

Ethique et construction collaborative de données lexicales par des GWAP

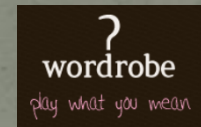
quelques leçons tirées de l'expérience JeuxDeMots

M. Lafourcade & N. Le Brun
LIRMM- Univ. Montpellier 2 & Imaginat



Problématique

- Dans le contexte des GWAP (Games With A Purpose) [Ahn et al.]
 - Rappel : jeux conçus pour que l'activité des joueurs génère des données
 - Pour la création de données pour le TAL
 - Phrase Detective [Chamberlain et al.]
 - WordRobe [Bos et al.]
 - ZombiLingo [Fort et al.]
 - JeuxDeMots [Lafourcade et al.]
 - ...
- et tous les autres GWAP hors TAL



Quel cadre éthique ? Que peut-on faire et ne pas faire ?

Attention pas de morale ici, mais une approche pragmatique

Annnonce des objectifs du projet

- Être clair sur la nature du projet et le but du jeu
 - certains joueurs attirés par le jeu manifestent rapidement intérêt et curiosité... pour le projet
 - se sentir utile et impliqué dans un projet scientifique peut être une motivation
- Rendre accessibles les données du projet  
 - données **effectivement** accessibles, gratuites, libres de droit et lisibles
 - facilité d'accès (lien pas mort), mise à jour fréquente...
[JDM Diko]
 - peu de GWAP le font
 - peur du vol
 - peur de rendre accessibles des données pas si bonnes que ça...



Annnonce des objectifs du projet

- Rendre accessibles les données du projet

- certains prétendent le faire, mais...

chemin de croix pour y accéder

ex : envoyer une lettre de motivation aux administrateurs et remplir un formulaire où on jure sur l'honneur qu'on ne dira pas du mal des données – réponse sous 6 mois...

Diko



- La plupart n'évoquent même pas les données produites c'est le cas notamment de Foldit, EteRNA, Nanocrafter, Fraxinus, Eyewire, Citizen Sort, Nightjar Project, Nanodoc, Dizzez, The Cure, Malaria Training Game, etc.



Données personnelles protégées et organisateurs du jeu identifiés

- Ne jamais divulguer les données personnelles
 - adresse mail, genre, âge
en pratique, elles sont en général sans intérêt
 - À l'inscription, inviter à choisir un pseudo
éviter identifiant = adresse mail
 - ne pas rendre possible l'identification d'un participant
à partir des données, notamment
 - ex de JDM : les données validées dans le réseau lexical
ne portent aucune trace de qui a fait quoi...
 - déclaration à la CNIL si nécessaire.



Données personnelles protégées et organisateurs du jeu identifiés

- Avoir une charte claire
 - que les joueurs doivent lire et approuver
même s'ils ne la lisent pas souvent
 - qui identifie les contacts, organisateurs, auteurs,
institutions, etc.



Eviter la publicité dans le jeu

- Renforce la crédibilité
 - prouve l'absence de visée commerciale
 - augmente la confiance du joueur
- La publicité peut gêner le joueur
 - au mieux le joueur y est indifférent
 - au pire, il ne jouera pas par peur du spam
- L'argument « on doit financer le projet » ne tient pas
 - coût des GWAP faible
sinon sans intérêt
 - Les GWAP sont le plus souvent proposés par des institutions universitaires
>> pas de financement possible par de la pub



Aucun gain en numéraire ou lots quelconques

- Payer les joueurs est à proscrire
 - même de façon déguisée *via* des lots de valeur
 - lots symboliques (stickers, magnets, etc.) **semblent ok**
 - favoriser les récompenses sous d'autres formes, non matérielles
- Un joueur payé a tendance à produire des données mauvaises
 - exemple de projets : SMS4Sciences
envoi massif de données sans intérêt
 - cas d'AMT (cas de pure myriadisation) [Sagot et al.]
 - surtout si on récompense la quantité... sachant que la qualité est plus dure à évaluer



Aucun gain en numéraire ou lots quelconques

- Moralement douteux
 - à la limite de l'emploi caché ...
 - problème légaux possibles
- Dérives fréquentes
 - exemple avec les lots : Phrase Detective, ...



Transparence de la méthode de création des données

- Donner les grandes lignes du principe
 - > éviter de développer une impression d'arbitraire
- *mais ne pas nécessairement entrer dans les détails de la mécanique mise en œuvre*
 - > éviter de donner prise à l'exploitation de défauts ou bugs
- Projet collectif
 - associer les joueurs au projet, communauté ouverte
 - publications collectives (FoldIt, Eterna, EyeWire)
 - données accessibles (bis répétita)

Conclusion

- Nécessité de transparence, faute de quoi
risque de mort rapide du projet faute de joueurs
risque de suspicion (légitime)
- Rendre les données accessibles aux joueurs
et aussi à tous les autres...
non seulement légitime mais utile
- Faire des projets sincères
pour produire des données utiles
*et non pour faire des publis, faire passer une thèse ... et
après plus rien...*
- Toutes les données du TAL ne se prêtent pas aux
GWAP et faire un GWAP certes **ne coûte pas cher** mais
n'est pas facile