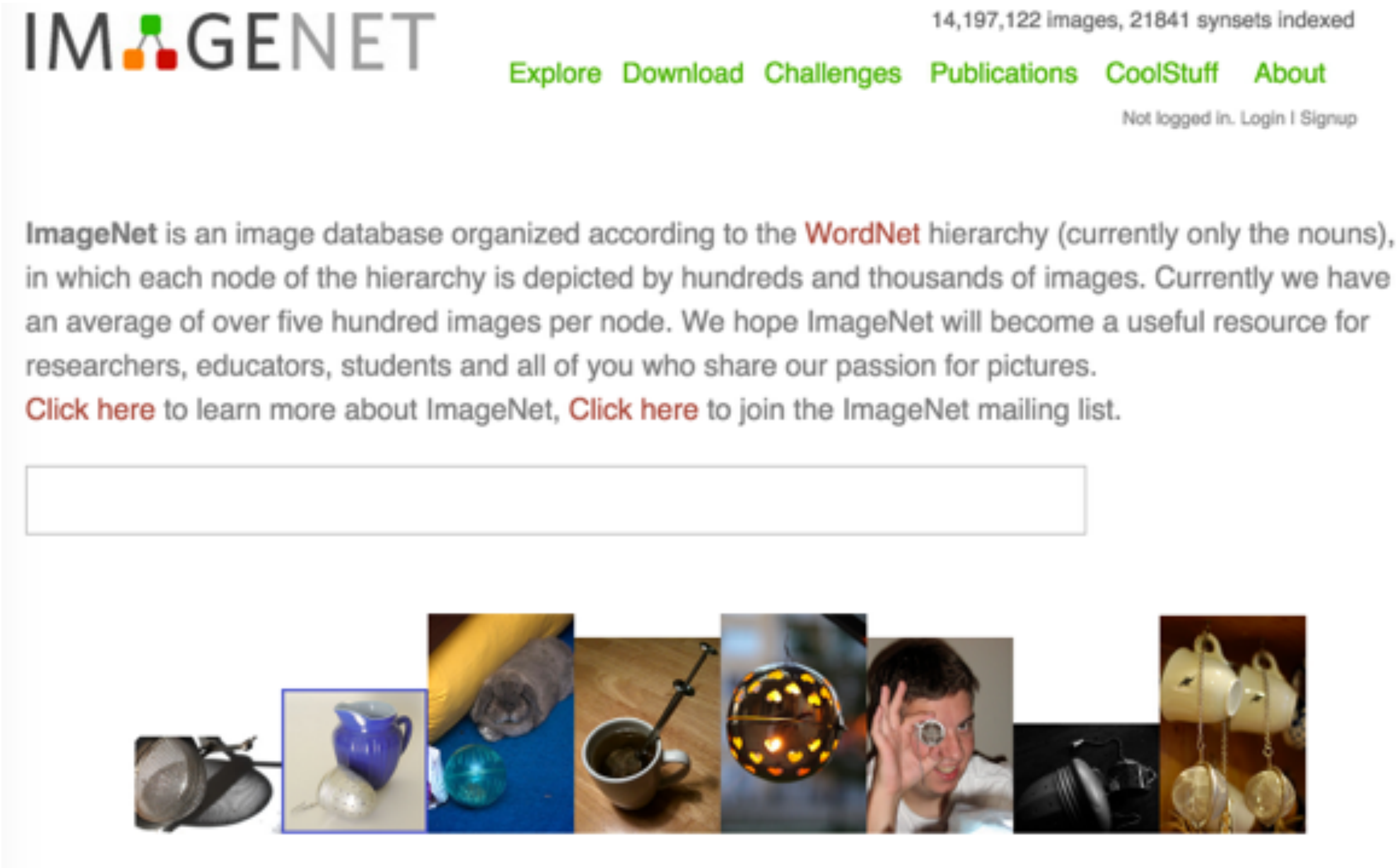
Abstract

Crowd-sourcing approaches such as Amazon's Mechanical Turk (MTurk) make it possible to annotate or collect large amounts of linguistic data at a relatively low cost and high speed. However, MTurk offers only limited

can predict not just the presence and location of certain objects in an image, but also the relations between objects, their attributes, or the actions and events they participate in. Such information can neither be obtained from standard computer vision data sets such as the COREL collection nor from

NAACL HLT

ImageNet



an image database organized according to the WordNet hierarchy

NLP uses of M-turk

- Translation
- Summarization
- Information Extraction
- Document relevance
- Word-sense disambiguation
- Figure captions
- Labeling sentiment, intent, style…
- Getting user information and associated text

Anything in NLP
- generating data or labels
- assessing quality

Schedule

- Taxonomy of crowdsourcing and human computation
- The Mechanical Turk crowdsourcing platform
  - How to set up and run an experiment
- **Break**
- Quality control (and Statistical analysis?)
- Limits of Mechanical Turking (and Ethics?)
- Case Studies in NLP
  - Machine translation
  - Information extraction
  - Word sense disambiguation
  - Computational social science
- Wrap-up