Annotation collaborative de corpus : légendes et réalités du *crowdsourcing*

Karën Fort

karen.fort@sorbonne-universite.fr / https://www.schplaf.org/kf/

27 novembre 2020

LETTRES SORBONNE UNIVERSITÉ

Qu'est-ce que le *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. [Wikipedia, 2 déc. 2010]*

Définition (K. Fort, 2015)

activité qui consiste à faire produire (des idées, des annotations, un dessin, un vote,...) à une masse de gens, aujourd'hui principalement via Internet.

Un mot-valise très expressif...

*The term "crowdsourcing" is a portmanteau of "***crowd***" and "***outsourcing***", first coined by Jeff Howe in a June 2006 Wired magazine article "The Rise of Crowdsourcing". Howe explains that because technological advances have allowed for cheap consumer electronics, the gap between professionals and amateurs has been diminished. Companies are then able to take advantage of the talent of the public, and Howe states that "It's not outsourcing; it's crowdsourcing."*

[Wikipedia, 2 déc. 2010]

. . . qui se délave à la traduction

- Grand Dictionnaire terminologique du Québec : externalisation ouverte
- Journal Officiel : production participative
- G. Adda dans [Sagot et al., 2011] : myriadisation

Différent types de productions

Toutes sortes, dont :

- crowdvoting : votes sur une question, un produit, etc
- crowdcreation : compétition d'idées
- crowdwisdom : réponses à des questions (`Yahoo! questions`)
- crowdfunding : récolte de fonds pour un projet artistique, une campagne politique, etc
- crowddata (?) : production de données

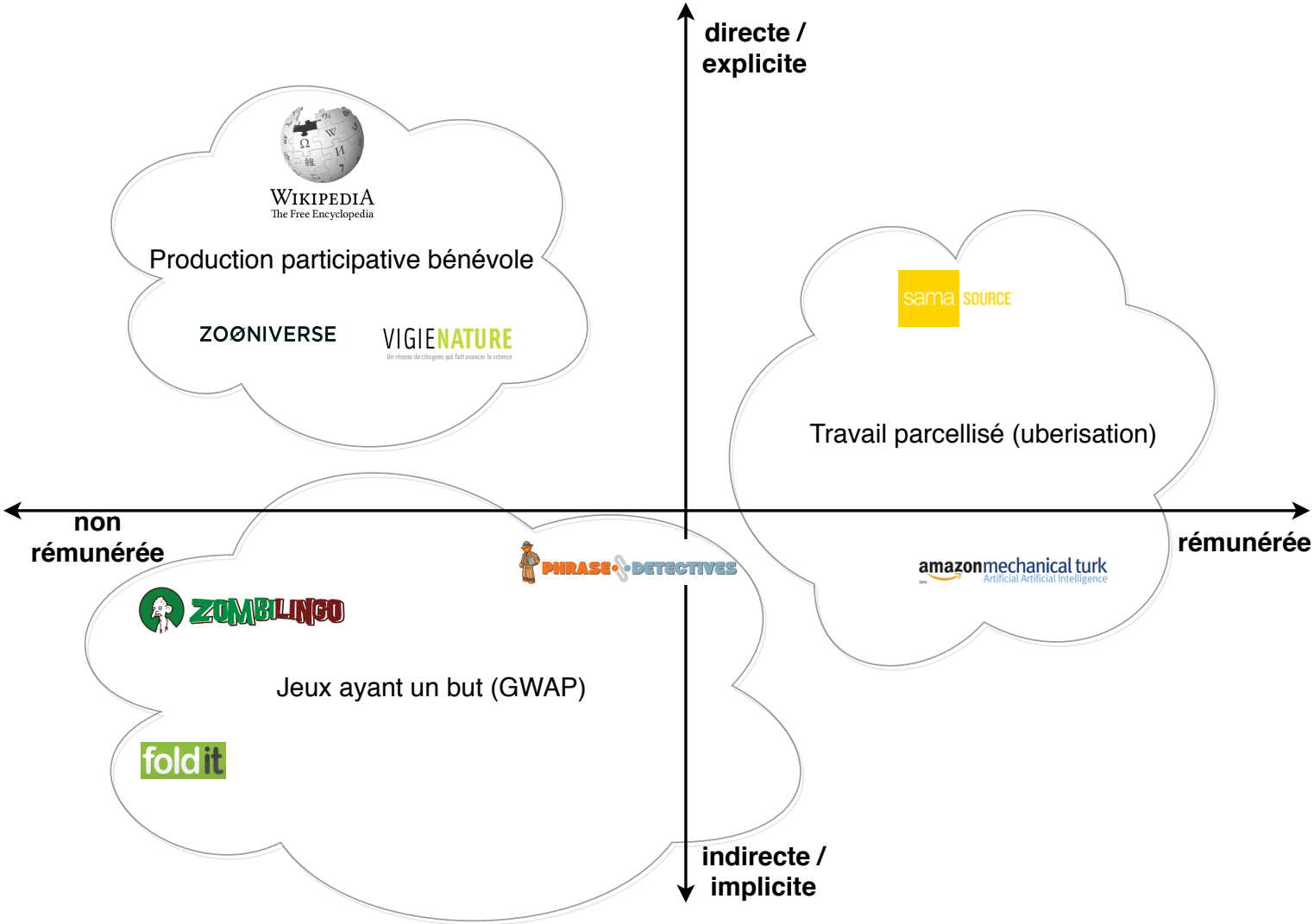
Taxinomie(s) ?

Autant de taxinomies que de points de vue sur la myriadisation

Voir [Geiger et al., 2011] pour un état de l'art détaillé

Une taxinomie (simplifiée) de la myriadisation

parmi d'autres, voir [Geiger et al., 2011]



Des réussites remarquables !

`Wikipédia`[1] (déc. 2014) :

- plus de **30 millions d'articles** en **241** langues
- plus de 8 millions de vues par heure pour la version anglaise, 800 000 pour la version française

`Distributed Proofreaders` (`Projet Gutenberg`)[2] :

- **36 294 livres** numérisés et corrigés
- 624 participants actifs durant les 7 derniers jours

Numérisation des déclarations de conflits d'intérêts des élus[3] :

- 11 095 extraits de déclarations saisis **en 10 jours**
- près de 8 000 participants

---

1. http://stats.wikimedia.org/EN/Sitemap.htm
2. http://www.pgdp.net/c/stats/stats_central.php
3. http://regardscitoyens.org/interets-des-elus/

Visibilité. . .

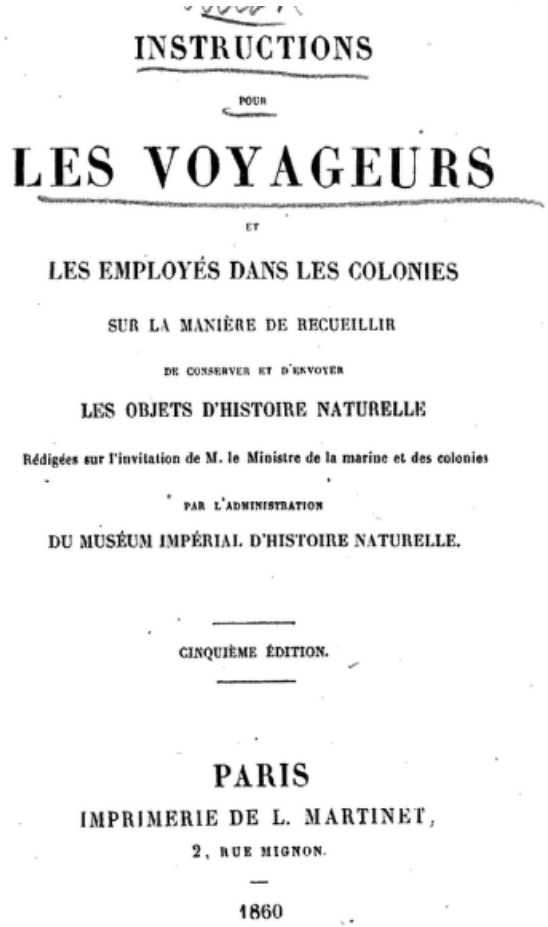
Depuis l'avènement du Web 2.0 :

- accès facilité à une masse de personne inédite
- possibilité d'interagir avec la page visitée (Web 2.0, dit ≪social≫)

Exemples : `Wikipédia`, `Projet Gutenberg` (*Distributed proofreaders*)

. . . n'est pas découverte

Instructions pour les voyageurs et les employés des colonies



INSTRUCTIONS

POUR

LES VOYAGEURS

ET

LES EMPLOYÉS DANS LES COLONIES

SUR LA MANIÈRE DE RECUEILLIR

DE CONSERVER ET D'ENVOYER

LES OBJETS D'HISTOIRE NATURELLE

Rédigées sur l'invitation de M. le Ministre de la marine et des colonies

PAR L'ADMINISTRATION

DU MUSÉUM IMPÉRIAL D'HISTOIRE NATURELLE.

CINQUIÈME ÉDITION.

PARIS

IMPRIMERIE DE L. MARTINET,

2, RUE MIGNON.

1860

Science participative :

- publié par le Museum National d'Histoire Naturelle
- pour que les voyageurs et les employés des colonies :

  "[fassent] connaître les résultats de leurs propres expériences, afin d'en profiter et d'en faire profiter le monde savant"

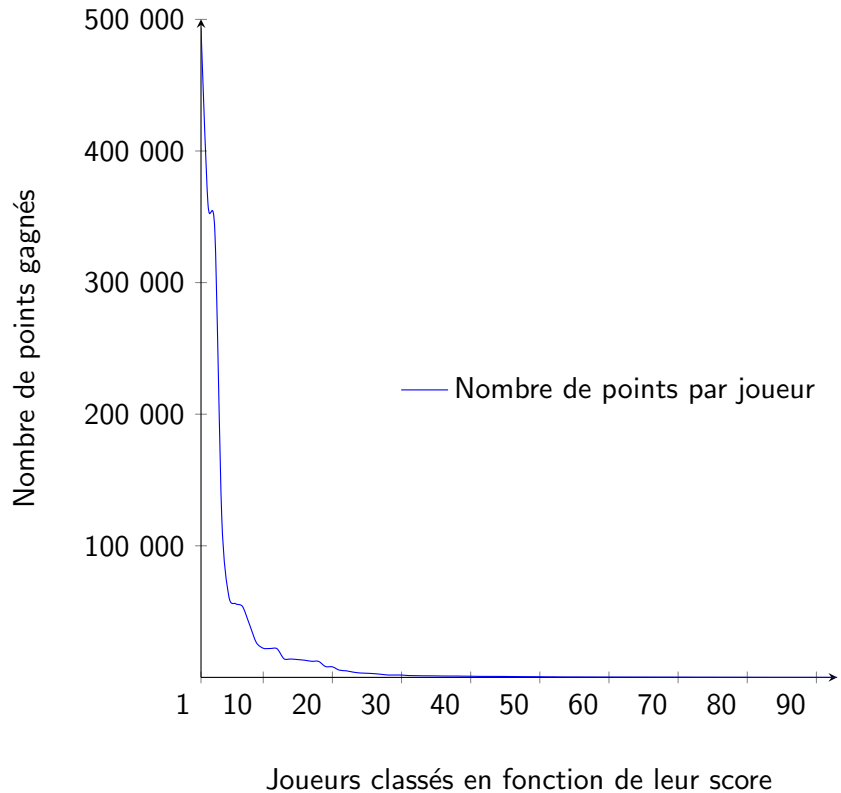
- première édition : **1824**

Autres exemples

- `Ligue de Protection des Oiseaux`[4] :
  - suivis des populations d'oiseaux
  - depuis plus d'un siècle
  - 5 000 bénévoles actifs
- `Longitude Prize`[5] (1714) :
  - prix octroyé par le gouvernement britannique à qui inventerait une méthode simple et pratique permettant de déterminer la longitude d'un navire
  - existe encore : thématique de 2014 = ≪Global antibiotics resistance≫

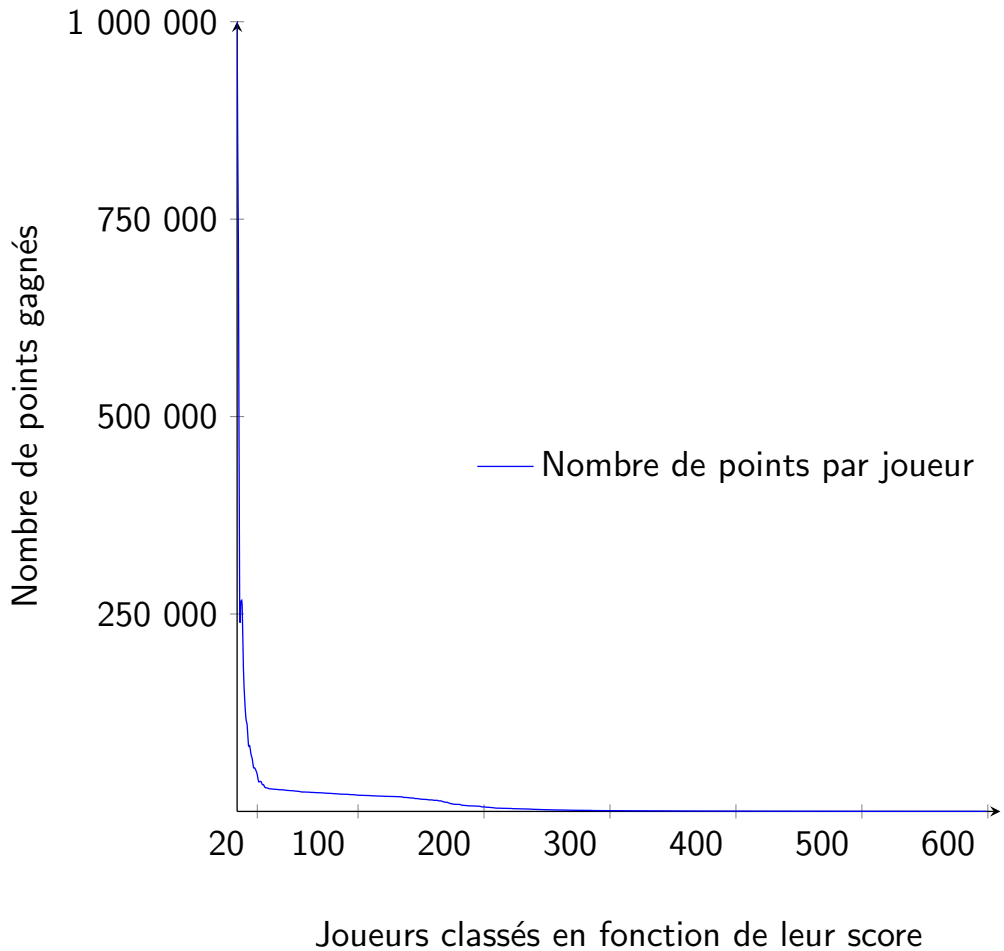
4. https://www.lpo.fr/partager-vos-observations/partagez-vos-observations
5. https://longitudeprize.org/

Une foule de joueurs ? Phrase Detectives [Chamberlain et al., 2013]



Nombre de joueurs sur *Phrase Detectives* en fonction de leur classement en points (février 2011 - février 2012)

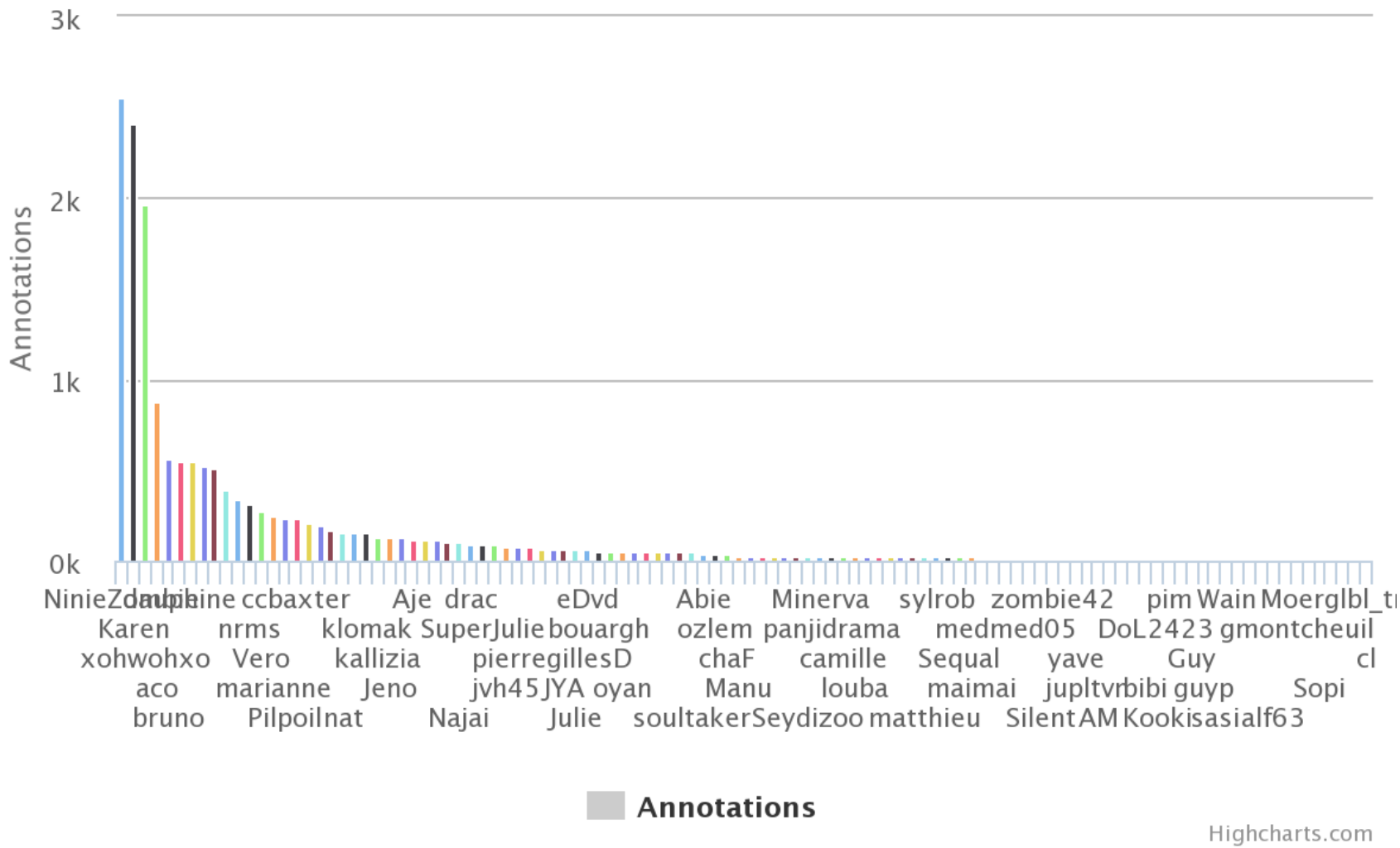
Une foule de joueurs ? JeuxDeMots



Nombre de joueurs sur *JeuxDeMots* en fonction de leur classement en points

(source : http://www.jeuxdemots.org/generateRanking-4.php)

Une foule de joueurs ? ZombiLingo



Une foule de travailleurs ? [Fort et al., 2011]

Nombre de *Turkers* actifs sur `Amazon Mechanical Turk` (MTurk) :

- nombre enregistré sur le site : plus de 500 000
- 80 % des tâches (HIT) sont réalisées par les 20 % de *Turkers* les plus actifs [Deneme, 2009]

$\Rightarrow$ réellement actifs : entre 15 059 et 42 912

Experts *vs* non-experts

Exemple de l'annotation en entités nommées dans un corpus de microbiologie :

- experts du domaine ?
  - du corpus (microbiologie) ?
  - de l'application (TAL) ?

Experts *vs* non-experts

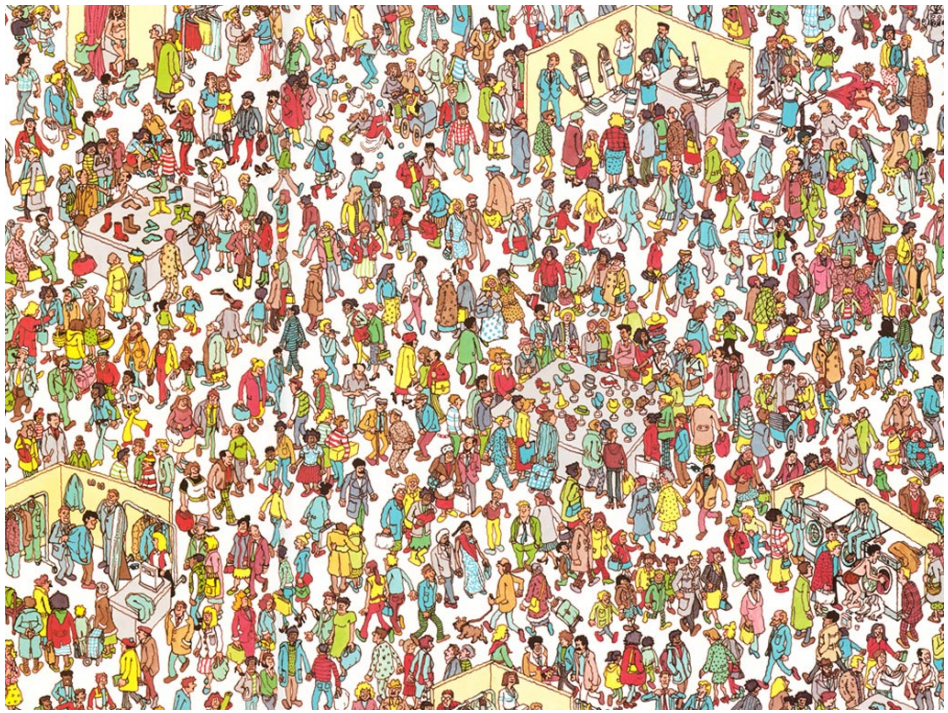
Exemple de l'annotation en entités nommées dans un corpus de microbiologie :

- experts du domaine ?
  - du corpus (microbiologie) ?
  - de l'application (TAL) ?

$\rightarrow$ experts de la tâche

Myriadisation

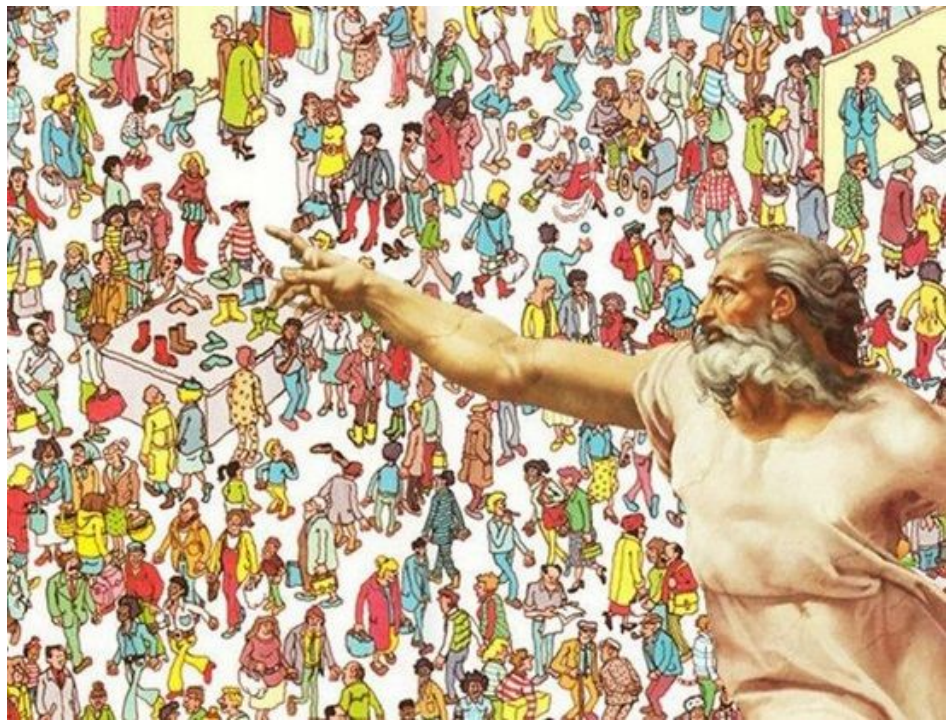
Faire annoter des « non-experts » ?



Myriadisation

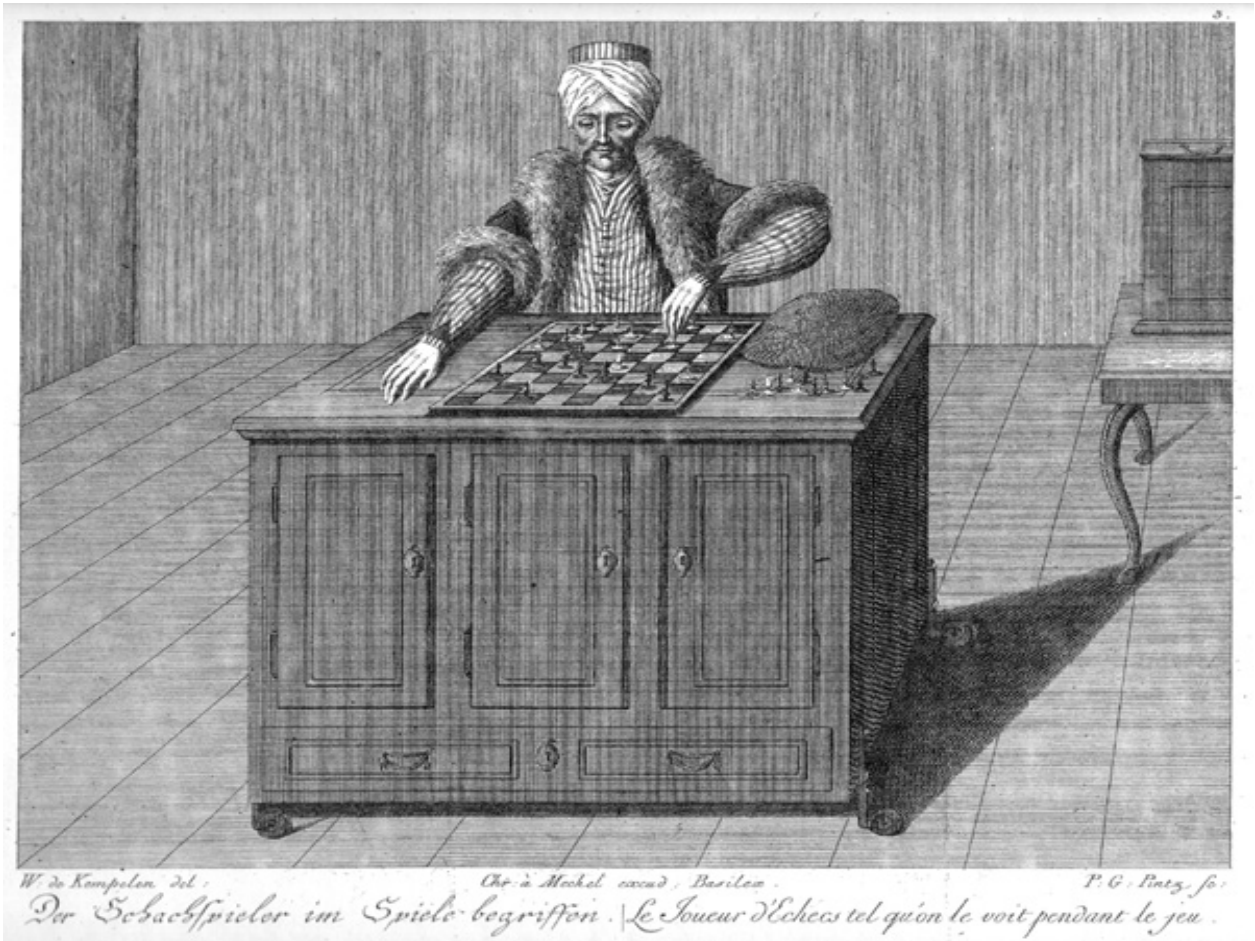
~~Faire annoter des « non-experts » ?~~

→ Trouver/former des experts (de la tâche) dans la foule



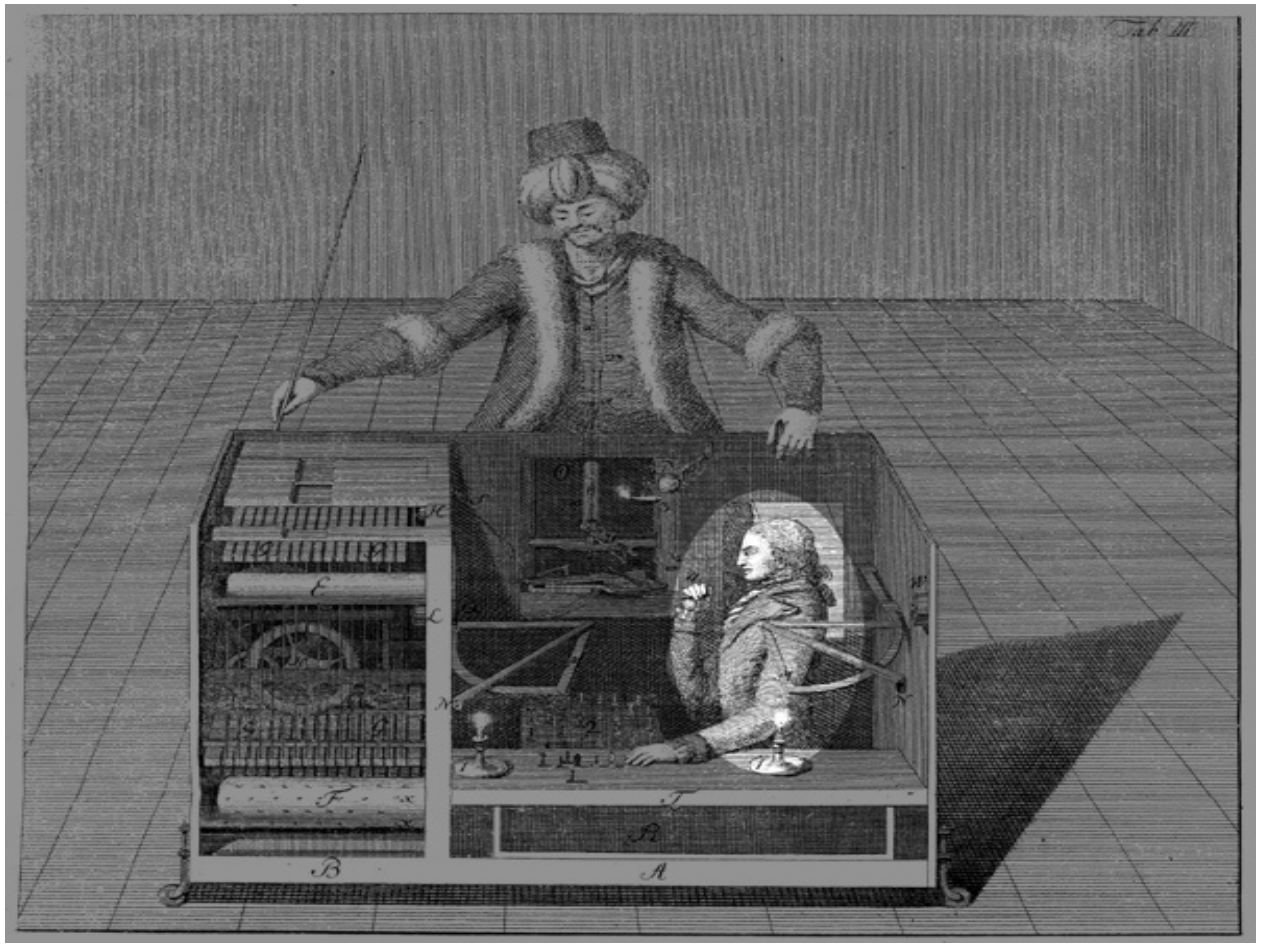
Le «Turc mécanique» de von Kempelen

Un joueur d'échecs mécanique créé par J. W. von Kempelen en 1770 :



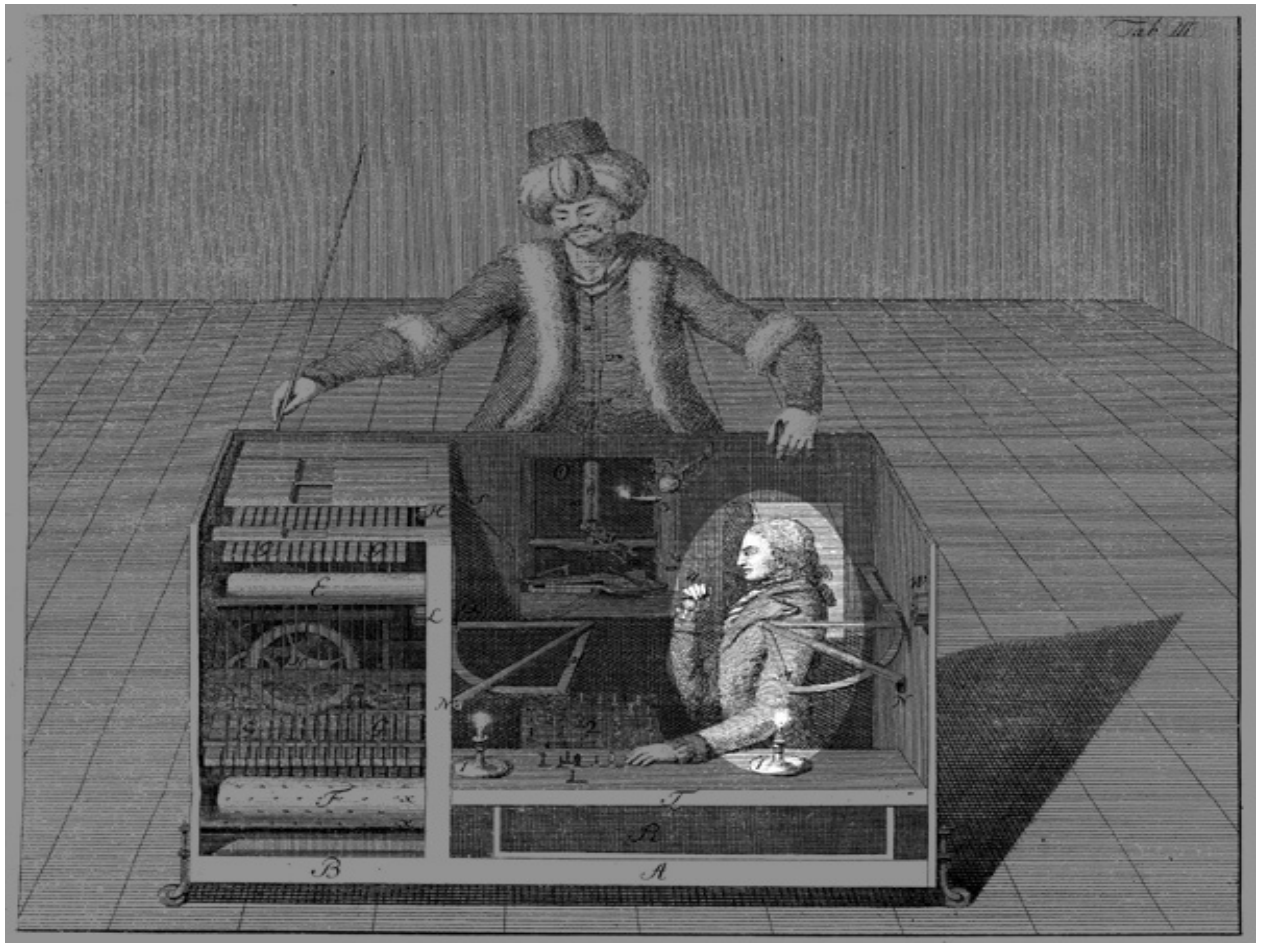
Le «Turc mécanique» de von Kempelen

En fait, un maître d'échecs était caché dans la machine :



Le «Turc mécanique» de von Kempelen

C'est l'intelligence artificielle **artificielle** !



Et Amazon créa AMT

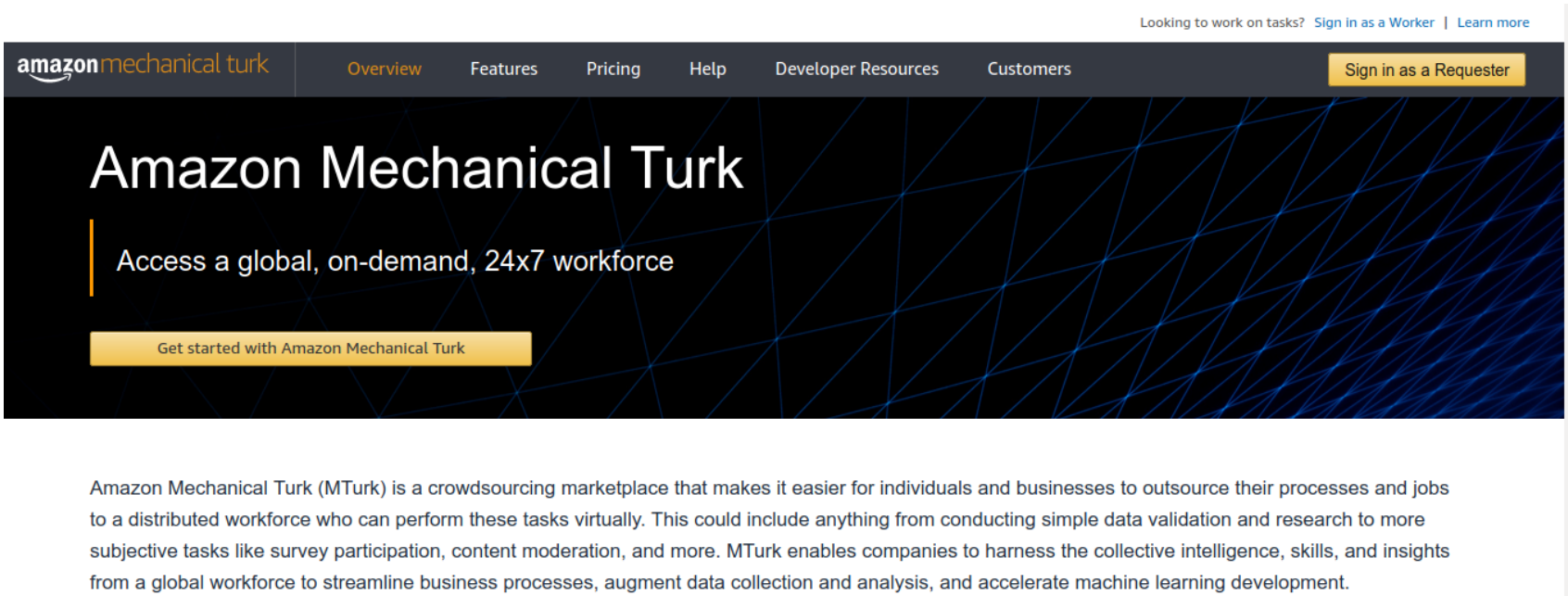
Amazon crée pour ses propres besoins une

plate-forme de travail parcellisé

et en ouvre l'accès en 2005 (moyennant X % des transactions)

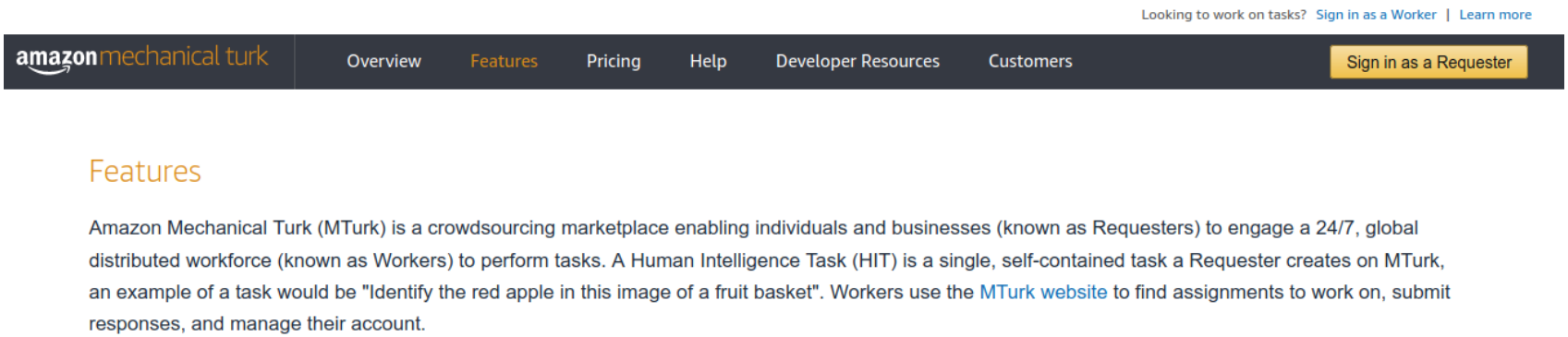
Amazon Mechanical Turk

MTurk



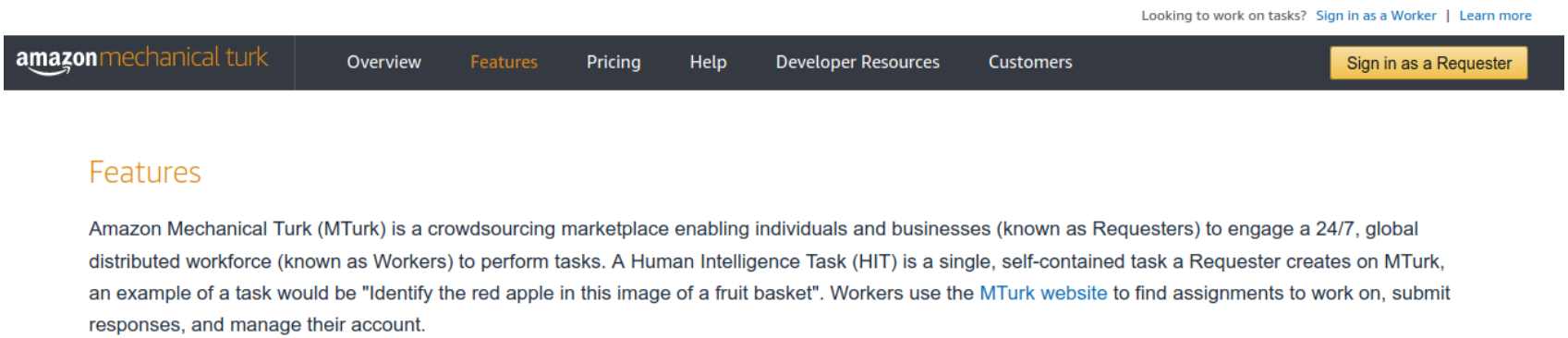
Amazon Mechanical Turk

MTurk est une plate-forme de myriadisation : le travail est *externalisé* via le Web et réalisé par de nombreuses personnes (la *foule*), ici les ~~Turkers~~ **workers**



Amazon Mechanical Turk

MTurk est une plate-forme de myriadisation du travail parcellisé : les tâches sont découpées en sous-tâches (HIT) et leur exécution est payée par les **Requesters**



Amazon Mechanical Turk

MTurk est une plate-forme de myriadisation du travail parcellisé : les tâches sont découpées en sous-tâches (HIT) et leur exécution est payée.

Amazon Mechanical Turk

MTurk est une plate-forme de myriadisation du travail parcellisé : les tâches sont découpées en sous-tâches (HIT) et leur exécution est payée.

How are Workers paid?

Workers will be paid and Amazon Mechanical Turk (MTurk) fees will be charged when you approve submitted work. If you reject the work, the Worker is not paid and you are not charged the MTurk fees. MTurk Prepaid HITs are subject to Participation Agreement. You can review MTurk pricing here.

Caractéristiques d'AMT

Rémunération :

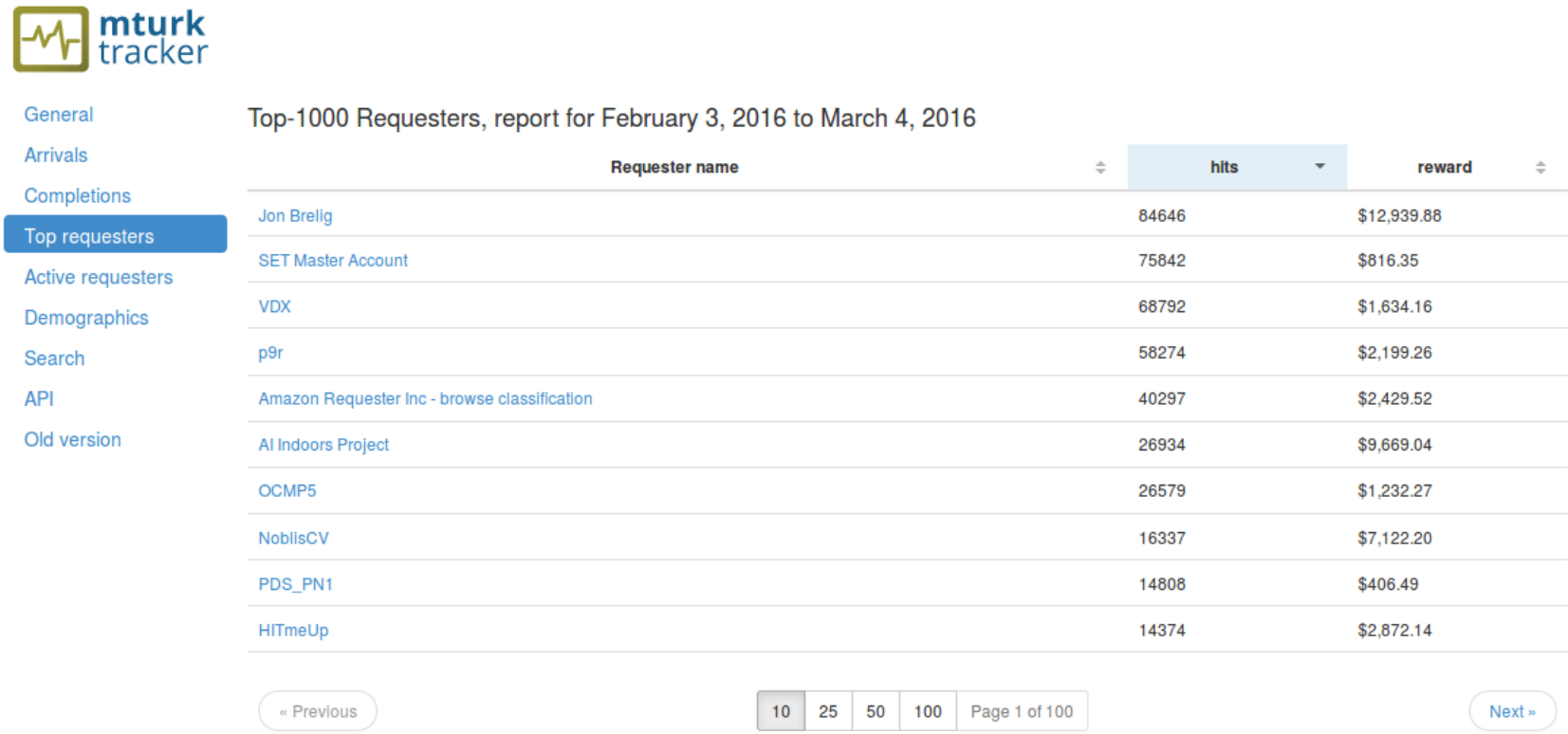
- à la tâche (*illégal* en France sauf (rares) exceptions) : moins de 2 $/h
- pas de relation explicite entre les *Turkers* et les *Requesters*

Tâches :

- nouveaux usages : par exemple, des créations artistiques, comme `http://www.thesheepmarket.com/`
- des tâches traditionnellement réalisées par des employés salariés : transcription, traduction (agences LDC, ELDA), etc

Types de tâches les plus courants sur AMT

`http://www.mturk-tracker.com`, février 2016



- identifications diverses sur des tickets de caisse
- classification binaire d'images (est/n'est pas sur l'image)
- ... et transcription (*CastingWords*), dialogue (*Dialogue Systems*), etc.

AMT : le rêve devenu réalité ?

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

[Snow et al., 2008]

AMT : le rêve devenu réalité ?

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

[Snow et al., 2008]

C'est très peu cher, rapide, de bonne qualité
et c'est un hobby pour les *Turkers* !

AMT permet de réduire les coûts

Très basse rémunération $\Rightarrow$ coûts faibles ? Oui, mais. . .

- coût de mise au point de l'interface
- coût de création de protections contre les spammers
- coût de validation et de post-traitement

certaines tâches (par exemple, la traduction du pachto vers l'anglais) génèrent des coûts similaires aux coûts habituels dans le domaine, du fait du manque de Turkers qualifiés [Novotney and Callison-Burch, 2010].

Quand Amazon se sert...

> Amazon is doubling the fee it collects from “requesters,” those seeking laborers to perform online tasks, to 20% beginning July 21. And for tasks requiring at least 10 people, Amazon will charge an additional 20%, a new fee.

[blog du Wall Street Journal, 23 juin 2015]

AMT permet de produire des ressources de qualité ?

- permet de produire des ressources de qualité dans certains cas précis (par exemple, la transcription simple)
- mais :
  - la qualité est insuffisante lorsque la tâche est complexe (par exemple, le résumé [Gillick and Liu, 2010])
  - l'interface d'AMT pose parfois problème [Tratz and Hovy, 2010]
  - les *Turkers* posent parfois problème (tricheurs, spammers)
  - le modèle de rémunération à la tâche pose problème [Kochhar et al., 2010]
- pour certaines tâches simples les outils de TAL produisent de meilleurs résultats qu'AMT [Wais et al., 2010].

Les HIT (*Human Intelligence Task*) : des tâches simplifiées

Pas de possibilité de se former à la tâche sur AMT :

⇒ Simplification des tâches :

- une ≪ vraie ≫ tâche d'annotation en inférences textuelles (inférence, neutre, contradiction) est réduite à 2 phrases et une question :
  ≪ Would most people say that if the first sentence is true, then the second sentence must be true ? ≫ [Bowman et al., 2015]

AMT : un passe-temps pour les *Turkers* ?

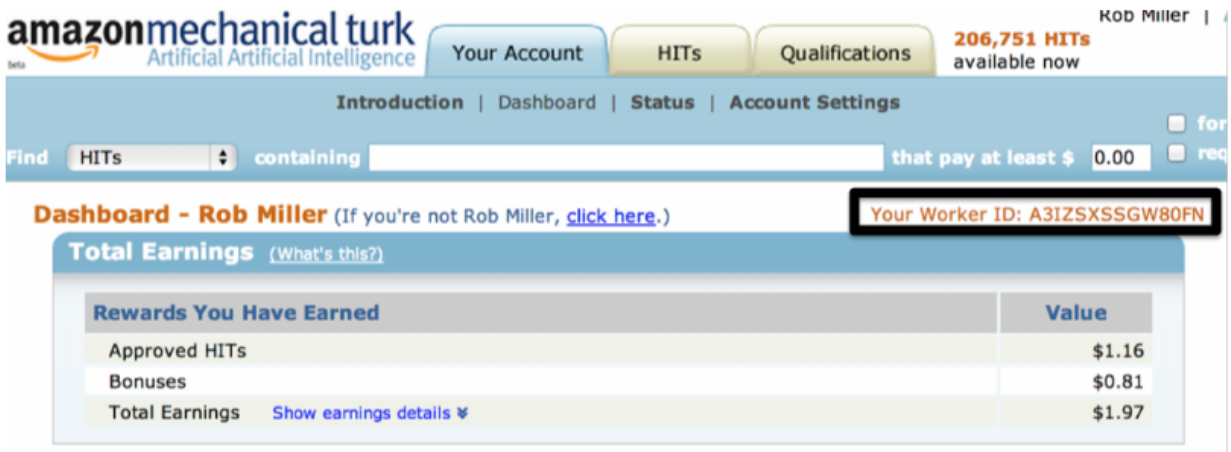
[Ross et al., 2010, Ipeirotis, 2010] montrent que :

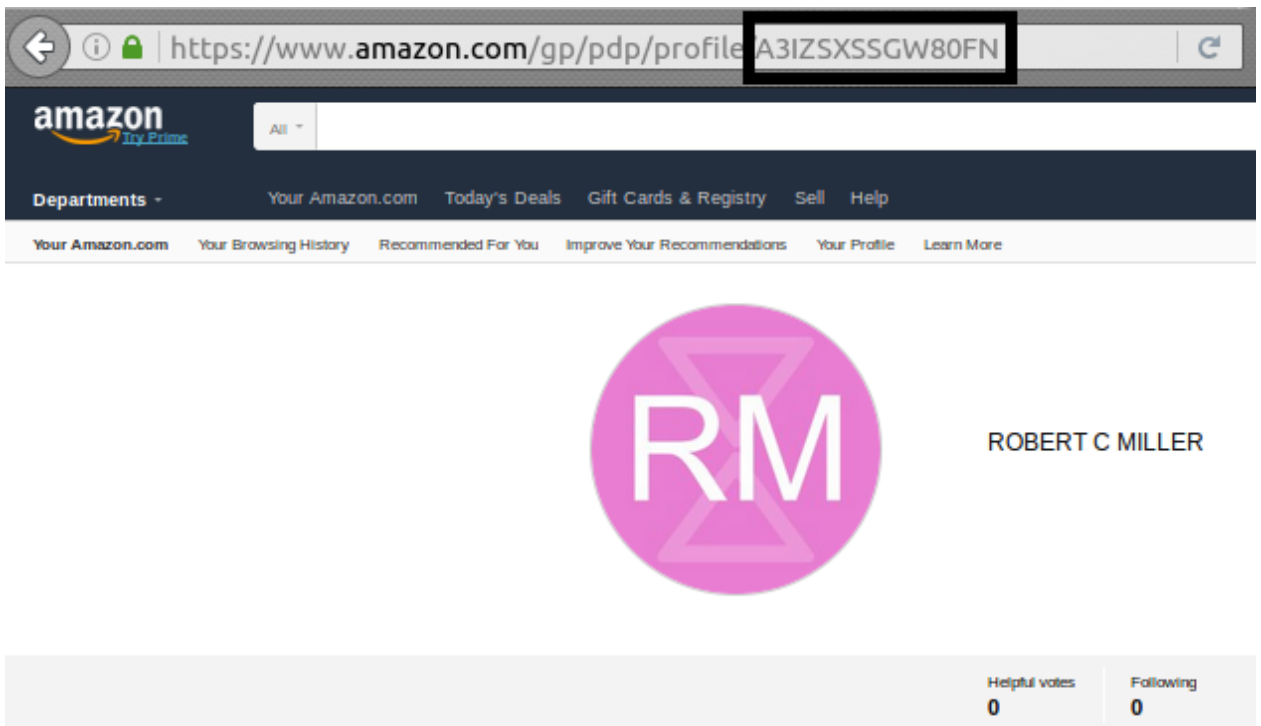
- les *Turkers* sont avant tout motivés par l'**argent** (91 %) :
  - 20 % considèrent AMT comme leur source de revenu primaire ;
  - 50 % comme leur source de revenu secondaire ;
  - l'aspect loisir n'est important que pour une minorité (30 %).
- 20 % des *Turkers* passent plus de 15 h par semaine sur AMT, et contribuent à 80 % des tâches
- le salaire horaire moyen observé est inférieur à 2 $

[Gupta et al., 2014] : en l'absence de possibilité de formation, un important travail caché est réalisé par les *Turkers*

Les *Turkers* ne sont pas anonymes [Lease et al., 2013]

L'id des *Turkers* correspond à leur id client `Amazon`





Est-ce qu'AMT est éthique et/ou légal ?

Éthique :

- pas d'identification : pas de lien officiel entre *Requesters* et *Turkers* et entre *Turkers*
- pas de possibilité de se syndiquer, pour protester contre des manquements des *Requesters* ou ester en justice
- pas de salaire minimum (< 2 \$/h en moyenne)
- possibilité de refuser de payer les *Turkers*

Est-ce qu'AMT est éthique et/ou légal ?

How are Workers paid?

Workers will be paid and Amazon Mechanical Turk (MTurk) fees will be charged when you approve submitted work. If you reject the work, the Worker is not paid and you are not charged the MTurk fees. MTurk Prepaid HITs are subject to Participation Agreement. You can review MTurk pricing here.

Est-ce qu'AMT est éthique et/ou légal ?

Légalité :

- accord de licence d'`Amazon` : les *Turkers* sont considérés comme des travailleurs indépendants $\Rightarrow$ ils sont supposés se déclarer comme tels et payer les cotisations afférentes
- illusoire, vus le niveau de rémunération

$\Rightarrow$ les États perdent une source de revenus légitime

Dépendance à une plateforme externe

Impossibilité de maîtriser :

- les coûts
- les conditions de travail des participants
- la sélection des participants
- les conditions d'expérience

Faire des choix

- **Autres systèmes de myriadisation** :
  - Éthiques : `SamaSource`, une ONG qui permet à des gens, partout dans le monde, d'être formés et payés correctement pour réaliser certaines tâches
  - Jeux ayant un but (Games With A Purpose) : `ESP Game` [von Ahn, 2006], `Phrase Detectives` [Chamberlain et al., 2008], etc

- **Autres solutions** :
  - approches non supervisées, semi supervisées, faiblement supervisées
  - pré-annotation
  - utilisation de ressources existantes (mais peu connues ou oubliées)

Inciter à utiliser d'autres moyens, plus éthiques

`Charte Éthique et Big Data`[6] :

- charte auto-déclarée
- outil pour les agences de moyen (politique de sélection)
- outil de documentation :
  - traçabilité
  - qualité
  - participants
  - licence(s)

6. http://wiki.ethique-big-data.org/

Jouer. . . sérieusement

Jouez à ZombiLUDik : https://zombiludik.org/


Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D. (2015).
A large annotated corpus for learning natural language inference.
arXiv preprint arXiv :1508.05326.


Chamberlain, J., Fort, K., Kruschwitz, U., Lafourcade, M., and Poesio, M. (2013).
Using games to create language resources : Successes and limitations of the approach.
In Gurevych, I. and Kim, J., editors, The People's Web Meets NLP, Theory and Applications of Natural Language Processing, pages 3–44. Springer Berlin Heidelberg.


Chamberlain, J., Poesio, M., and Kruschwitz, U. (2008).
Phrase Detectives : a web-based collaborative annotation game.
In Proceedings of the International Conference on Semantic Systems (I-Semantics'08), Graz, Autriche.


Deneme (2009).
How many turkers are there ?
http ://groups.csail.mit.edu/uid/deneme/.


Fort, K., Adda, G., and Cohen, K. B. (2011).
Amazon Mechanical Turk : Gold mine or coal mine ?
Computational Linguistics (editorial), 37(2) :413–420.


Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., and Schader, M. (2011).
Managing the crowd : Towards a taxonomy of crowdsourcing processes.
In AMCIS 2011 Proceedings.


Gillick, D. and Liu, Y. (2010).
Non-expert evaluation of summarization systems is risky.
In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, CSLDAMT '10, pages 148–151, Stroudsburg, PA, USA. Association for Computational Linguistics.

Gupta, N., Martin, D., Hanrahan, B. V., and O'Neill, J. (2014).
Turk-life in india.
In Proceedings of the 18th International Conference on Supporting Group Work, GROUP '14, pages 1–11, New York, NY, USA. ACM.

Ipeirotis, P. (2010).
The new demographics of mechanical turk.
http ://behind-the-enemy-lines.blogspot.com/2010/03/new-demographics-of-mechanical-turk.html.

Kochhar, S., Mazzocchi, S., and Paritosh, P. (2010).
The anatomy of a large-scale human computation engine.
In Proceedings of Human Computation Workshop at the 16th ACM SIKDD Conference on Knowledge Discovery and Data Mining, KDD 2010, Washington D.C.


Lease, M., Hullman, J., Bigham, J. P., Bernstein, M. S., Kim, J., Lasecki, W., Bakhshi, S., Mitra, T., and Miller, R. C. (2013).
Mechanical turk is not anonymous.
Technical report.


Novotney, S. and Callison-Burch, C. (2010).
Cheap, fast and good enough : automatic speech recognition with non-expert transcription.
In Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), HLT'10, pages 207–215, Stroudsburg, PA, USA. Association for Computational Linguistics.


Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. (2010).
Who are the crowdworkers ? : shifting demographics in mechanical turk.

In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, CHI EA '10, pages 2863–2872, New York, NY, USA. ACM.

Sagot, B., Fort, K., Adda, G., Mariani, J., and Lang, B. (2011).
Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé.
In Actes de Traitement Automatique des Langues Naturelles (TALN), Montpellier, France.
12 pages.

Snow, R., O'Connor, B., Jurafsky, D., and Ng., A. Y. (2008).
Cheap and fast - but is it good ? evaluating non-expert annotations for natural language tasks.
In Proceedings of EMNLP 2008, pages 254–263.

Tratz, S. and Hovy, E. (2010).
A taxonomy, dataset, and classifier for automatic noun compound interpretation.

In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 678–687, Uppsala, Suède. Association for Computational Linguistics.

von Ahn, L. (2006).
Games with a purpose.
IEEE Computer Magazine, pages 96–98.

Wais, P., Lingamneni, S., Cook, D., Fennell, J., Goldenberg, B., Lubarov, D., Marin, D., and Simons, H. (2010).
Towards building a high-qualityworkforce with mechanical turk.
In Proceedings of Computational Social Science and the Wisdom of Crowds (NIPS).