

Université de Montréal

WikiGames : une plateforme de jeux dédiée à la validation d'une base de connaissances produite à partir de techniques d'extraction d'information ouverte

par
Kevin Forand

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Mémoire présenté à la Faculté des arts et des sciences
en vue de l'obtention du grade de Maître ès sciences (M.Sc.)
en informatique

Août, 2017

© Kevin Forand, 2017.

RÉSUMÉ

L'extraction d'information ouverte permet la création de larges collections de triplets relationnels à partir de corpus de textes non structurés. Ces larges collections de triplets extraits contiennent souvent une grande quantité de triplets bruités n'apportant aucune information utile. Ces collections peuvent atteindre des tailles rendant la validation manuelle trop longue pour être réalisées par un petit groupe de personnes en un temps convenable et il serait dans bien des cas trop dispendieux pour ces équipes d'engager le personnel nécessaire pour cette tâche.

L'utilisation de jeux à des fins de production participative a permis, lors de tâches similaires, de recueillir un grand ensemble de bénévoles. Nous nous sommes donc intéressés à inférer, à partir d'une de ces collections de triplets bruités qui fut précédemment générée à partir de techniques d'extraction d'information ouverte, un ensemble de connaissances potentiellement utiles et pertinentes et avons ensuite amorcé la validation de cette base de connaissances par l'intermédiaire d'une plateforme de jeux.

Mots clés: traitement automatique des langues naturelles, base de connaissances, apprentissage automatique, jeux.

ABSTRACT

Open information extraction techniques can generate a large amount of relation triplets from unstructured corpus of texts. These large collections of triplets often contain a good portion of noisy triplets that brings little to no usable information. These collections of triplets can become too large to be manually validated by most small teams in a reasonable amount of time and hiring the number of validators required for such task would be too costly for most teams. The use of games as a crowdsourcing tool has shown great success in acquiring a large pool of volunteer for the realization of similar tasks.

We have therefore looked into the extraction of a set of useful knowledge from a rather large and noisy relation triplets collection that was previously extracted using an open information extraction tool. We have then started the process of validating the resulting knowledge base with the help of a games with a purpose platform.

Keywords: natural language processing, knowledge base, machine learning, games with a purpose.

TABLE DES MATIÈRES

RÉSUMÉ	ii
ABSTRACT	iii
TABLE DES MATIÈRES	iv
LISTE DES TABLEAUX	vii
LISTE DES FIGURES	ix
LISTE DES ANNEXES	x
LISTE DES SIGLES	xi
REMERCIEMENTS	xii
CHAPITRE 1 : INTRODUCTION	1
CHAPITRE 2 : BASE DE CONNAISSANCES	4
2.1 Données initiales	4
2.1.1 Triplets relationnels	4
2.1.2 Origine des données initiales	5
2.1.3 Données présentes dans la collection	5
2.1.4 Triplets problématiques	6

2.2	Préfiltrage	8
2.2.1	Pronoms	8
2.2.2	Symboles	9
2.2.3	Fréquentiel	10
2.2.4	État de la base post filtrage	11
2.3	Inférences	13
2.3.1	Classification	13
2.3.2	Profil relationnel	15
2.3.3	Déduction des relations importantes	20
2.4	Similarités	21
2.4.1	Similarités entre termes et catégories	22
2.4.2	Similarités entre les catégories	22
2.4.3	Similarités entre termes	24
2.4.4	Distributions des Arg2	24
CHAPITRE 3 : GAMES WITH A PURPOSE		26
3.1	Jeux de production participative	26
3.1.1	GWAPS dans le domaine du TALN	27
3.1.2	GWAPS dans les autres domaines	28
3.1.3	Études sur les GWAPS	30
3.2	Conception de WikiGames	31
3.2.1	Principes de conceptions	31

3.2.2	Questionnaires associatifs	33
3.2.3	Traitement de joueurs adversatif	39
3.2.4	Choix de conceptions	42
3.3	Plateforme WikiGames	44
3.3.1	Structure générale des jeux	44
3.3.2	Écran d'accueil	45
3.3.3	Simple	46
3.3.4	Pinball	47
3.3.5	Élimination	48
3.3.6	Écran de fin de partie	49
CHAPITRE 4 : EXPÉRIENCES		51
4.1	Dévoilement des jeux	51
4.2	Données récoltées pour chaque partie	51
4.3	Analyse des résultats	53
4.3.1	Analyse quantitative des résultats	53
4.3.2	Analyse qualitative des résultats	57
4.4	Retour sur les résultats	60
4.4.1	Perspectives futures	60
CHAPITRE 5 : CONCLUSION		63
BIBLIOGRAPHIE		64

LISTE DES TABLEAUX

I	Exemples triplets extraits	6
II	Types triplets problématiques	7
III	Exemples triplets contenant au moins un pronom	9
IV	Triplets contenant un champ ne contenant que des symboles	10
V	Exemples triplets hapax	11
VI	Échantillon aléatoire des triplets restants	12
VII	Exemples catégories retenues	14
VIII	Exemples catégories éliminées	14
IX	Extrait du profil relationnel du terme <i>Obama</i>	16
X	Termes associés à la catégorie <i>un chanteur</i>	17
XI	Extrait du profil relationnel de la catégorie <i>un chanteur</i>	18
XII	Extrait du profil relationnel moyen	19
XIII	Exemple de relations importantes pour la catégorie <i>un chanteur</i>	20
XIV	Extrait de similarités entre le terme <i>Thomas Edison</i> et catégories	22
XV	Extrait de similarités entre la catégorie <i>un poète</i> et l'ensemble des catégories	23
XVI	Distribution des Arg2 pour la relation <i>être</i> et la catégorie <i>un auteur</i>	25
XVII	Paramètres pour les questionnaires d'assignation de catégories	36
XVIII	Exemple de sortie pour un questionnaire d'assignation de catégories	36

XIX	Paramètres pour les questionnaires profil relationnel de terme . . .	37
XX	Exemple de sortie pour un questionnaire associé au terme "Charles de Gaulle"	37
XXI	Paramètres pour les questionnaires profil relationnel de catégorie .	38
XXII	Exemple de sortie pour un questionnaire associé à "un monument"	39
XXIII	Description des données recueillies pour chaque partie	52
XXIV	Statistique sur les parties jouées	54
XXV	Statistiques sur les questionnaires	54
XXVI	Statistiques sur les jeux	54
XXVII	Temps passé sur les jeux	55
XXVIII	Annotations par heure	57
XXIX	Exemples d'assignations validées	58
XXX	Exemple de catégories considérées non valides pour une entité . .	58
XXXI	Résultats de questionnaires de profils relationnels	59

LISTE DES FIGURES

3.1	Écran d'accueil WikiGames	45
3.2	Jeu Simple	46
3.3	Jeu Pinball	47
3.4	Jeu Élimination	48
3.5	Écran fin de partie	50

LISTE DES ANNEXES

LISTE DES SIGLES

GWAP Game with a purpose

GWAPS Games with a purpose

TALN Traitement automatique des langues naturelles

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche Philippe Langlais pour tout le savoir, la supervision, l'aide et les judicieux conseils qu'il m'a apportés tout au long de ma maîtrise.

Je tiens à également remercier tous les membres du RALI et du DIRO pour toutes les rencontres et discussions qui ont contribué à l'enrichissement de mes connaissances.

Je tiens à finalement remercier ma famille pour tout le support qu'ils m'ont apporté tout au long de mes études.

CHAPITRE 1

INTRODUCTION

L'acquisition de connaissances chez l'être humain est un processus débutant dès la jeune enfance et est grandement renforcée quelques années plus tard avec l'acquisition de l'habileté de lecture. Cette faculté de reconnaître une série de symboles et d'en extraire une sémantique nous permet au cours d'une vie d'assimiler une quantité astronomique d'informations à partir de multiples sources telles que : livres, manuels scolaires, pages web, messages textes, journaux, etc. Cet ensemble de ressources textuelles sans cesse grandissant motive la recherche et le développement d'outils informatiques tentant de s'approcher de ce processus cognitif. Bien que ces outils possèdent des capacités de raisonnement nettement inférieures à celle que possède un être humain, on a pu observer des progrès substantiels au cours des dernières années quant à la qualité et la quantité d'informations que ces outils permettent d'extraire et d'inférer à partir de ressources textuelles.

Ces progrès combinés à l'explosion du nombre et de la taille de corpus disponibles qu'a amenée la prolifération de textes via internet ont permis la création de bases de connaissances de plus en plus massives qui afin de maintenir un certain niveau de qualité nécessite généralement des phases de validation de plus en plus extensive. Ces processus de contrôle de qualité nécessitent donc l'investissement d'un nombre d'heures-personnes de plus en plus imposant empêchant les équipes de petite taille à réaliser cette tâche dans

un temps raisonnable. Pour combler ce manque de ressources humaines, de nombreux projets se sont tournés vers l'utilisation de méthodes de production participative. Ces techniques proposent de faire appel au grand public pour le recrutement d'un bassin de volontaires dans le but de réaliser de larges tâches étant trop incombant pour un petit groupe d'individus en décomposant celles-ci en une série de sous-tâches pouvant être individuellement réalisable et distribuant celles-ci parmi les volontaires. Pour recruter la quantité de volontaires nécessaire, ces méthodes reposent sur l'utilisation d'un ou plusieurs incitatifs. L'incitatif le plus simple est de miser sur le désir des gens d'aider la science en contribuant au projet. Cet incitatif peut donner des résultats très variables en fonction de la tâche qu'elle vise à compléter et n'est généralement pas suffisant pour recruter la quantité nécessaire de volontaires pour compléter l'ensemble des sous-tâches. Il est donc courant de faire appel à un second incitatif pour réussir à attirer la masse de contributeurs nécessaire. L'utilisation de service tel que Amazon Mechanical Turks est donc monnaie courante en proposant un incitatif financier à ses utilisateurs en récompensant chaque tâche complétée avec une petite compensation financière. Cependant lorsque le nombre de sous tâches à compléter est relativement élevé, ces petits montants s'accumulent rapidement pour donner un total non négligeable limitant l'utilisation d'un tel incitatif à un petit groupe d'individus ayant à leur disposition le budget nécessaire. Pour contrôler ces coûts et empêcher qu'ils ne deviennent exorbitants, certains projets ont opté pour un incitatif ludique et décidé d'investir leurs temps et budget dans le développement de jeux dont les mécaniques de jeux mènent à la réalisation des sous-tâches.

Ce mémoire s'intéresse donc à l'utilisation de ce type de jeux pour la validation d'une base de connaissances dont le contenu fut extrait et inféré automatiquement à partir d'une collection de données textuelles. Nous présentons donc premièrement une telle base de connaissances et proposons ensuite une structure de questionnaires associatifs permettant de faire le pont entre les informations contenues dans une base de connaissances et une plateforme de jeux sérieux. Cette structure intermédiaire a comme objectif d'augmenter l'indépendance entre les données à valider et les jeux utilisés. Cette stratégie pourrait donc permettre de développer des jeux sérieux sans devoir connaître au préalable le contenu de la collection à valider et permettrait également de réutiliser ces jeux pour la validation d'autres collections et types de données. Finalement nous introduisons la plateforme de jeux Wikigames permettant d'instancier cette structure de questionnaires associatifs vers un ensemble de jeux visant à recueillir les réponses de ses joueurs.

CHAPITRE 2

BASE DE CONNAISSANCES

2.1 Données initiales

L'acquisition et le regroupement de connaissances sont des tâches ayant fait l'objet de multiples études qui à leur tour ont fait naître une panoplie d'outils tentant d'accomplir cette tâche. On note entre autres la plateforme NELL (Carlson et al., 2010) parcourant le Web continuellement pour en extraire une quantité grandissante d'informations. On retient également l'outil Reverb (Fader et al., 2011) qui en utilisant des techniques d'extraction d'information ouverte (Banko et al., 2007) permet l'acquisition d'une collection de triplets relationnels et plus récemment le projet Knowledge Vault (Dong et al., 2014) de Google combinant de multiples modèles à priori avec la sortie de plusieurs extracteurs pour bâtir une base de connaissances de taille impressionnante.

2.1.1 Triplets relationnels

Notre projet s'intéresse plus particulièrement à l'information contenue ou pouvant être inférée à partir d'une collection de triplets relationnels telle que celle produite par l'utilisation d'outils d'extraction d'information ouverte (Banko et al., 2007). Ces outils d'extractions prennent un corpus de texte(s) en entrée et à partir du texte produisent une collection de triplets relationnels (E_1, \mathbf{R}, E_2) définissant une relation \mathbf{R} entre deux

entités E_1 et E_2 . Par exemple, (Etzioni et al., 2008) note que l’outil TextRunner lorsque présenté avec la phrase "Kafka, a writer born in Prague, wrote The Metamorphosis." extrait le triplet (*Kafka, born in, Prague*).

2.1.2 Origine des données initiales

Au lieu de procéder à l’extraction d’une nouvelle collection de triplets, nous avons opté pour récupérer une collection déjà existante. Cette dernière, décrite dans (Gotti et Langlais, 2016), fut précédemment extraite à partir d’un outil adaptant les règles d’extractions présentes dans l’outil Reverb (Fader et al., 2011) pour qu’elles s’appliquent au français et fut ensuite utilisé sur un corpus contenant 1.5 million d’articles qui proviennent de la version française de Wikipédia datant de juin 2014.¹

2.1.3 Données présentes dans la collection

Cette collection contient initialement un total de 30.4 millions de triplets relationnels étant sous la forme (**Arg 1** == **Relation** == **Arg 2**) tel que l’on peut observer dans l’échantillon présenté dans la table I. Cependant parmi ces millions de triplets une portion non négligeable n’apporte que très peu ou parfois aucune information utile tels que ceux présentés en section 2.1.4. Cibler ces triplets problématiques et procéder à leur élimination pourrait être bénéfique pour rehausser la qualité de la base de connaissances. En effet, leur inclusion en plus d’alourdir inutilement la collection risque de bruyier et ainsi réduire la qualité des inférences impliquant ces triplets.

¹<http://rali.iro.umontreal.ca/rali/?q=fr/wikipedia-dump>

Arg 1	Relation	Arg 2
Vaira Vīķe-Freiberga	être réélire en	2003
d' un Conseil européen extraordinaire	se tenir sous	présidence espagnole
Ce crabe	être doter de	longues pattes fines
The Show	must	go
Symphony X	écrire	une chanson
Venise	aider militairement	l' Empire byzantin
elle	arrêter	l' émission
un logiciel libre	écrire en	Vala
Anetia briarea	être	un grand papillon
Il	léguer	200 livres

Tableau I – Exemples triplets extraits

2.1.4 Triplets problématiques

Comme mentionné précédemment, l'outil d'extraction qui fut à l'origine de la collection de triplets a extrait des triplets problématiques ou jugés non informatifs. En effet, cet outil étant basé sur Reverb, il hérite malheureusement de certains problèmes identifiés dans (Fader et al., 2011). De plus comme mentionné dans (Gotti et Langlais, 2016), cette version adaptée pour le français met une emphase sur le rappel au détriment de la précision et capture donc plus de relations entre les entités, mais également plus de bruit. L'élimination de ces triplets pourrait être effectuée à partir de méthodes de science citoyenne tel que les jeux que nous réalisons au cours de ce projet. Cependant, puisque le nombre de validations/heures que nous pouvons obtenir de nos jeux n'est pas illimité, nous réservons l'utilisation de ces jeux pour la validation d'information ne pouvant pas être facilement traitée automatiquement. De plus, nous considérons qu'une bonne partie de ces triplets pouvant être éliminés automatiquement serait peu intéressante à évaluer pour un utilisateur ce qui pourrait éventuellement causer certains joueurs à quitter la

plateforme. Nous considérons qu'il est en effet généralement moins intéressant pour un utilisateur d'évaluer la pertinence de triplets tels que (*il == être == le*) ou (*\$ == avoir ==!*) comparé à d'autre tel que (*Star Wars == être réaliser par == Georges Lucas*) demandant à l'utilisateur de faire appel à ses connaissances. Nous avons donc identifié à l'intérieur de la collection certaines catégories de triplets problématiques et pour chacune de ces catégories nous avons tenté de déterminer un moyen pour détecter et éliminer les triplets lui étant associés. Nous présentons dans le tableau II les principaux types de triplets problématiques que nous avons identifiés et séparons ces types de triplets entre ceux qui seront traités automatiquement par la phase de préfiltrage et ceux nécessitant une attention plus particulière.

Types traités par la phase de préfiltrage			
Type	Exemple		
Pronoms	Il	être	le
Arguments formés uniquement de symboles	L	= \mathbb{R} [X]/(X ² +1
Alphabet non français	la ville	être connaître sous	共青城
Arguments hapax	<i>le GR II36</i>	compter toujours	<i>4 MB174</i>

Types n'étant pas traités par la phase de préfiltrage			
Type	Exemple		
Coréférences non pronominales	sa population	être de	103
Mauvaise identification des args	Montréal	faire	naufage
Fausses informations	la terre	être considérer comme	plate
Fautes d'orthographe récurrente	Il	étudier	le <i>cinema</i>

Tableau II – Types triplets problématiques

2.2 Préfiltrage

Dans cette section, nous décrivons les différentes mesures et étapes qui furent utilisées lors de la phase de préfiltrage ayant permis la détection et l'élimination de triplets appartenant à au moins une des catégories problématiques présentes dans le tableau II.

2.2.1 Pronoms

L'absence d'une bonne chaîne de résolution de coréférence cause une bonne quantité des triplets extraits à contenir des coréférences non résolues. Il est sans grande surprise qu'une bonne portion de ces coréférences impliquent l'utilisation de pronoms. Malheureusement, l'information pouvant être extraite de triplets comme (Il == être == un président) est minime. Leur inclusion risquerait donc de bruyé le calcul des statistiques qui seront utilisées pour les inférences décrites dans la section 2.3. L'élimination de ces triplets peut s'effectuer en utilisant une liste de pronoms et en comparant chaque pronom P présent dans cette liste à l'ensemble des triplets en appliquant le simple modèle de sélection (($P \parallel \mathbf{R} \parallel \mathbf{Arg2}$) OR ($\mathbf{Arg1} \parallel \mathbf{R} \parallel P$)) et en éliminant tous ceux retournés. L'application de ce simple filtre nous a permis d'éliminer 10 millions de triplets contenant des pronoms tels que ceux présentés dans le tableau III représentant environ le tiers de la taille de la collection originale.

Arg1	Relation	Arg2
on	estimer	il
il	conserver à	la France
il	vivre	il
elle	devenir	un duché de l' Empire germanique
il	aller	l' instruire de ses vérités divines
On	estimer	chaque année presque
ce	induire	de nombreux disservices tels
Elle	donner	des infos
tu	leur donner à	toutes

Tableau III – Exemples triplets contenant au moins un pronom

2.2.2 Symboles

Nous avons également constaté une grande présence de symboles n'étant pas alphanumériques à l'intérieur de la collection et bien que des triplets comme (*Julien Villeneuve == recevoir == deux contraventions de 637\$*) sont potentiellement utiles puisqu'on peut associé une certaine sémantique à ces symboles à partir du contexte, cependant d'autres triplets comme ceux ne contenant que des symboles dans au moins un des arguments ou les symboles n'étant pas reconnu par l'engin d'analyse nous sont peu utiles. Nous avons donc décidé de procéder à l'élimination de ces deux types de triplets impliquant des symboles. Pour ce faire, nous avons séparé le filtrage en deux phases : nous avons premièrement éliminé tous les triplets contenant des symboles étant hors de l'alphabet français et nous avons ensuite dans un deuxième temps éliminé ceux qui ont au moins un des champs ne contenant aucun caractère appartenant à l'ensemble alphanumérique comme ceux présents dans la table IV. Ces deux opérations ont permis de filtrer environ 3 millions de triplets bruités.

Arg 1	Relation	Arg 2
Rur@rt	&	Ars Numerica
1844	—	[17
L	= \mathbb{R} [X]/(X ² +1
Clermont-Ferrand	–	Jean Pierre Fontana
Élizabeth Margoni	***	tournage
Lawrence Metal	&	Marine Works
Architecture	<	1,000 Square

Tableau IV – Triplets contenant un champ ne contenant que des symboles

2.2.3 Fréquentiel

Enfin, nous complétons la phase de préfiltrage en appliquant un filtre fréquentiel sur les arguments présents dans la collection pour cibler tous les triplets contenant des arguments hapax tels que ceux présentés dans le tableau V. Ces arguments n'apparaissant dans aucun autre triplet, ils n'apportent que très peu d'informations lors des calculs de statistiques. De plus, ce filtre permet de capturer une bonne quantité d'erreurs d'orthographe puisque, à l'exception des cas impliquant des erreurs fréquentes et récurrentes, les arguments les contenant apparaissent généralement qu'une seule fois. Contrairement aux autres filtres précédemment présentés, nous n'éliminons pas complètement les triplets retournés par ce filtre puisque ceux-ci peuvent contenir de l'information utile qui, dans l'éventualité d'une extension de la base de connaissances avec de nouveaux triplets, ces arguments pourraient ne plus être uniques. Nous conservons donc les triplets capturés dans un second ensemble qui ne sera pas utilisé lors du calcul des statistiques.

Arg1	Relation	Arg2
<i>C.H.H.</i>	t	<i>.XIV; p. 15</i>
<i>Ch'a</i>	être utiliser pour représenter	<i>le son [t]^h</i>
<i>frontière sauvage du Tennessee</i>	y combattre	<i>l Indian Removal</i>
<i>le GR II36</i>	compter toujours	<i>4 MB174</i>
<i>le Théorème de Bernoulli</i>	démontrer sur	<i>une ligne de courant</i>
<i>termine 11e de division I</i>	rejoindre	la D3

Tableau V – Exemples triplets hapax

2.2.4 État de la base post filtrage

À la suite de toutes ces opérations de filtering le compte de triplets fut drastiquement réduit passant d'un total initial de 30.4 millions de triplets à un sous-ensemble filtré contenant 11.3 millions de triplets. Malgré cette phase de filtrage agressive, il reste encore une quantité non négligeable de bruit. En effet comme mentionné précédemment, notre phase de préfiltrage couvre uniquement la détection des triplets problématiques dont l'analyse de pertinence est relativement simple à effectuer algorithmiquement ne demandant pas une analyse de la sémantique. Il reste donc dans la collection une certaine portion de triplets étant vides de sens ou étant inexacts. Puisque ces triplets furent extraits à partir d'articles provenant de Wikipédia on pourrait s'attendre à ce que cette quantité de triplets inexacts soit négligeable, malheureusement, les outils d'extraction d'information ouverte n'étant pas encore totalement au point ils ont parfois des problèmes à identifier les bons arguments étant en relation. Par exemple, l'outil étant à l'origine de la collection utilisée au cours de ce projet a généré le triplet (Montréal == constituer == un échec financier) lorsque présenté avec la phrase "Malgré les statistiques impressionnantes, les jeux à Montréal constituaient un échec financier".

La détection de la validité de tels triplets étant plus complexe à effectuer automatiquement, nous nous tournerons vers l'utilisation des jeux pour tenter de réduire l'impact de ces triplets sur nos inférences futures.

Arg1	Relation	Arg2
Paul	souhaiter purger	sa peine
ParcExtension	45	31 39
l École	archéologique	américaine d Athènes
just	baby	renvoie
Notre Ukraine	ensuite former	une coalition
la ville	être entourer depuis	la fin du XIXe siècle
Un pirate	essayer de déchiffrer	le message
l édifice	ne ni perdre	son style
Diomedea	epomophora dans	l ordre Procellariiformes
George Martin	concert à	Los Angeles
La version japonaise du jeu	contenir moins de	50 niveaux
La forêt	dominer largement	le parc
la réalité	empirique de	l espace
MauriceYvan Sicard	journaliste de	gauche
Dieu	le aider	le garde
Bucarest	long de	111 km
folle	lui faire voir	le néant de tout
le groupe	ne pas pouvoir jouer de	concert
Le Portugal	présenter uniquement	deux athlètes
Victor Bunel	réimprimer	le Journal officiel
un programme	se exécuter avec	succès
La chanson	lauréat devenir	un tube planétaire
le 2 juin 2006	ne pas être concrétiser de	fait de son rachat
un espace de Hilbert	avoir priori non	commutative
DONG	commencer à étendre	ses activités
murs	déborder dans	le fossé afin
Louvois	faire bombarder	le fort de Landskrone
L empereur Constance II	faire enlever de	Sénat
un moine	finir par mettre à	point

Tableau VI – Échantillon aléatoire des triplets restants

2.3 Inférences

Nous nous sommes ensuite intéressés à l'enrichissement de la base de connaissances en y effectuant une série d'inférences à partir de diverses statistiques calculées sur la collection de triplets nous permettant ainsi d'étendre la base de connaissances en y ajoutant un plus grand éventail de types d'informations.

2.3.1 Classification

Une des opérations les plus courantes lorsque confronter à de grands ensembles de données est celle de la classification. En effet, cette tâche, consistant à regrouper les données jugées semblables à partir des propriétés qu'elles partagent et qui permettent de les rendre discernables du reste de la collection, est fortement attrayante puisque ce regroupement nous offre une grande puissance de généralisation permettant de se concentrer sur les catégories au lieu de chaque donnée individuelle et permet donc de faire ressortir une certaine structure de cet amas de données.

2.3.1.1 Sélection de catégories

Un aspect central d'un tel système de catégorisation est de décider comment sera déterminé l'ensemble des catégories possibles. Pour ce projet, nous avons décidé d'opter pour une approche conservant l'idée derrière l'extraction ouverte d'information et avons laissé les données parler d'elles-mêmes et définir les catégories. Nous obtenons ainsi un ensemble de catégories fortement lié à la collection sans devoir connaître au préalable

le contenu de la collection. Procéder ainsi nous offre la possibilité future d'appliquer ce processus sur différentes collections provenant de différents domaines sans devoir manuellement définir pour chacune de ces collections une liste de catégories propre au domaine de la collection. Pour atteindre cet objectif, nous avons défini le modèle de sélection suivant (**Arg1** == être == (un/lune) **catégorie**) nous permettant d'extraire un ensemble de catégories et pour chacune de ces catégories une liste d'entités **Arg1**. Ce modèle nous retournant un grand nombre de résultats, nous avons décidé de nous concentrer sur les catégories apparaissant le plus fréquemment dans cet ensemble et avons conservé les 1000 catégories plus fréquentes. Finalement, nous avons épuré cette

Catégories		
une danse	un processus	un jeu de cartes
un acteur	un site archéologique	un romancier
un compositeur	un éditeur	une organisation internationale
une comédie musicale	un personnage de fiction	une discipline
un peintre	une poétesse	une tragédie

Tableau VII – Exemples catégories retenues

Catégories	
une église catholique située	un prénom masculin désuet
un personnage du feuilleton télévisé Plus belle	une censusdesignated place
une commune suisse du canton de Fribourg	une rivière de Russie
un syndicaliste	un peintre allemand
une localité du comté de Nordland	un tournant
une voie du 16e arrondissement de Paris	une nouvelle fois

Tableau VIII – Exemples catégories éliminées

liste de catégories en procédant à l'élimination manuelle de catégories peu intéressantes telles que celles étant trop précises ou étant mals formées, nous laissant avec un total de

542 catégories jugées intéressantes.

2.3.2 Profil relationnel

Une des questions principales auxquelles notre base de connaissances tente de répondre est de déterminer, étant donné une entité ou une catégorie, quelle combinaison de relations peut la définir. Pour répondre à cette question, nous introduisons le concept de profil relationnel prenant la forme de vecteurs bi dimensionnels $(x, P(x))$ qui, étant donné un sous-ensemble E de triplets, nous donne pour chacune des relations x présentes dans ce sous-ensemble sa fréquence d'occurrence normalisée $P(x)$. Ces vecteurs représentent donc une distribution sur les relations étant présentes dans ce sous-ensemble et permettent donc, lorsque celui-ci est choisi judicieusement, d'obtenir une vue d'ensemble sur les relations associées à un terme ou une catégorie et leurs importances.

2.3.2.1 Profil de terme

Nous nous sommes premièrement intéressés au calcul des profils relationnels des termes (**Arg1**) présents dans notre collection de triplets. Le calcul d'un de ces termes se fait tout simplement en récupérant l'ensemble des triplets dans lequel ce terme est présent et en calculant ensuite le compte relatif de chaque relation présente dans ce sous-ensemble et finalement en normalisant par le nombre de triplets dans cet ensemble.

Plus formellement, soit C une collection de triplets ayant comme format $\{\mathbf{Arg1}, \mathbf{R}, \mathbf{Arg2}\}$ et soit $ArgsI$ l'ensemble des $\mathbf{Arg1}$ distincts dans C . Soit $T \in ArgsI$ le terme pour lequel nous voulons calculer le profil et soit $T_{triplets}$ l'ensemble des triplets conformes au modèle de sélection $\{T, \mathbf{R}, \mathbf{Arg2}\}$ permettant d'obtenir tous les triplets impliquant T . Soit $T_{relations} = \{rel_1, rel_2, \dots, rel_s\}$ l'ensemble des relations distinctes présentes dans l'ensemble $T_{triplets}$ et $Freq(rel_i)$ la fréquence d'occurrence de la relation rel_i . Nous obtenons ensuite pour chaque relation rel_i sa fréquence relative $FreqRel_i = \frac{Freq(rel_i)}{|T_{triplets}|}$. Nous constituons finalement le profil relationnel de T en accumulant toutes les relations rel_i et toutes leurs fréquences relatives respectives dans un vecteur pour donner $Profile(T) = \{(rel_1, FreqRel_1), (rel_2, FreqRel_2), \dots, (rel_s, FreqRel_s)\}$

Relation	%
être	3.8575
devenir	2.0771
avoir	1.1869
annoncer	1.1869
choisir	0.8902
mentionner	0.8902
signer	0.8902
déclarer	0.8902
affirmer	0.5934
confirmer	0.5934

Tableau IX – Extrait du profil relationnel du terme *Obama*

2.3.2.2 Profil de catégorie

Le calcul du profil relationnel d'une catégorie s'effectue en récupérant l'ensemble des profils relationnels étant des termes qui furent associés à cette catégorie par l'heuristique décrite précédemment en section 2.3.1.1. Ensuite, nous sommions tous ces profils ensemble et finalement nous normalisons le vecteur résultant par le nombre de profils relationnels impliqués dans cette somme. Procéder de cette façon, au lieu de récupérer les triplets associés à la catégorie, nous permet de nous assurer que chaque terme apporte la même part de contribution lors du calcul et donc que les profils de termes ayant un très grand nombre de triplets lui étant associé ne dominent pas les autres ayant beaucoup moins de triplets.

Arg 1	Relation	Arg 2
Lazarus McCoy	être	un chanteur
Ray Charles Robinson	être	un chanteur
80 ans	être	un chanteur
Dave Mustaine	être	un chanteur
Rio de Janeiro	être	un chanteur
Daniel Bélanger	être	un chanteur
Gilbert Bécaud	être	un chanteur
Arturo Navas Sosa	être	un chanteur
Johnny	être	un chanteur

Tableau X – Termes associés à la catégorie *un chanteur*

Plus formellement, soit C une collection de triplets ayant comme format $\{\mathbf{Arg1}, \mathbf{R}, \mathbf{Arg2}\}$ et soit Cat la catégorie pour laquelle nous calculons le profil et soit $Cat_{entities} = \{ent_1, ent_2, \dots, ent_s\}$ l'ensemble des entités **ent** conformes au modèle de sélection suivant $\{\mathbf{ent}, \mathbf{être}, \mathbf{Cat}\}$. Le profil relationnel de la catégorie Cat est alors donné en accumulant

tout les profils relationnel des entités ent_i et en normalisant le vecteur résultant pour donner le profil $Profile(C) = \sum_i^s \frac{Profile(ent_i)}{s}$.

Relation	%
être	37.4926
faire	2.3026
avoir	1.5854
sortir	1.2507
devenir	1.2314
enregistrer	1.1862
participer à	0.7370
quitter	0.6685
dire	0.5630
commencer	0.5558

Tableau XI – Extrait du profil relationnel de la catégorie *un chanteur*

2.3.2.3 Surreprésentation de la relation être dans les catégories

Le verbe "être" étant généralement le plus fréquent dans la grande majorité des corpus français, on peut donc s'attendre à une grande représentation au travers de la collection de la relation être. On peut effectivement observer cette tendance, cependant on remarque également que sa proportion à l'intérieur des profils de catégories est anormalement élevée. On explique cette anomalie par le biais inséré par notre heuristique de sélection de catégories, présentée en section 2.3.1.1, puisque cette dernière force chaque profil de terme étant impliqué dans le calcul du profil de catégorie à contenir au moins une occurrence de la relation être. Ainsi pour ne pas bruite les inférences qui seront faites sur ces profils, nous avons décidé d'ignorer cette relation lors des calculs subséquents impliquant ces profils, mais nous avons cependant décidé de tout de même la

conserver à l'intérieur des profils pour ne pas perdre d'information.

2.3.2.4 Profil moyen

En plus des profils relationnels de termes et de catégories, nous avons également jugé pertinent d'ajouter un profil moyen étant calculé à partir de l'ensemble des triplets présent dans la base de connaissances. Ce profil permet d'obtenir une vue d'ensemble des relations présentes dans la collection et de leur fréquence d'apparition.

Relation	%
être	16.0305
avoir	2.0139
faire	1.4473
être de	1.2085
devenir	1.1986
prendre	0.6707
être à	0.5926
compter	0.4969
pouvoir être	0.4898
posséder	0.4353

Tableau XII – Extrait du profil relationnel moyen

Nous calculons ce profil moyen ainsi :

Soit C l'ensemble des triplets dans la collection et $C_{relations} = \{rel_1, rel_2, \dots, rel_s\}$ l'ensemble des relations distinctes et $Freq(rel_i)$ la fréquence de la relation rel_i . Nous normalisons ensuite ce vecteur de relations pour obtenir $FreqRel_i = \frac{Freq(rel_i)}{|C|}$. Finalement, nous formons le profil moyen **Profil(Moyen)** = $\{(rel_1, FreqRel_1), (rel_2, FreqRel_2), \dots, (rel_s, FreqRel_s)\}$

2.3.3 Déduction des relations importantes

On remarque que les relations considérées les plus fréquentes selon notre profil moyen se retrouvent également comme les plus fréquentes d'une bonne partie des catégories. En effet, ces relations sont très générales et peuvent se retrouver dans un très grand nombre de contextes différents et nous donnent donc peu d'informations sur les relations étant plus spécifiques à la catégorie en question. Cependant lorsqu'on analyse ces profils de catégorie plus en profondeur on peut observer que ces relations plus significatives sont cependant bien présentes et ont une plus grande probabilité d'apparition au sein de la catégorie que dans le profil moyen. Nous pouvons faire émerger ces relations distinguant une catégorie de la moyenne en prenant son profil relationnel et en lui soustrayant le profil moyen et en reclassant ensuite le vecteur obtenu selon les valeurs absolues les plus élevées.

Relation	%
être	21.4621
enregistrer	1.1110
sortir	1.0412
faire	0.8552
quitter	0.4498
participer à	0.4237
interpréter	0.4200
composer	0.4164
commencer	0.3809
débuter	0.3601

Tableau XIII – Exemple de relations importantes pour la catégorie *un chanteur*

Comme on peut l'observer dans le tableau XIII les relations apparaissant les plus fréquemment après l'application de cette opération sur la catégorie "Un chanteur" semblent

plus significatives comparativement à celles avant cette opération que l'on peut observer dans le tableau XI. On peut effectivement observer que des relations telles que "composer" et "interpréter" ont monté en tête de liste pour remplacer des relations telle que "avoir" et "devenir".

2.4 Similarités

Les profils relationnels nous donnent une vue d'ensemble sur les relations et leur fréquence d'utilisation étant donné une entité ou un groupe d'entités. Il peut être intéressant de pousser cette analyse et de comparer ces profils entre eux. Ces profils relationnels possédant une structure de vecteurs normalisés représentant une distribution sur les relations, nous pouvons facilement comparer ces derniers en utilisant une des nombreuses mesures de distance existante s'appliquant à cette structure de donnée et nous avons donc décidé d'utiliser pour ce projet la mesure de similarité cosinus.

Le calcul de la similarité entre deux profils relationnels est donc effectué ainsi : Soit P_1 et P_2 les deux profils pour lesquels nous voulons calculer la similarité et soit $R = \{rel_1, rel_2, \dots, rel_s\}$ un ensemble formé en récupérant toute les relations présentes dans les profils relationnels P_1 et P_2 à l'exception de la relation "être". La similarité est donc

$$\text{donnée par } \mathbf{Sim}(P_1, P_2) = \frac{\sum_i^s P_1(rel_i)P_2(rel_i)}{\sqrt{\sum_i^s P_1(rel_i)^2} \sqrt{\sum_i^s P_2(rel_i)^2}}.$$

2.4.1 Similarités entre termes et catégories

Le premier type de similarité que nous avons calculé est celui impliquant le profil relationnel de terme et celui d'une catégorie nous permettant d'obtenir le niveau d'appartenance pour ce terme et cette catégorie et lorsque répéter pour toutes les catégories présentes dans la collection permet d'obtenir quelles sont les catégories étant les plus similaires a ce terme. Ces catégories fortement similaires peuvent nous permettre d'étendre la classification initiale effectuée précédemment en proposant le terme concerné comme candidat potentiel appartenant à ces catégories. Nous avons appliqué ce processus sur les 100 000 termes les plus fréquents pour obtenir la similarité entre chacun de ces termes et chacune des catégories.

Catégorie	%
un inventeur	44.9178
un scientifique	25.7381
un chirurgien	25.3737
un ingénieur	25.0748
un pionnier	23.0974
un entrepreneur	18.9159
un mathématicien	18.4367
une langue	18.1690
un art	18.0267
une machine	17.7711

Tableau XIV – Extrait de similarités entre le terme *Thomas Edison* et catégories

2.4.2 Similarités entre les catégories

En plus de tenter de déterminer quels sont les termes étant les plus similaires à une catégorie nous nous sommes également intéressés à trouver quelles sont les autres caté-

gories lui étant similaires. Nous avons donc calculé pour chaque catégorie son niveau de similarité avec toutes les autres catégories nous permettant d’obtenir les catégories lui étant le plus similaires. Comme on peut l’observer dans le tableau XV pour le cas de la catégorie "*un poète*", ces catégories ayant des profils relationnels similaires sont pour la plupart également similaires sémantiquement.

Catégorie	%
un poète	100
un écrivain	91.9699
un journaliste	82.4342
un auteur	78.8863
un nom de famille	78.5174
un historien	77.5534
un patronyme	77.0628
une journaliste	76.0513
un romancier	74.8200
une romancière	74.6809

Tableau XV – Extrait de similarités entre la catégorie *un poète* et l’ensemble des catégories

Il est à noter que bien que ces vecteurs de similarités possèdent quelques éléments bruités ou indésirables tels que les catégories "*un nom de famille*" et "*un patronyme*" dans l’exemple présenté, les résultats obtenus semblent nettement moins bruités que ceux provenant des similarités impliquant des termes. Nous estimons que cette hausse de qualité peut être attribuée au nombre de triplets indirectement impliqués dans la génération des profils relationnels de catégories. Ces derniers étant constitués de plusieurs profils de termes ils contiennent donc généralement un plus grand éventail de relations et une meilleure approximation de la distribution souhaitée.

2.4.3 Similarités entre termes

Nous avons également envisagé le calcul des similarités entre les différents termes, cependant les résultats préliminaires étaient fort décevants et nous avons donc décidé de ne pas inclure ce type de similarité dans la base de connaissances. Ces résultats décevants peuvent s'expliquer par la faible quantité de relations distinctes qu'un profil de terme contient comparativement au bassin de relations distinctes possibles. On peut effectivement constater que les profils de terme contiennent en moyenne 6.1 relations distinctes provenant d'un bassin total de 222 960 relations distinctes. Nous nous retrouvons donc avec des profils de terme ayant très peu de relations en commun et dont la plupart se trouvent à être les relations les plus fréquentes au travers la base de connaissances tel qu' "*avoir*" ou "*faire*". Le calcul de similarité dégénère donc dans bien des cas à la simple comparaison de la fréquence d'une ou deux relations ce qui peut expliquer les résultats décevants que nous avons observés.

2.4.4 Distributions des Arg2

Nous avons jugé qu'il pourrait être également intéressant de tenter de distinguer les différents contextes d'utilisation des relations en fonction de chaque catégorie en examinant les **Args2** utilisés ainsi que leurs fréquences d'utilisation. En effet, on pourrait s'attendre à rencontrer ces arguments dans différentes proportions dépendant de la catégorie. Par exemple, pour la relation "*écrire*" on pourrait s'attendre à rencontrer l'**Arg2** "*un livre*" plus fréquemment dans le cas de la catégorie "*un auteur*" que pour la caté-

gorie "un chanteur" pour laquelle on s'attend à rencontrer plus souvent l'argument "une chanson". Ainsi nous avons calculé la probabilité $P(\text{Arg2} \mid \mathbf{R}, \text{Cat})$ nous donnant pour chaque paire de catégories **Cat** et relation **R** une distribution sur les **Arg2** présents.

Arg 2	%
la préface	0.007272727
une lettre	0.007272727
la présence de Cruella	0.003636364
un immigrant autrichien	0.003636364
Mary Jo Nie	0.003636364
le personnage de MarieAnge	0.003636364
le maire de la NouvelleOrléans	0.003636364
Benjamin Jollivet	0.003636364

Tableau XVI – Distribution des Arg2 pour la relation *être* et la catégorie *un auteur*

Malheureusement, cette tendance ne semble généralement pas se manifester dans la grande majorité des distributions que nous avons calculées. Ces dernières semblent contenir une très grande majorité de fréquences unaires causant une grande quantité des distributions à être très proches d'une distribution uniforme et apportent donc très peu d'information utilisable à des fins de différenciation. Cette grande proportion de fréquence unaire s'explique par la très grande quantité d'arguments étant trop spécifiques rencontrant par exemple des arguments du style de "son troisième album intitulé" apparaissant qu'une seule fois dans le contexte de la catégorie "un chanteur" et de la relation "écrire".

CHAPITRE 3

GAMES WITH A PURPOSE

3.1 Jeux de production participative

La popularité des jeux vidéos a augmenté considérablement ces dernières années. Les jeux prennent de plus en plus de place dans la vie quotidienne de bien des gens. Leur facilité d'accès grandissante a contribué à attirer une portion de la population qui ne jouait pas ou très peu précédemment et a eu pour effet d'augmenter le nombre d'heures passées sur ce mode de divertissement pour une portion de joueurs. En effet, ce mode de divertissement, qui à l'époque était limité à une petite portion de la population puisqu'il nécessitait du matériel spécifique et dispendieux, est de nos jours disponible en tout temps et à la portée de main de la majorité de la population étant donné la grande quantité de jeux présents sur les téléphones intelligents et sur le web facilitant l'initiation et la consommation de ce médium. Cet engouement pour les jeux vidéos fut remarqué par la communauté scientifique et mise à profit en 2004 avec le dévoilement de **ESP** (Von Ahn et Dabbish, 2004). Ce jeu a réussi à attirer plus de 200 000 utilisateurs en leur proposant de se divertir tout en contribuant à la science. Ce dernier avait comme objectif principal d'utiliser les mécaniques de jeu pour mettre à contribution ce considérable bassin de joueurs afin de procéder à l'étiquetage d'une collection contenant plus de 50 millions d'images provenant du web.

Le grand succès de **ESP** ainsi que celui rencontré par **Verbosity** deux ans plus tard (Von Ahn et al., 2006) ayant permis la construction d'une large collection de "faits de sens commun" ont piqué l'intérêt de la communauté scientifique en démontrant l'efficacité des GWAPS (Games with a purpose) comme outil de science participative afin de réaliser certaines tâches étant jugées complexes à réaliser algorithmiquement, mais étant relativement simples à effectuer pour un utilisateur humain.

3.1.1 GWAPS dans le domaine du TALN

Le traitement de langues naturelles est l'un des premiers domaines d'études à s'être grandement intéressé à l'utilisation de jeux pour la réalisation d'une multitude de tâches. On peut effectivement constater que de nombreuses tâches du domaine bénéficient de l'inclusion de ressources annotées ou validées manuellement afin d'améliorer ou assurer une certaine qualité. Cet intérêt a donné naissance à de nombreux projets de grande envergure tels que le projet **Jeux de mots** (Lafourcade et Joubert, 2008) qui en utilisant une série de jeux a permis de bâtir un vaste réseau lexical dédié au français et contenant à ce jour plus de 90 millions de relations impliquant plus de 1.4 million termes ou encore le projet **PhrasesDetective** (Chamberlain et al., 2008) ayant mis de l'avant l'utilisation d'un jeu pour l'annotation des références anaphoriques présentes dans un ensemble de textes. Plus récemment on a pu voir la naissance de **ZombiLingo** (Fort et al., 2014) un jeu dédié à l'annotation de corpus en syntaxe de dépendances en utilisant des utilisateurs afin de valider les sorties produites par des analyseurs syntaxiques. Les projets **Voice**

Race (McGraw et al., 2009) et **Voice Scatter** (Gruenstein et al., 2009) se sont intéressés quant à eux à l'utilisation de jeux pour l'acquisition de la parole en demandant à leurs utilisateurs d'interagir avec le jeu en utilisant un microphone afin de dicter certains groupes de mots apparaissant sur l'écran de jeu permettant ainsi la création de corpus mettant en parallèle les échantillons vocaux d'utilisateurs avec les passages sous forme textuelle.

On note également le projet présenté dans (Jurgens et Navigli, 2014) combinant les résultats de deux jeux pour permettre de réaliser leur tâche. Le premier jeu nommé **Puzzle Racer** vise à créer une liste d'associations entre certains sens présents sur WordNet (Fellbaum, 1998), une ressource lexicale populaire pour la langue anglaise, et des images ayant la même sémantique. Le deuxième jeu **Ka-boom**, quant à lui, vise à procéder à la désambiguïsation sémantique et utilise les associations produites par le premier jeu pour donner une représentation visuelle aux sens possibles. Ces deux jeux ont de plus réussi à s'éloigner des interfaces principalement textuelles se retrouvant dans la grande majorité des GWAPS en TALN en remplaçant ces dernières par des interfaces graphiques plus semblables à celles présentes dans les jeux vidéos traditionnels.

3.1.2 GWAPS dans les autres domaines

Le TALN n'est pas le seul domaine ayant produit une grande quantité de jeux à des fins de recherche. En effet, on a pu constater de nombreux jeux ayant vu le jour dans le domaine de la bio-informatique et de la médecine avec des projets tels que **FOLDIT**

(Cooper et al., 2010) proposant à ses utilisateurs de contribuer à une banque de données sur le repliement des protéines en interagissant avec une interface ludique ou le jeu **Phylo** (Kawrykow et al., 2012) s'intéressant à l'alignement de séquences multiples ou également le projet **MalariaSpot** (Luengo-Oroz et al., 2012) demandant à ses utilisateurs d'identifier certains parasites dans des échantillons sanguins ou plus récemment encore le projet **Nanocrafter** (Barone et al., 2015) dédié à la découverte et la simulation d'appareils nanotechnologiques construits à partir de séquences d'ADN.

En plus de ces domaines ayant utilisé une bonne quantité de jeux comme moyen de science participative, on note quelques autres domaines ayant également brièvement exploré l'utilisation de jeux. Parmi tous ces projets, nous retenons des jeux forts intéressants tels que le jeu **PhotoCity** (Tuite et al., 2010) qui utilise les mécaniques de jeu pour diriger ses utilisateurs vers certains lieux à proximité et demande ensuite à ses utilisateurs d'utiliser leur téléphone cellulaire pour prendre en photo certains édifices sous certains angles afin d'obtenir une représentation virtuelle de ces édifices en trois dimensions. Le jeu **Urbanopoly** (Celino et al., 2012) demande également à ses utilisateurs d'interagir avec le monde non virtuel en leur posant des questionnaires à propos de lieux publics les entourant afin de récolter certaines informations sur ces endroits tels que les heures d'ouverture ou le type de cuisine dans le cas de restaurants. On note également une bonne quantité de jeux dédiés à l'identification ou la classification d'images telle que **KissKissBan** (Ho et al., 2009) reprenant l'idée derrière **ESP** et tente d'éliminer certains biais présents dans ce dernier ou encore le projet **GalaxyZoo** (Lintott et al., 2010)

dans le domaine de l'astronomie dédié à la classification d'images de galaxies dans différentes catégories ou encore dans le domaine de l'art le jeu **Karido** (Steinmayr et al., 2011) pour l'étiquetage d'images provenant d'oeuvres d'art.

3.1.3 Études sur les GWAPS

Cette grande quantité de jeux ayant vu le jour dans le domaine scientifique a également fait émerger un bon lot d'études supplémentaires analysant divers aspects liés à ceux-ci tels que les stratégies employées lors de la conception et du déploiement de ces jeux ou analysant la qualité des résultats récoltés. Dans leur étude (Sabou et al., 2013) ont dressé une comparaison entre l'utilisation de jeux et l'utilisation de services de recrutement de participants tel que **Amazon Mechanical Turks** en comparant ces deux méthodes d'acquisition de participant en les analysant sous divers aspects tels que les coûts initiaux engendrés par la mise en place ainsi que ceux étant récurrents pour chaque tâche accomplie, la qualité des données recueillies, la diversité des participants ou encore les problèmes éthiques. De leur côté (Chamberlain et al., 2013), dresse un bilan général des succès et limitations liés à l'élaboration et la promotion de jeux ayant pour but la création de ressources linguistiques. Dans leur article (Siu et al., 2014) effectuent une comparaison entre les jeux dits coopératifs et ceux compétitifs en analysant ces derniers sur divers aspects tels que l'exactitude des résultats ou encore la popularité de ces derniers.

3.2 Conception de WikiGames

Nous appuyant sur ces travaux antérieurs ayant démontré le potentiel de l'utilisation de jeux pour la réalisation de tâches de grande ampleur, nous avons amorcé la conception de la plateforme **WikiGames** ayant pour but la validation des connaissances résultant des inférences que nous avons effectuées sur la collection de triplets. La conception d'une telle plateforme nécessite une certaine planification afin de déterminer certains aspects entourant le développement et demande que l'on se questionne sur de nombreux points. Ces questionnements passent de sujets techniques tels que le choix de l'outil de développement ou également déterminer comment sera accédée la plateforme et à partir de quel type d'appareil électronique ou encore des questionnements plus ludiques tels que déterminer comment rendre la tâche amusante à réaliser. Dans cette section nous tenterons de répondre à ces questions en dévoilant les principes de conception et choix majeurs ayant influencé le développement de la plateforme **WikiGames**.

3.2.1 Principes de conceptions

3.2.1.1 Facilité d'extension

Un des défauts majeurs de l'utilisation de jeux à des fins de science citoyenne comparativement à des alternatives telles que **Amazon mechanical turks** est que le temps nécessaire pour la conception et la mise en place de ces jeux est nettement supérieur et entraîne donc indirectement des coûts initiaux. Fort heureusement, ces coûts peuvent être

amortis par une grande quantité de tâches effectuées par l'entremise des jeux étant donné les coûts presque nuls une fois les jeux déployés. Pour tenter d'amortir encore davantage ce temps investi, nous avons porté une attention particulière lors de la conception à une prise de décision permettant de s'assurer qu'une extension future de la plateforme soit possible et relativement simple à effectuer.

3.2.1.2 Expérience utilisateur

Le succès de l'utilisation de jeux à des fins de science citoyenne repose en grande partie sur ses joueurs et plus particulièrement sur la quantité et la qualité de leurs contributions. Comme le font remarquer (Chamberlain et al., 2013) pour réussir à obtenir une grande quantité de contributions de qualité il n'est pas suffisant de seulement attirer les joueurs vers la plateforme, il faut ensuite également s'assurer de maintenir ces nouveaux joueurs intéressés après leur introduction aux jeux afin qu'ils jouent de multiples parties. En effet, ces joueurs récurrents permettent de non seulement réduire les efforts et montants dédiés à la publicité des jeux, mais permettent également d'obtenir des contributions de qualité supérieure puisqu'à mesure qu'ils accumulent des parties ces joueurs actifs tendent à devenir meilleurs pour la réalisation des tâches qui leur sont demandées. La conception de **WikiGames** fut donc grandement guidée par le désir d'offrir une expérience agréable et de qualité à ses joueurs. Pour atteindre ce but, nous nous sommes appuyés sur les critères contribuant au plaisir énoncé par le modèle **GameFlow** (Sweetser et Wyeth, 2005) et avons incorporé certaines de leurs recommandations. Ce modèle

basé sur le modèle plus général **Flow** (Csikszentmihalyi, 1990) adapte ce dernier pour qu'il s'applique aux jeux vidéos et identifie huit éléments principaux étant liés au plaisir des joueurs et pour chacun de ces éléments dresse une liste de critères et recommandations permettant de s'assurer que ces éléments sont bien présents dans le jeu.

3.2.1.3 Jeux à usages multiples

En examinant la grande quantité de jeux ayant une vocation scientifique, nous avons constaté qu'une grande majorité de ceux-ci sont développés pour répondre à une unique tâche bien particulière lui étant associée et est généralement le seul jeu associé à la réalisation de cette tâche. Nous croyons qu'il est pourtant faux de croire que tous les joueurs sont attirés par les mêmes genres de jeux. De plus, étant donné le temps investi au développement d'un jeu nous estimons qu'il pourrait être avantageux de réutiliser un jeu pour la réalisation de plus d'une tâche.

Nous nous éloignons donc du paradigme usuel du "Jeu ayant un but" mettant en relation un jeu et son problème et explorons l'utilisation de jeux ayant un format d'entrée commun pour obtenir des "Jeux avec buts multiples" mettant en relation une quantité arbitraire de problèmes avec une autre quantité arbitraire de jeux.

3.2.2 Questionnaires associatifs

Pour réussir à faire le pont entre les différents types de données que nous désirons faire valider avec les jeux et faciliter le développement de ces derniers, nous introduisons

une structure abstraite intermédiaire que nous appellerons "Questionnaire associatif". Ces structures sont composées de deux ensembles : le premier étant un ensemble E d'éléments à associer et le second un ensemble A de valeurs d'assignation possibles pour ces éléments. Ces deux ensembles initialement vides seront dynamiquement peuplés lors de l'initialisation d'une session de jeu en interrogeant la base de connaissances à partir de requêtes et paramètres propres à chaque type de questionnaire et ciblant différentes parties de la base de connaissances.

Nous séparons ainsi la tâche de validation en deux sous tâches : génération de questionnaires associatifs à partir des données que nous voulons valider et ensuite création de jeux utilisant ces questionnaires. Procéder ainsi nous permet d'ajouter une certaine indépendance entre les jeux et les données qu'ils visent à valider et permettra dans un premier temps de facilement ajouter et interchanger les jeux utilisant ce format et pourrait éventuellement permettre de réutiliser ces jeux pour la validation d'ensembles de données complètement différents et réduirait la tâche à la définition de règles générant des questionnaires associatifs à partir de ces données.

Pour procéder à la validation des inférences effectuées sur la collection de triplets, nous avons amorcé la plateforme à partir de trois types de questionnaires associatifs ciblant chacun une portion différente de la base de connaissances et appliquant leur ensemble de règles respectifs pour peupler les ensembles A et E .

3.2.2.1 Questionnaire d'assignation de catégorie

L'objectif de ce type de questionnaire est de s'assurer que les utilisateurs partagent le même avis que nos vecteurs de similarités sur les catégories étant considérées les plus significatives pour chaque entité. Nous tentons d'atteindre cet objectif en peuplant l'ensemble E avec des entités et demande à l'utilisateur d'assigner chacune d'elle à la catégorie qu'il considère la plus significative présente dans l'ensemble A . Procéder à la génération de ces deux ensembles par sélection entièrement aléatoire en sélectionnant parmi toutes les entités et catégories présentes dans la base de connaissances est possible, mais malgré que cette méthode permet d'obtenir l'avis d'utilisateur pour chaque paire d'entité et de catégorie elle a également comme effet secondaire de créer un ensemble de questionnaires dont la plupart des questionnaires présents contiennent un ensemble E dont toutes les entités ne semblent avoir aucun bon candidat d'assignation dans A . Afin de ne pas faire fuir nos joueurs et les incommoder par ce flot de questionnaires que nous jugeons peu intéressants, nous avons adopté un ensemble de règles de génération tentant de faire émerger des ensembles A et E contenant une grande quantité de liens entre les éléments y étant présents. De plus, afin de ne pas forcer le joueur à devoir choisir une assignation pour une entité lorsqu'il ne connaît pas cette dernière ou lorsqu'il considère qu'aucune des assignations proposées dans le questionnaire n'est bonne, nous ajoutons également les choix "*Ne sais pas*" et "*Autre*" à l'ensemble A .

Les paramètres et règles de génération pour ce type de questionnaire sont les suivantes :

Paramètres	Description
C_1 (default="")	Catégorie initiale
R (default=3)	Nombre d'itérations
X (default=6)	Quantité minimale d'entités
Z (default=3)	Quantité minimale de catégories
Y (default=3)	Profondeur maximale pour l'acquisition d'entités similaires à une catégorie
Q (default=3)	Profondeur maximale pour l'acquisition de catégories similaires a une entité

Tableau XVII – Paramètres pour les questionnaires d'assignation de catégories

1. Si le paramètre de catégorie initiale C_1 n'est pas fourni l'initialiser aléatoirement à partir de l'ensemble des catégories et l'ajouter à A
2. Tirer aléatoirement $\lceil \frac{X}{R} \rceil$ entités parmi les Y étant considéré les plus similaires à la dernière catégorie insérée dans A et les ajouter à E
3. Tirer aléatoirement $\lceil \frac{Z}{R} \rceil$ catégories parmi les Q étant considéré les plus similaires à la dernière entité insérée dans E et les ajouter à A
4. Répéter les étapes 2 et 3 R itérations
5. Ajouter à la liste d'assignations possibles A les choix "*Ne sais pas*" et "*Autre*"

Ensemble E	Ensemble A
Michael Jackson	Un auteurcompositeur
Chris Brown	Un chanteur
Madonna	Un groupe
le duo	Autre
Cohen	Ne sais pas
Britney Spears	

Tableau XVIII – Exemple de sortie pour un questionnaire d'assignation de catégories

3.2.2.2 Questionnaire profil relationnel de terme

Ces questionnaires contiennent dans l'ensemble E des relations tirées d'un profil relationnel associé à un terme et contiennent dans l'ensemble d'assignation A un ensemble de valeurs de pertinence servant à l'évaluation des relations dans E par rapport au terme.

Paramètres	Description
T (défault="")	Terme à évaluer
V (défault=(Bon, Neutre, Mauvais))	Ensemble de valeurs de pertinence
Z (défault=6)	Quantité de relations à tirer
Y (défault=15)	Profondeur maximale pour l'exploration du profil relationnel

Tableau XIX – Paramètres pour les questionnaires profil relationnel de terme

Ensemble E	Ensemble A
annoncer	Bon
faire	Neutre
visiter	Mauvais
annoncer à	Autre
déclarer	Ne sais pas
prononcer	

Tableau XX – Exemple de sortie pour un questionnaire associé au terme "Charles de Gaulle"

Processus de génération des questionnaires :

1. Si T n'est pas spécifié choisir aléatoirement un terme
2. Calculer le profil relationnel du terme
3. Récupérer les Y relations les plus fréquentes dans le profil relationnel et ensuite tirer aléatoirement Z relations parmi celles-ci pour former E

4. Récupérer le paramètre V fourni en entrée contenant les valeurs de pertinence qui seront disponibles pour valider les relations dans E et y ajouter les valeurs "Ne sais pas" et "Autre" pour former A

3.2.2.3 Questionnaire de profil relationnel de catégorie

Ces questionnaires ont un objectif semblable à ceux dédiés au profil relationnel de terme étant également dédié à l'évaluation de la pertinence d'un ensemble de relations, mais utilisent des règles de génération légèrement différentes pour la sélection de ces relations. On note premièrement que, contrairement à ceux dédiés à l'évaluation de profil de terme, les profils relationnels de catégories sont déjà présents dans la base de connaissances au lieu d'être calculés. De plus, comme mentionné précédemment, les profils de catégorie ont généralement comme relations fréquentes des relations étant très générales et peu discriminantes. Nous faisons donc appel à la méthode énoncée précédemment dans la section 2.3.3 tentant de faire émerger les relations jugées discriminantes pour ce profil consistant à soustraire de ce dernier le profil relationnel moyen.

Paramètres	Description
C (default="")	Catégorie à évaluer
V (default=(Bon, Neutre, Mauvais))	Ensemble de valeurs de pertinence
Z (default=6)	Quantité de relations à tirer
Y (default=15)	Profondeur maximale pour l'exploration du profil relationnel

Tableau XXI – Paramètres pour les questionnaires profil relationnel de catégorie

Processus de génération des questionnaires :

1. Si C n'est pas spécifié choisir aléatoirement une catégorie.
2. Récupérer le profil de C et y soustraire le profil moyen et conserver les Y relations les plus discriminantes
3. Tirer de cet ensemble Z relations et les ajouter a E
4. Récupérer le paramètre V fourni en entrée contenant les valeurs de pertinence qui seront disponibles pour valider les relations dans E et y ajouter les valeurs "Ne sais pas" et "Autre" pour former A

Ensemble E
être situer à
se situer à
se composer de
être ériger à
avoir
être inaugurer

Ensemble A
Bon
Neutre
Mauvais
Autre
Ne sais pas

Tableau XXII – Exemple de sortie pour un questionnaire associé à "un monument"

3.2.3 Traitement de joueurs adversatif

Un des problèmes étant commun à l'ensemble des techniques de production participative faisant appel à l'utilisation du public est la présence, parmi cette masse de gens, d'utilisateurs adversatifs qui volontairement ne respectent pas les instructions et risquent de bruiteer les résultats. En effet, même si dans le cas de l'utilisation de jeux nous n'avons pas à traiter le cas d'utilisateurs tentant de maximiser leurs gains monétaires en répondant n'importe quoi le plus rapidement possible, nous devons tout de même considé-

rer les utilisateurs malicieux cherchant à vandaliser la base de connaissances ainsi que les joueurs curieux testant le système afin d'observer les répercussions de soumettre de mauvaises réponses. Pour traiter ces joueurs nous faisons appel à deux stratégies : une préventive visant à détecter et éliminer les réponses de ces joueurs avant qu'elles ne soient enregistrées dans la base de connaissances et une corrective visant à faciliter l'élimination a posteriori des réponses provenant de joueurs malicieux ayant déjoué la phase préventive.

3.2.3.1 Phase d'entraînement des joueurs

Pour la stratégie préventive, nous nous sommes inspirés d'une stratégie utilisée par d'autres jeux tels que **PhrasesDetectives** (Chamberlain et al., 2008) en attribuant initialement à nos joueurs le statut "en entraînement". Pendant cette phase initiale, les questionnaires qui sont utilisés par les jeux ne sont pas générés en interrogeant dynamiquement la base de connaissances, mais proviennent plutôt d'un petit ensemble de questionnaires générés manuellement et pour lesquels nous encodons également un ensemble de réponses idéales permettant de comparer celles fournies par l'utilisateur. Pour évaluer un joueur nous lui attribuons un score d'entraînement qui est augmenté lorsqu'il fournit une bonne réponse et est réduit pour chaque mauvaise réponse produite. L'utilisateur quitte donc ce statut d'entraînement lorsque ce score atteint un certain seuil lui donnant alors accès au plein potentiel de génération de questionnaires. Si l'utilisateur épuise cependant toutes les questions d'entraînement disponibles pour ce type de ques-

tionnaire, nous mettons fin à sa session de jeu et prenons tout de même la peine de le remercier pour son intérêt pour le projet. Cette phase en plus de laisser à l'utilisateur le temps de bien comprendre les instructions et de se familiariser avec l'interface avant que ses réponses ne soient considérées, agit également comme barrière tentant de bloquer l'accès aux utilisateurs malveillants vandalisant volontairement la base de connaissances par l'introduction de réponses erronées. De plus, nous estimons que cette phase a le potentiel de laisser une meilleure première impression à l'utilisateur puisque l'on peut facilement s'assurer d'une certaine qualité pour les questionnaires étant manuellement générés contrairement à ceux générés dynamiquement.

3.2.3.2 Profilage par métadonnées

Bien que notre stratégie préventive nous offre une certaine protection élémentaire elle n'est pas suffisante car facilement déjouable par un utilisateur étant conscient de l'existence d'une telle mesure. De plus, certains utilisateurs malicieux ignorant qu'ils sont testés risquent de tout de même de passer la phase d'entraînement. En effet, comme mentionné dans (Gadiraju et al., 2015) une portion des utilisateurs malicieux offrent initialement des réponses valides et après un certain point ce comportement bascule et leur intention malicieuse se révèle. Pour tenter de faciliter l'élimination des parties provenant de ces joueurs nous ajoutons un ensemble de métadonnées qui, malgré l'absence de création d'un compte par l'utilisateur, tente de regrouper et analyser les parties provenant d'un même utilisateur par l'entremise de métadonnées telles que son adresse IP

ainsi que l'heure marquant le début et fin de chaque partie ou la durée moyenne de ses parties permettant de détecter par exemple les utilisateurs répondant anormalement vite et ne prenant donc pas le temps de répondre convenablement.

3.2.4 Choix de conceptions

3.2.4.1 Compte utilisateur

Exiger la création d'un compte utilisateur pour chaque utilisateur peut permettre un suivi plus précis des parties que celui-ci effectue et de ces habitudes de jeu. Cependant, une telle exigence peut également faire fuir certaines personnes étant curieuses par le projet, mais ne désirant pas passer par le processus d'inscription et donner de potentielles informations personnelles telles que leur adresse email. Nous avons donc décidé de ne pas inclure un tel système d'authentification afin de potentiellement obtenir plus de joueurs.

3.2.4.2 Engin de jeu

Pour l'implémentation des jeux, nous avons opté pour l'utilisation de l'outil de développement de jeu **Unity3d**¹. Cet outil inclus de nombreuses fonctionnalités qui nous ont permis d'accélérer ce processus tel que la visualisation en temps réel des changements ainsi que l'inclusion d'un moteur physique sans oublier la possibilité de déployer les jeux vers de multiples plateformes.

¹<https://unity3d.com/>

3.2.4.3 Application web

Afin d'augmenter la masse de joueurs potentiels, il est important de s'assurer que la plateforme soit facile d'accès et s'assurer qu'elle soit disponible pour la majorité des gens en supportant une grande variété de configuration matérielle. Nous avons donc développé la plateforme **WikiGames** comme une application web étant disponible en accédant à la page web² dédiée à partir d'un fureteur internet ayant implémenté de la technologie **WebGL**. Cette interface, étant de nos jours présente de manière native dans la majorité des fureteurs, nous permet d'offrir à l'utilisateur des jeux ayant le potentiel d'être riche graphiquement directement à même une page web sans devoir télécharger de plug-ins supplémentaires.

Il est à noter que, bien que les jeux ont été développés pour être contrôlés initialement par l'utilisation d'une souris et un clavier, une attention particulière fut accordée lors de l'implémentation des mécanismes de contrôles de chacun des jeux pour qu'ils soient facilement transposables sur des écrans tactiles dans l'éventualité de la réalisation d'une application mobile.

²<http://rali.iro.umontreal.ca/rali/en/oie-kb-using-gwap>

3.3 Plateforme WikiGames

Ayant établi les lignes directrices et accompagnées par les choix de conception qui en ont découlé telles que l'utilisation de questionnaires associatifs comme format d'entrée commun, nous avons commencé le développement de **WikiGames** et nous présentons ici les jeux que nous avons mis à la disposition de nos utilisateurs ainsi que les divers systèmes et interfaces permettant d'interagir avec ces derniers.

3.3.1 Structure générale des jeux

La création d'un jeu est un processus relativement complexe. Il est donc facile de se perdre dans les innombrables possibilités d'implémentation. Ayant l'ambition de peupler la plateforme **WikiGames** avec initialement trois jeux et pour faciliter le développement de jeux associatifs futurs, il nous a apparu judicieux d'implémenter une structure générale de conception exploitant le format de nos questionnaires associatifs afin de simplifier le processus de création et ainsi réduire la complexité de concevoir un nouveau jeu à la réalisation de quatre sous-taches :

1. Choisir l'endroit où seront générées les entités présentes dans l'ensemble E
2. Décider où seront générées les zones d'assignations présentes dans A
3. Choisir et implémenter les différentes mécaniques de jeu permettant à l'utilisateur de manoeuvrer les entités vers les zones d'assignation

4. Décorer le jeu en définissant pour chaque composante présente et événement déclenchés au cours d'une partie certaines propriétés audiovisuelles telles que : son, couleur, image, transparence, animation, etc.

3.3.2 Écran d'accueil

Lorsque l'utilisateur démarre l'application, il est accueilli par un écran d'accueil tel que celui présenté dans la figure 3.1 lui permettant de choisir à quel jeu il désire jouer ainsi que le type de donnée qui sera validé par ce jeu en sélectionnant quel type de questionnaires sera utilisé lors de la génération du jeu.



Figure 3.1 – Écran d'accueil WikiGames

3.3.3 Simple

Le premier jeu que nous présentons est le jeu Simple illustré dans la figure 3.2 demandant à l'utilisateur de faire glisser les entités générées dans la partie centrale de l'écran vers les zones d'assignation générées dans le bas de l'écran. Puisque l'entité que l'utilisateur manipule réagit aux collisions avec les autres entités et les propulse avec une force proportionnelle à celle de l'impact, il doit être vigilant lors de ces manipulations pour ne pas propulser des entités vers les mauvaises zones d'assignations. Ce jeu, comparativement aux autres que nous présenterons, a l'avantage d'être très simple à apprendre et risque de plaire à un public étant plus intéressé à contribuer à améliorer la base que par son côté ludique. De plus, le niveau de difficulté du jeu étant relativement bas, nous estimons que les erreurs de manipulation devraient être nettement inférieures.

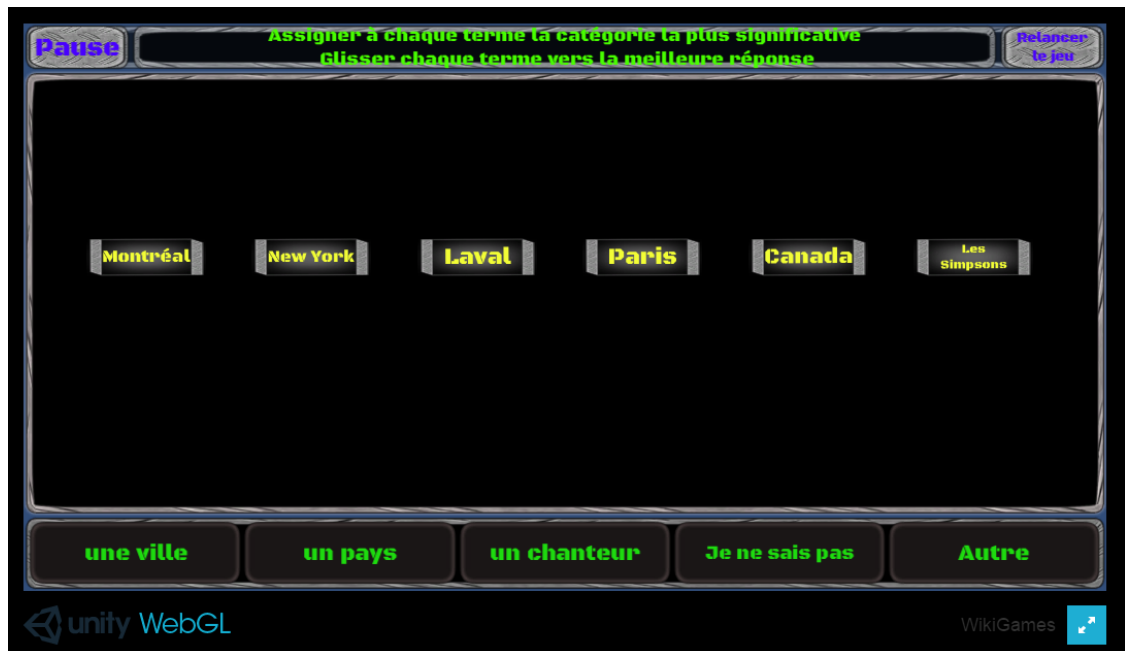


Figure 3.2 – Jeu Simple

3.3.4 Pinball

Ce jeu présenté dans la figure 3.3 s'inspire des machines à boules en générant les termes dans le coin supérieur droit sous la forme de balle que l'utilisateur, en utilisant les palettes et s'aidant des divers éléments interactifs présents sur l'espace de jeu, doit manipuler vers la zone d'assignation qu'il considère la plus appropriée pour l'entité étant représentée par la balle en jeu. La difficulté de ce jeu risque d'occasionner une plus grande proportion d'erreurs dues à de mauvaises manipulations. Nous espérons cependant que la phase d'entraînement plus longue résultant de ces erreurs permettra aux joueurs d'aiguiser leurs habilités pour que ces erreurs, si non signalées comme erreurs dans l'écran de fin de partie, soient suffisamment rares pour se perdre dans la masse de réponses.



Figure 3.3 – Jeu Pinball

De plus, cette difficulté supplémentaire a le potentiel d'attirer les joueurs plus tradi-

tionnels pour qui le désir d'aider la science est secondaire à celui d'être diverti. En effet, comme mentionné dans le modèle **Gameflow** (Sweetser et Wyeth, 2005), un des facteurs importants impliqués dans le divertissement qu'apportent les jeux est la rencontre de situations étant suffisamment difficiles pour poser un défi au joueur et lui permettre de tester ses habilités en lui donnant la possibilité de les améliorer.

3.3.5 Élimination

Comme on peut l'observer dans la figure 3.4, dans ce jeu les entités présentes dans l'ensemble E sont attachées à des ballons et après avoir été générées volent progressivement vers la droite de l'écran. La tâche de l'utilisateur est alors de déterminer parmi ces entités celles qu'il considère comme étant intruses et d'éliminer ces dernières en cliquant sur leurs ballons respectifs pour les faire tomber avant qu'elles n'atteignent la droite de l'écran et ne soit assignées comme étant valides.

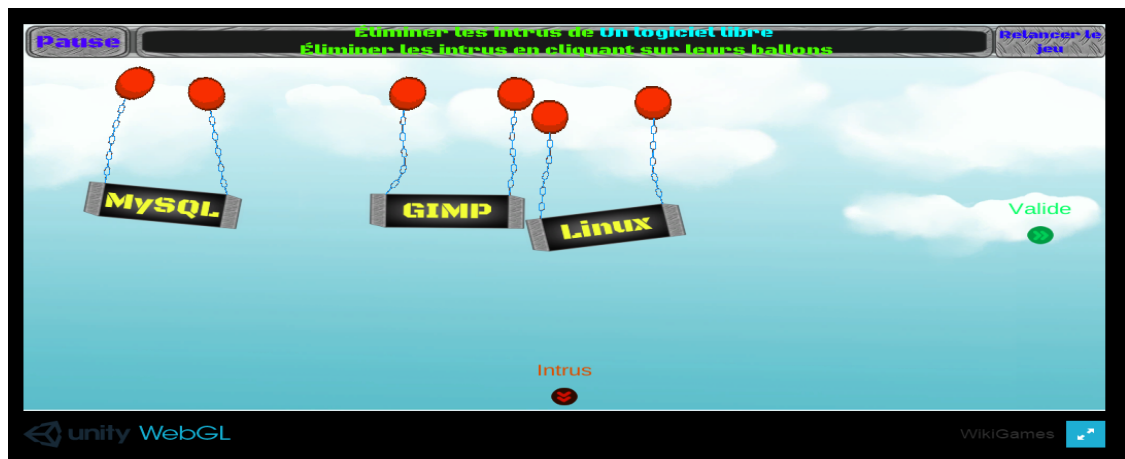


Figure 3.4 – Jeu Élimination

Ce jeu a comme avantage sa simplicité ne demandant pas une grande dextérité et

n'offrant à l'utilisateur que deux options de réponse soit valide ou invalide rendant le processus de décision pour l'utilisateur plus simple.

3.3.6 Écran de fin de partie

Une fois la partie terminée, il se fait transférer vers l'écran de fin de partie telle que celle présente dans la figure 3.5. Cet écran affiche un récapitulatif des assignations qui furent effectuées lors de la dernière partie et invite l'utilisateur à confirmer ses réponses en éliminant celles ayant été assignées par inadvertance lors de la partie. Afin de conserver une trace de ces réponses éliminées, nous conservons à la fois l'ensemble des réponses originales de l'utilisateur contenant les assignations erronées ainsi que le sous-ensemble de réponses validées. Conserver ces deux ensembles de réponses peut nous permettre d'obtenir une meilleure évaluation des habitudes de jeu de nos utilisateurs et de cibler certains de ces joueurs ayant un comportement indésirable tel que les joueurs ne corrigeant pas les assignations que nous savons erronées (car provenant de questions d'entraînement) ou encore de cibler les joueurs ayant un comportement suspect ne corrigeant jamais aucune des réponses qu'ils fournissent. De plus, puisqu'une bonne portion des joueurs tendent à fermer directement l'application lors de la dernière partie qu'ils jouent au cours d'une session, nous n'obtenons que très rarement une validation pour les dernières parties effectuées. Ainsi si nous ne conservions que les parties validées nous perdrons l'information contenue dans ces dernières parties et puisque les premières parties qu'un joueur joue sont en fait des parties d'entraînement ces dernières

parties éliminées constituent une bonne portion des parties utiles à des fins de validation.



Figure 3.5 – Écran fin de partie

CHAPITRE 4

EXPÉRIENCES

4.1 Dévoilement des jeux

La plateforme de jeux fut mise en ligne le 2 avril 2017 et deux jours plus tard fut dévoilée lors d'une présentation à l'atelier Games4NLP 2017 durant laquelle nous avons recueilli un lot de commentaires positifs, suggestions d'améliorations et travaux futurs. Malheureusement, malgré l'intérêt que semble avoir provoqué la présentation, nous n'avons pas enregistré de nouvelle partie dans les heures qui ont suivi cette présentation probablement dû au fait que la conférence était donnée en anglais devant un public ne parlant pas nécessairement français tandis que nos jeux valident des données en français. Quelques heures plus tard, nous avons fait une deuxième annonce en faisant parvenir un email aux membres présents sur une liste de diffusion liée aux séminaires RALI-OLST.

4.2 Données récoltées pour chaque partie

Comme mentionné brièvement, en plus de collecter les assignations que l'utilisateur effectue au cours d'une partie, nous récoltons également un lot de métadonnées entourant ces parties. En effet, dès le moment que l'utilisateur démarre l'application nous lui assignons un identificateur unique de session *sessionID* ayant pour rôle le regroupement

de l'ensemble des parties consécutives jouées par ce joueur jusqu'au moment où il ferme l'application. Chacune de ces parties reçoit ensuite à leur tour un identificateur unique *gameID* pour regrouper les résultats et métadonnées entourant cette partie. Nous maintenons la structure de triplet adoptée depuis le début du projet en conservant ces données reliées aux parties sous le format { *gameID :gameID* , *property :Nom_de_propriété* , *Valeur* } ayant comme nom de propriété possible un parmi ceux décrits dans le tableau XXIII. Pour ce qui est des assignations générées lors des parties, nous faisons intervenir des noeuds intermédiaires permettant de contenir des résultats ayant un format de quadruplet { *gameID :gameID* *property :result* [*entité assignation*] } à l'intérieur de deux triplets { *gameID :gameID* , *property :result* , *nodeID ://nodeID* } et { *nodeID ://nodeID* , *entité* , *assignation* } .

Propriété	Valeur
gameID	I.D unique de partie
sessionID	I.D unique de session
playerHost	Adresse i.p de l'utilisateur
version	Version de Wikigames associé a la partie
gameStartTime	Date et heure du début de la partie
gameEndTime	Date et heure de la fin de la partie
gameType	Type de jeu associé a la partie
gameDataType	Type de données associées a la partie
catInGame	Catégorie présente dans la partie
result	Noeud intermédiaire représentant une assignation

Tableau XXIII – Description des données recueillies pour chaque partie

4.3 Analyse des résultats

Dans cette section, nous présentons les résultats recueillis à partir de notre plateforme. Nous présenterons premièrement une analyse quantitative de ses résultats et ensuite nous analyserons ces derniers selon un aspect qualitatif.

4.3.1 Analyse quantitative des résultats

4.3.1.1 Statistiques sur l'utilisation de la plateforme

L'annonce effectuée sur la liste de diffusion RALI-OLST a permis d'attirer, dans le mois qui a suivi, 19 joueurs (uniques), qui comme on peut l'observer dans la table XXIV ont participé à 51 sessions de jeux ayant résulté à un total de 306 parties ayant permis d'accumuler un total de 1617 assignations d'entités. Il est cependant important de considérer qu'environ le tiers de ces parties étaient dédiées à l'entraînement des joueurs et comportait donc des assignations dont nous connaissons déjà les réponses, et rend donc ces parties de contrôles et les assignations qu'elles ont générées peu utiles pour valider la base de connaissances. Cependant, puisque l'utilisateur ignore quelles parties sont des parties d'entraînement et celles qui ne le sont pas, nous considérons que ces parties peuvent tout de même être utilisées lors du calcul des statistiques liées à la popularité des jeux.

Statistiques	
Joueurs	19
Parties jouées	306
Sessions de jeu	51
Assignations	1617
Parties d'entraînement	114
Parties utiles	192
Assignations utiles	974

Tableau XXIV – Statistique sur les parties jouées

4.3.1.2 Statistiques sur les paramètres choisis par le joueur

En plus du nombre total de parties effectuées, il est également intéressant d'observer comment sont distribuées ces parties en fonction du type de questionnaires XXV et de jeux XXVI qui furent choisies par les utilisateurs. On remarque à partir de ces statistiques que la plupart des joueurs ont choisi d'utiliser les paramètres par défaut et seulement un petit ensemble de ces joueurs a exploré les autres types de questionnaires et jeux.

Type de questionnaire	Sessions	Parties	Joueurs	Assignations
Classification	43	182	16	950
Profil de terme	9	57	8	319
Profil de catégorie	12	67	6	348

Tableau XXV – Statistiques sur les questionnaires

Type de jeu	Sessions	Parties	Joueurs	Assignations
Simple	48	251	19	1343
Pinball	9	27	6	125
Balloon Elimination	11	28	8	149

Tableau XXVI – Statistiques sur les jeux

4.3.1.3 Temps passé sur les jeux

Puisque nous n'avons pas directement enregistré l'information concernant la durée d'une partie, nous utilisons les métadonnées associées à l'heure de début et fin de la partie pour tenter de la calculer. Malheureusement bien que cette méthode permet généralement d'obtenir une très bonne approximation du temps qu'un joueur passe à jouer une partie, elle inclut également le temps où l'utilisateur a mis le jeu en pause. Bien que nous estimons ce temps supplémentaire négligeable ou nul dans la grande majorité des parties, nous observons tout de même dix parties ayant des durées anormalement élevées que nous attribuons à ce biais. Ces dix parties ayant des durées de plus de 10 minutes, dont une approchant même 18 heures, risque de fortement biaiser le calcul des statistiques les impliquant. En effet, comme on peut observer dans la table XXVIII, le temps moyen d'une partie étant de 48 secondes est nettement inférieur à la durée de ces parties, nous avons donc ignoré ces parties et leurs assignations lors des calculs impliquant la durée des parties.

Temps(HH :MM :SS)	Global	Basic	Pinball	Elimination
Total	03 :57 :00	02 :56 :06	00 :19 :50	00 :41 :04
Par joueur	00 :12 :28	00 :09 :16	00 :03 :18	00 :05 :08
Par partie	00 :00 :48	00 :00 :43	00 :00 :44	00 :01 :28

Tableau XXVII – Temps passé sur les jeux

4.3.1.4 Débit de production

Une des statistiques fréquemment utilisées pour l'évaluation de la performance d'un GWAP est son débit de production (Von Ahn et Dabbish, 2008) représentant le nombre d'annotations effectuées par heure-personne. Cette statistique permet de simplifier la comparaison de la performance de différents GWAPS en réduisant l'impact que pourrait avoir la publicité liée au jeu et le temps passé depuis le lancement du jeu. En calculant cette statistique pour nos jeux, on remarque que notre plateforme obtient en moyenne 409 annotations par heure-personne. Pour ce qui est de la performance individuelle des jeux, on remarque que Pinball et Basic obtiennent un débit s'approchant de celui qui fut déclaré par **Phrases Detective** (Chamberlain et al., 2013) tandis que notre jeu Élimination obtient un débit nettement inférieur ne permettant d'assigner qu'environ la moitié d'entité pour le même nombre d'heures-personnes. Ce résultat peut être surprenant lorsque l'on considère que la charge cognitive de ce jeu semble être inférieure aux deux autres puisque l'utilisateur ne possède que deux choix de réponse possible (*Valide*, *Non valide*). On pourrait donc s'attendre à ce que l'utilisateur passe moins de temps à réfléchir et qu'il produise au moins un nombre similaire d'annotations par heure. Il faut cependant prendre en considération que contrairement aux deux autres jeux, l'utilisateur peut difficilement influencer la durée du jeu puisque les énoncés se déplacent automatiquement et sont générés à un intervalle fixe. Augmenter la vitesse de déplacement et de génération des entités afin de réduire l'écart entre ces débits seraient donc à première vue souhaitable, cependant puisque différents énoncés et différents joueurs peuvent né-

cessiter un temps de réflexion étant également différent, il est donc difficile de trouver une vitesse appropriée pour toutes les situations.

WikiGames	Basic	Pinball	Elimination	ESP	Phrases Detective	Jeux de mots
409	441	378	200	233	450	648

Tableau XXVIII – Annotations par heure

4.3.2 Analyse qualitative des résultats

En plus d’analyser les jeux et leur capacité à récolter des assignations, il est également important de s’intéresser à la qualité de ces résultats pour s’assurer que les informations recueillies puissent être utilisables. Bien que la quantité d’annotations que nous avons accumulée est pour le moment nettement insuffisante pour procéder à la validation de la base de connaissances, nous analysons les données recueillies jusqu’à présent pour démontrer le potentiel de la plateforme.

4.3.2.1 Assignment d’entités

Pour obtenir des résultats de qualité et éviter le biais lié à l’opinion d’un joueur et afin d’également réduire l’effet que pourraient avoir les mauvaises manipulations, il est préférable d’obtenir l’avis de plusieurs utilisateurs. Cependant dues à l’énorme quantité de données à valider comparativement au nombre d’assignations que nous avons récoltées jusqu’à présent, nous n’avons qu’une faible portion d’assignations contenant plus d’un exemplaire. De plus, une portion de ces résultats sont des assignations d’entités vers

la catégorie "Autres" et malgré le fait que ces dernières peuvent être utiles pour déduire les catégories considérées invalides pour cette entité en observant les autres catégories présentes, elles nous informent cependant très peu sur les assignations idéales pour ces entités. Nous avons donc dû être très cléments quant à la fréquence d'apparition d'une réponse pour que nous la considérions comme valide. Nous avons donc accepté comme étant valide toute assignation apparaissant au moins deux fois et n'étant pas contredite par d'autres joueurs ayant choisi comme réponse la catégorie "Autres".

Entité	Catégorie assignée
Voltaire	un écrivain
Paris	un nom de lieu
Paul	un prénom
Smith	un nom de famille
Whisper	un mot anglais
PCI	un organisme
L'avenue du général	une voie
Xavier Dolan	un réalisateur
L'artiste	un peintre
L'église	un monument
Mozart	un compositeur

Tableau XXIX – Exemples d'assignations validées

Entité	Catégorie
américain	un poète
américain	un écrivain
américain	un nom de famille
l'orifice du puits	une commune urbaine
l'aéroport RoissyCharlesdeGaulle	une commune urbaine
Étoile	une place
Étoile	une voie
Étoile	une organisation internationale

Tableau XXX – Exemple de catégories considérées non valides pour une entité

Nous présentons dans le tableau XXIX un échantillon d'assignations considérées valides par nos utilisateurs et dans le tableau XXX un échantillon d'assignations considérées non valides. On remarque que l'information que nous avons recueillie semble prometteuse, mais on observe qu'il existe tout de même une certaine ambiguïté dans cette catégorisation. Par exemple, nos joueurs ont considéré la catégorie *une place* non valide pour l'entité "*Étoile*", mais il existe cependant un lieu à Paris portant ce nom.

4.3.2.2 Profils relationnels

Les résultats liés aux profils souffrent également de la quantité d'annotations recueillies jusqu'à présent, en effet ces questionnaires ayant été sélectionnés par les utilisateurs moins souvent que ceux de classifications ils ont donc généré encore moins d'assignations ayant plus d'une occurrence. Pour tenter de pallier à ce problème, nous incluons également les résultats provenant de parties d'entraînements puisque les questionnaires utilisés pour ces parties furent utilisés plus fréquemment et contiennent donc une plus grande quantité d'assignations récurrentes.

Catégorie	Relation	Évaluation
un poisson	lire	Non Valide
un poisson	nager	Valide
un poisson	goûter	Valide
un site touristique	attirer	Valide
un site touristique	être subventionner	Valide
un site touristique	accueillir	Valide
une ville	chanter	Non Valide
une ville	accueillir	Valide
une ville	composer	Valide

Tableau XXXI – Résultats de questionnaires de profils relationnels

4.4 Retour sur les résultats

Comme on a pu l'observer, notre plateforme permet à partir des jeux que nous avons développés de recueillir l'avis d'utilisateurs sur la validité d'inférences que nous avons effectuées en interrogeant la base de connaissances à partir de questionnaires associatifs. Cependant malgré que nos jeux obtiennent un débit similaire à d'autres GWAP dans le domaine, la quantité de joueurs et donc indirectement celle d'avis recueillis jusqu'à présent est nettement inférieure à celle nécessaire pour valider la taille de la base de connaissances que nous avons générée. Nous hypothétisons ce taux de participation par le faible effort publicitaire et montant investi pour populariser les jeux. En effet, comme le font remarquer (Chamberlain et al., 2013), pour transformer un GWAP en succès, il n'est pas suffisant de développer des jeux attrayants, encore faut il mettre en place des méthodes de publicité efficaces.

4.4.1 Perspectives futures

Notre plateforme pourrait potentiellement être améliorée et étendue de multiple façons. Nous explorons ici quelques pistes pour ce projet.

4.4.1.1 Publicité

Comme mentionné précédemment, la plateforme pourrait grandement bénéficier d'un plus grand influx de joueurs pour augmenter le nombre de connaissances validées ainsi que le niveau de confiance sur les validations déjà effectuées par l'acquisition de plu-

sieurs avis provenant de multiples sources. De plus, puisque l'annonce du jeu a uniquement été effectuée sur des petits groupes ayant tous un intérêt pour le domaine du traitement des langues naturelles et un niveau de scolarité plutôt élevé, nous obtenons un groupe de joueurs très homogène. Il pourrait donc être intéressant de publiciser la plateforme à un plus grand public et obtenir une plus grande diversité de joueurs.

4.4.1.2 Intégration aux réseaux sociaux

Dans le but d'attirer un plus grand bassin de joueurs, il pourrait être intéressant de procéder à une intégration de réseaux sociaux à la plateforme pour permettre aux usagers de partager leurs expériences et inviter leurs amis à également contribuer et pourrait donc réduire l'effort publicitaire requis en utilisant nos joueurs comme ressource publicitaire supplémentaire.

4.4.1.3 Ajout de jeux et niveaux de difficulté

En plus des améliorations possibles pour attirer de nouveaux joueurs vers les jeux, on pourrait également tenter d'améliorer la plateforme directement pour offrir une meilleure expérience aux utilisateurs et donc potentiellement obtenir un meilleur taux de rétention et obtenir plus de parties par joueurs. Une des pistes pour atteindre ce but est d'offrir un plus grand éventail de jeux pour permettre aux joueurs de choisir le style de jeu qui lui plait ou encore l'ajout de niveau de difficulté au jeu pour leur permettre d'adapter les jeux à leurs habiletés et leur donner un sentiment d'accomplissement à mesure qu'ils

progressent à travers les différents niveaux de difficulté.

4.4.1.4 Ajout de questionnaires

Ayant développé la plateforme et les jeux qu'elle contient autour du principe de l'utilisation de questionnaires associatifs comme format d'entrée, il pourrait être intéressant de mettre en place une plus grande variété de questionnaires et permettre de traiter un plus grand éventail de données. En effet, nous considérons que de nombreuses tâches pourraient être facilement transposables vers notre structure de questionnaire associatif et pourraient donc bénéficier de l'utilisation de la plateforme pour recueillir l'avis du public.

CHAPITRE 5

CONCLUSION

Au cours de cette expérience, nous avons récupéré une collection de triplets générés à partir de techniques d'extraction d'information ouverte et, après avoir effectué une phase de préfiltrage pour éliminer automatiquement les triplets apportant peu d'information utile, nous avons amorcé la construction d'une base de connaissances en peuplant celle-ci à partir d'informations inférées à partir de la collection de triplets tels que : la désignation de catégorie, les profils relationnels et le calcul de similarités entre ces profils. Nous avons ensuite défini une structure de questionnaires associatifs ayant pour but la validation de ces inférences en procédant à une interrogation ciblée des connaissances contenues dans la base et avons ensuite amorcé une plateforme de jeux contenant trois jeux interchangeables utilisant ce type de questionnaire comme format d'entrée pour ludifier le processus de validation. Ces jeux ont permis à la suite d'une annonce sur une petite liste de diffusion interne d'obtenir l'avis d'un petit groupe de joueurs et d'amorcer la validation des informations contenues dans la base de connaissances.

BIBLIOGRAPHIE

- [1] Michele Banko, Michael J Cafarella, Stephen Soderland, Matthew Broadhead et Oren Etzioni. Open information extraction from the web. Dans *IJCAI*, volume 7, pages 2670–2676, 2007.
- [2] Jonathan Barone, Colin Bayer, Rowan Copley, Nova Barlow, Matthew Burns, Sundipta Rao, Georg Seelig, Zoran Popovic, Seth Cooper et N Players. Nanocrafter : Design and evaluation of a dna nanotechnology game. Dans *FDG*, 2015.
- [3] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr et Tom M Mitchell. Toward an architecture for never-ending language learning. Dans *AAAI*, volume 5, page 3, 2010.
- [4] Irene Celino, Dario Cerizza, Simone Contessa, Marta Corubolo, Daniele Dell’Aglia, Emanuele Della Valle et Stefano Fumeo. Urbanopoly—a social and location-based game with a purpose to crowdsource your urban data. Dans *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 910–913. IEEE, 2012.
- [5] Jon Chamberlain, Karën Fort, Udo Kruschwitz, Mathieu Lafourcade et Massimo Poesio. Using games to create language resources : Successes and limitations of the approach. Dans *The People’s Web Meets NLP*, pages 3–44. Springer, 2013.

- [6] Jon Chamberlain, Massimo Poesio et Udo Kruschwitz. Phrase detectives : A web-based collaborative annotation game. Dans *Proceedings of the International Conference on Semantic Systems (I-Semantics' 08)*, pages 42–49, 2008.
- [7] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović et al. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, 2010.
- [8] Mihaly Csikszentmihalyi. *Flow : The psychology of optimal performance*, 1990.
- [9] Xin Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmman, Shaohua Sun et Wei Zhang. Knowledge vault : A web-scale approach to probabilistic knowledge fusion. Dans *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 601–610. ACM, 2014.
- [10] Oren Etzioni, Michele Banko, Stephen Soderland et Daniel S Weld. Open information extraction from the web. *Communications of the ACM*, 51(12):68–74, 2008.
- [11] Anthony Fader, Stephen Soderland et Oren Etzioni. Identifying relations for open information extraction. Dans *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545. Association for Computational Linguistics, 2011.
- [12] Christiane Fellbaum. *WordNet*. Wiley Online Library, 1998.

- [13] Karën Fort, Bruno Guillaume et Hadrien Chastant. Creating zombilingo, a game with a purpose for dependency syntax annotation. Dans *Proceedings of the First International Workshop on Gamification for Information Retrieval*, pages 2–6. ACM, 2014.
- [14] Ujwal Gadiraju, Ricardo Kawase, Stefan Dietze et Gianluca Demartini. Understanding malicious behavior in crowdsourcing platforms : The case of online surveys. Dans *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1631–1640. ACM, 2015.
- [15] Fabrizio Gotti et Philippe Langlais. Harnessing open information extraction for entity classification in a french corpus. Dans *Canadian AI 2016*. Springer International Publishing Switzerland, Springer International Publishing Switzerland, 2016.
- [16] Alexander Gruenstein, Ian McGraw et Andrew Sutherland. A self-transcribing speech corpus : collecting continuous speech with an online educational game. Dans *International Workshop on Speech and Language Technology in Education*, 2009.
- [17] Chien-Ju Ho, Tao-Hsuan Chang, Jong-Chuan Lee, Jane Yung-jen Hsu et Kuan-Ta Chen. Kisskissban : a competitive human computation game for image annotation. Dans *Proceedings of the acm sigkdd workshop on human computation*, pages 11–14. ACM, 2009.
- [18] David Jurgens et Roberto Navigli. It’s all fun and games until someone annotates :

- Video games with a purpose for linguistic annotation. *Transactions of the Association for Computational Linguistics*, 2:449–464, 2014.
- [19] Alexander Kawrykow, Gary Roumanis, Alfred Kam, Daniel Kwak, Clarence Leung, Chu Wu, Eleyine Zarour, Luis Sarmenta, Mathieu Blanchette, Jérôme Waldispühl et al. Phylo : a citizen science approach for improving multiple sequence alignment. *PloS one*, 7(3):e31362, 2012.
- [20] Mathieu Lafourcade et Alain Joubert. Jeuxdemots : un prototype ludique pour l'émergence de relations entre termes. Dans *JADT'08 : Journées internationales d'Analyse statistiques des Données Textuelles*, pages 657–666, 2008.
- [21] Chris Lintott, Kevin Schawinski, Steven Bamford, Anže Slosar, Kate Land, Daniel Thomas, Edd Edmondson, Karen Masters, Robert C Nichol, M Jordan Raddick et al. Galaxy zoo 1 : data release of morphological classifications for nearly 900 000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 410(1):166–178, 2010.
- [22] Miguel Angel Luengo-Oroz, Asier Arranz et John Freen. Crowdsourcing malaria parasite quantification : an online game for analyzing images of infected thick blood smears. *Journal of medical Internet research*, 14(6), 2012.
- [23] Ian McGraw, Alexander Gruenstein et Andrew M Sutherland. A self-labeling speech corpus : collecting spoken words with an online educational game. Dans *Interspeech*, pages 3031–3034, 2009.

- [24] Marta Sabou, Kalina Bontcheva, Arno Scharl et Michael Föls. Games with a purpose or mechanised labour ? : A comparative study. Dans *Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies*, page 19. ACM, 2013.
- [25] Kristin Siu, Alexander Zook et Mark O Riedl. Collaboration versus competition : Design and evaluation of mechanics for games with a purpose. Dans *FDG*, 2014.
- [26] Bartholomäus Steinmayr, Christoph Wieser, Fabian Kneißl et Fracois Bry. Karido : A gwap for telling artworks apart. Dans *Computer Games (CGAMES), 2011 16th International Conference on*, pages 193–200. IEEE, 2011.
- [27] Penelope Sweetser et Peta Wyeth. Gameflow : a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, 3(3):3–3, 2005.
- [28] Kathleen Tuite, Noah Snavely, Dun-Yu Hsiao, Adam M Smith et Zoran Popović. Reconstructing the world in 3d : bringing games with a purpose outdoors. Dans *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, pages 232–239. ACM, 2010.
- [29] Luis Von Ahn et Laura Dabbish. Labeling images with a computer game. Dans *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326. ACM, 2004.

- [30] Luis Von Ahn et Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.
- [31] Luis Von Ahn, Mihir Kedia et Manuel Blum. Verbosity : a game for collecting common-sense facts. Dans *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 75–78. ACM, 2006.