

ACADEMIE DE MONTPELLIER
UNIVERSITÉ MONTPELLIER II
Sciences et Techniques du Languedoc

THÈSE

présentée au Laboratoire d'Informatique de Robotique
et de Microélectronique de Montpellier pour
obtenir le diplôme de doctorat

Spécialité : **Informatique**
Formation Doctorale : **Informatique**
École Doctorale : **Information, Structures, Systèmes**

**Structuration de débats en ligne à l'aide d'Annotations
socio-sémantiques**
Vers une analyse de réseaux sociaux centrés sur l'interaction

par

Antoine SEILLES

Version du 1^{er} avril 2012

Directeur de thèse

M. Jean SALLANTIN, DR LIRMM, CNRS

Co-directrice de thèse

Mme. Nancy RODRIGUEZ, MCF LIRMM, UM2

Rapporteurs

Mme. Brigitte GRAU, Professeur LIMSI

M. Fabien GANDON, Chargé de recherche HDR INRIA, Sophia Antipolis, Nice

Examinateurs

Mme. Catherine GARBAY, DR CNRS laboratoire d'informatique de Grenoble

M. Francis CHATEAURAYNAUD, DR EHESS

Mme. Marie Laure MUGNIER, Professeur LIRMM

i

A Camille



Table des matières

Table des matières	iii
Remerciements	1
Résumé	3
Résumé étendu	6
Introduction	6
Background et état de l'art	7
Proposition	8
Expérimentations	8
Conclusion	9
1 Introduction	11
1.1 Problématique	11
1.2 Intermed	14
1.3 Les nouveaux citoyens sur le Web	16
1.4 Débats 2.0	17
1.4.1 Contenu généré par les utilisateurs	18
1.4.2 Débat à grande échelle	19
1.4.3 Exprimer et représenter des opinions	21
1.4.4 L'existant	24
1.4.5 Collectivités 2.0	26
1.4.6 Verrous des débats 2.0	27
1.5 Stratégie adoptée	29

2 Background	31
2.1 Notions intrinsèques aux débats	31
2.1.1 Visées énonciatives et Actes de paroles	32
2.1.2 Argumentation et Arguments	32
2.1.3 Parties Prenantes	33
2.2 Les usages d'Internet	34
2.2.1 Origine d'Internet	34
2.2.2 Le fonctionnement d'Internet et du Web	37
2.2.3 Internet et le travail collectif	45
2.2.4 Les nouveaux usages du Web	48
2.2.5 Web 2.0	51
2.3 L'interopérabilité	55
2.3.1 Web Sémantique	57
2.3.2 Web Science	60
2.4 Les réseaux sociaux	61
2.4.1 Histoire	61
2.4.2 Théorie des graphes	62
2.4.3 Caractéristiques des graphes de réseaux sociaux	65
2.4.4 Détection de communautés	65
2.4.5 Éléments de sciences sociales	68
3 Etat de l'art	71
3.1 La démocratie électronique	71
3.1.1 Histoire	72
3.1.2 Les domaines de la démocratie électronique	73
3.1.3 Enjeux et objectifs	74
3.1.4 Limites de la démocratie électronique	75
3.2 Les débats en ligne	76
3.2.1 Vers des débats 2.0	76
3.2.2 Les outils du Web Social	76
3.2.3 L'existant en matière d'expression d'opinion	84
3.2.4 L'existant en représentation des débats	87
3.2.5 Structuration des débats via les n-opposition	92
3.2.6 Synthèse des fonctionnalités pour les débats 2.0	94
3.3 L'annotation	95
3.3.1 Histoire	95
3.3.2 Les différentes formes d'annotation sur le Web	96
3.4 L'analyse des SocNets	112
3.4.1 Identité numérique	112
3.4.2 Terminologie	113
3.5 Limites de l'état de l'art	116

4 Proposition	119
4.1 Spécification d'un outil de débat 2.0	120
4.1.1 Scénarios d'usage	122
4.1.2 n-opposition et expression d'opinions	124
4.2 L'annotation	125
4.2.1 L'annotation socio-sémantique	126
4.2.2 Formes d'usage et expressivité de nos annotations	127
4.2.3 Définition formelle	130
4.3 Les réseaux sociaux	133
4.3.1 Réseau social calculé à partir d'interactions	133
4.3.2 Identité numérique	134
4.3.3 Déduction d'un réseau à partir de l'annotation	135
4.3.4 Classification	137
4.4 Interopérabilité	138
4.4.1 Représenter les données de la discussion	139
4.4.2 Représentation des réseaux sociaux	142
4.4.3 Synthèse du modèle de données	143
4.5 Exemple complet	145
5 Expérimentations	149
5.1 Premier test de l'annotation : Betapolitique	150
5.1.1 Description du dispositif	151
5.1.2 Structure d'oppositions pour les jugements	151
5.1.3 Analyse	152
5.2 Deuxième test de l'annotation : Xwiki	153
5.2.1 Description du dispositif	153
5.2.2 Analyse	155
5.3 Usages de l'annotation dans Intermed	157
5.3.1 Phases de coconception et évocation des besoins	157
5.3.2 Description du premier dispositif	161
5.3.3 Description du deuxième dispositif	162
5.3.4 Analyses	163
5.4 Moissonnage de données Facebook	165
5.4.1 Contexte et dispositif	165
5.4.2 Description des graphes	166
5.4.3 Résultats	167
5.4.4 Conclusion	168
5.5 Argumentea : Prototype actuel	168
5.5.1 Choix d'implémentation	169
5.5.2 Plateforme évolutive	169
5.5.3 Expérimentations Argumentea	170

5.5.4 Utilisation d'Argumentea par des tiers	176
6 Conclusion	177
6.1 Reformulation de la problématique	177
6.2 Perspectives	180
6.2.1 Perspectives théoriques	180
6.2.2 Perspectives applicatives	181
Glossaire	183



Remerciements

Ces un peu plus de trois années au LIRMM ont été l'occasion de rencontrer beaucoup de monde, venant de plusieurs domaines, pas que des chercheurs, et de beaucoup apprendre. Comme il est impossible que je pense à remercier tout ce monde nominalement, j'adresse un remerciement groupé à toutes les personnes avec lesquelles il m'a été possible d'échanger, de discuter, de partager, d'expérimenter ... pendant le temps de cette thèse. Bien sûr il y a tout de même des contacts privilégiés que je peux remercier nominalement : Jean et Nancy qui m'ont permis de réaliser une thèse ou tout a été construit ensemble du début à la fin. Alors oui, nous n'avons pas vraiment fait ce que nous pensions faire au début mais en tout cas mon sujet de thèse ne m'a pas été imposé, je suis responsable de mes choix et je pense avoir pris du niveau en autonomie. Martine qui a souvent été là pour soutenir les troupes. Eric, avec qui nous avons entamé un projet jamais terminé. Peut-être qu'un jour on s'y remettra. Mais en attendant on va pouvoir plus facilement aller boire des bières ensemble maintenant que tu es sur Montpellier. Stefano, je pense que nos discussions bien que trop rares m'ont permis de construire ma vision du Web Social et du Web Science. Abdel, pour m'avoir fait arrêté le café et ouvert la voie sur un champ disciplinaire qui m'intéresse tout particulièrement aujourd'hui. Et puis bien sur, Julien, compagnon de galère pendant trois ans, la pluridisciplinarité et les projets à trente-six acteurs c'est pas toujours la joie. Intactile Design, toute l'équipe, particulièrement efficace, la coconception bien organisée ça existe vraiment !

Je dois également remercier mes proches, ma famille, mes amis. Là encore je ne pourrai pas remercier tout le monde nominalement : mes parents, ma sœur, mes grands-parents, cousins, cousines, Fabien, l'association imagiLAB (rock'n roll dans l'âme), BlackPad, tous les doctorants du LIRMM, et mes nouveaux collègues Ben, Seb, Tristan, Ines.

Et bien sur, Camille, à qui je dédie cette thèse.



Résumé

Préambule

Ce résumé précise le propos de la thèse. Il en explique les composantes principales et précise à qui, quels lecteurs, cette thèse est dédiée.

Sommaire

Résumé étendu	6
-------------------------	---

Avant-propos

Internet, et plus particulièrement le Web, est devenu l'un des canaux majeurs de communication si ce n'est le canal majeur. Sur un plan technique, les protocoles et dispositifs nécessaires pour le fonctionnement d'Internet sont à peu près les mêmes qu'à l'origine dans les années 80. Bien sur il y a eu quelques innovations techniques comme XML et le Javascript par exemple, mais ces innovations, motivées par de nouveaux usages, ne sont que des couches intermédiaires s'intercalant dans les communications clients-serveurs, entre envoi de requêtes et envoi de réponses.

Dès le début d'Internet, le potentiel de celui-ci pour améliorer les processus démocratiques (en particulier la mise en place de discussion entre un grand nombre d'acteurs sur un grand nombre de sujets, faisant fi des distances et des problèmes de disponibilités de chacun, dans le but de participer à des prises de décision) était mis en exergue. On parlait déjà de démocratie électronique mais Internet allait en changer profondément les possibilités. Le Web originel et ses premiers usages, les Mails, les Bulletin Boards, les Newsgroups n'ont pas révolutionné la démocratie pour autant. Mais ces premiers usages en ont en-

gendré d'autres, donnant de plus en plus d'idées aux utilisateurs et développeurs jusqu'à aboutir au Web 2.0, cette soi-disant nouvelle version du Web. Aujourd'hui, la participation en ligne est impressionnante. L'utilisateur est en permanence créateur de données et ce dans de nombreuses activités (et parfois même sans en avoir conscience). La vraie nouveauté présentée du Web 2.0 est sa composante sociale. Quelles qu'elles soient, nos activités en ligne (écouter de la musique, discuter autour d'un intérêt commun, regarder des vidéos, faire ses courses, ...) se basent sur le partage de données avec d'autres personnes. Internet est bien un dispositif permettant de créer du lien social, exactement comme vers la fin des années 60 Licklider et Taylor [[Licklider et Taylor, 1968](#)] rêvaient l'informatique.

En maintenant près de 30 ans, Internet est devenu un mode de communication multi-usage quasiment incontournable. Et pourtant, la démocratie électronique en reste elle à ses balbutiements. La demande citoyenne pour plus de participation à la vie politique est forte, de nombreux collectifs ou associations proposent des sites web permettant de débattre, ils proposent même de plus en plus d'outils conçus ou modifiés par leurs soins, s'appuyant sur le mouvement du logiciel libre. Mais un frein essentiel persiste, la reconnaissance et l'adoption par les gouvernements de ces solutions dans de véritables processus de prise de décision. Ce frein est sûrement motivé par des craintes de perte de contrôle d'une part, d'usage non voulu des données de l'autre, mais aussi par le fait que les outils de démocratie électronique ne sont pas encore établis en tant que tels. Des solutions émergent, bricolées par des collectifs, associations, particuliers, chercheurs du domaine, mais quelles sociétés proposent des solutions fiables, robustes, éprouvées ? Non, la démocratie électronique reste malgré tout un domaine récent, expérimental.

Le constat est donc le suivant, la démocratie électronique est un domaine qui s'ancre parfaitement dans les usages du Web 2.0. Le Web 2.0 et l'étude des usages du Web 2.0 est une science pluridisciplinaire par nature (technologique bien sûr, mais aussi nécessitant des connaissances en droit, en sciences sociales, parfois même en physique et en biologie, ...). Et les premiers utilisateurs, les citoyens d'aujourd'hui et de demain, ne sont pas toujours bien informés des évolutions de leur outil principal de communication. Combien ont compris les enjeux des lois Hadopi et de l'ACTA aujourd'hui ? Qui sait ce que signifient et impliquent les fondamentaux d'Internet : la décentralisation des données, la commutation par paquets, l'accès à une adresse ip publique, la neutralité du réseau ...

Dans cette thèse, en réponse à ce constat, nous allons adopter un traitement proche du sens commun. Nous souhaitons que les chercheurs de domaines autres que l'informatique mais aussi des citoyens non chercheurs puissent tirer partie de ce document pour en apprendre d'avantage sur l'outil Internet et sur les enjeux de la démocratie électronique.

Objectifs de thèse et verrous

Cette thèse est réalisée dans le cadre du projet ANR Intermed. L'objectif du projet Intermed est de produire des outils d'aide à la concertation et à la gestion des zones côtières. Intermed regroupe plusieurs acteurs, des laboratoires de recherche et des entreprises, des

informaticiens, des chercheurs en sciences humaines et sociales, des ergonomes ... L'une des priorités du projet était de fournir un environnement, une plateforme permettant la mise en place d'expérimentations. L'un des partis pris du projet était de s'appuyer sur l'annotation comme outil de discussion.

Dans cette thèse, nous avons travaillé sur la conception et l'implémentation d'un outil de discussion basé sur l'annotation. Les axes principaux de recherche concernaient la structuration des discussions, l'interopérabilité des données issues des débats (comment faciliter l'importation, l'exportation et l'analyse des données relatives aux débats, aux discussions produites dans l'outil) et l'analyse de ces données en vue d'extraire un réseau social, le réseau des parties prenantes. L'exploitation de ce réseau étant ensuite laissé à d'autres applications comme des notifications, des recommandations, de la visualisation de données ... Nous avons donc travaillé à la création d'un système d'annotations que nous qualifions de socio-sémantiques permettant de structurer des discussions, des débats en vue d'une analyse permettant d'extraire un réseau social.

Les verrous rencontrés sont les suivants :

- L'annotation étant une nouvelle forme de discussion, il faut s'assurer de son acceptation auprès des utilisateurs ;
- Un grand nombre de participants débattant sur un grand nombre de sujets entraîne la production d'un grand nombre de données. Les outils de discussion pour passer à la grande échelle doivent proposer de nouvelles solutions pour structurer les discussions et en faciliter le traitement.
- Dans le contexte des débats publics, un débat ne sert à rien s'il n'aboutit pas sur une prise de décision. Si l'opinion publique est concertée mais n'est pas prise en compte, la concertation est contre productive. Elle risque de créer un désintérêt des citoyens pour la concertation et la vie politique en générale. Ce n'est pas à un outil de débat de proposer les décisions à prendre, mais d'autres outils peuvent s'appuyer sur des données de débats pour aider à prendre des décisions. Les outils de débats doivent donc favoriser le traitement et la réutilisabilité des données qu'ils produisent concernant les opinions des citoyens.
- Les groupes d'opinion ont tendance à se fermer sur eux-mêmes, à discuter au sein du groupe plutôt qu'entre groupes d'opinions divergentes. Ainsi, on observe une insularisation des groupes d'opinion sur Internet et avec celle-ci une radicalisation des opinions. Cette dynamique ne favorise pas l'acquisition de nouvelles informations par les individus, la naissance de nouvelles opinions alternatives et donc n'améliore pas la qualité des débats et ne permet pas de trouver de solutions consensuelles. Les outils de débats doivent donc favoriser l'ouverture d'esprit et des discussions entre les groupes plutôt que de favoriser l'insularisation.

Le système sur lequel nous avons travaillé propose donc une forme d'annotation structurante se basant sur l'usage de formulaires permettant d'exprimer des opinions. Le format des données produites par le système s'appuie sur l'état de l'art des technologies du web sémantique dédiées à la représentation des données produites par un réseau social et des

données propres à l'argumentation. Nous proposons l'utilisation conjointe de plusieurs de ces formats et les étendons pour couvrir l'ensemble des données dont nous avons besoin. Les indicateurs définis, se basant sur une première analyse des données produites par les utilisateurs, sont pensés pour l'analyse de réseaux sociaux en vue de lutter contre l'insularisation des parties prenantes ou groupes de pressions. Ces indicateurs vont permettre de caractériser des comportements et faciliter la définition de profils utilisateurs similaires ou complémentaires.

Résumé étendu

Introduction

Dans l'introduction de la thèse, nous revenons sur la problématique qui nous concerne :

- Comment faciliter l'indexation et la navigation dans la masse de données créées lors de discussions en ligne concernant la vie citoyenne ? L'une des sous-questions soulevées est celle de la définition d'indicateurs de pertinence dédiés aux débats.
- Comment permettre l'échange et la réutilisabilité des données produites par les débats en ligne ? Notamment en vu de simplifier la prise de décisions à partir de concertations citoyennes.
- Enfin, comment faciliter la discussion entre groupes d'opinions différentes voir opposées ? Nous soulignons que l'un des enjeux est de lutter contre l'insularisation des groupes et la radicalisation d'opinions que cela peut entraîner.

Nous revenons ensuite sur le contexte et les objectifs de la thèse, à savoir l'ANR Intermed, projet pluridisciplinaire, et l'objectif de développer des outils en ligne dédiés à la concertation publique. Nous en profitons aussi pour cibler les usagers qui nous concernent, les nouveaux citoyens, citoyens ayant toujours connu Internet. Enfin, nous définissons ce que nous entendons par Débats 2.0 en soulignant l'importance des contenus générés par les citoyens dans des discussions à grande échelle, l'échelle d'une collectivité territoriale. Nous présentons également l'idée au cœur de notre proposition, l'utilisation de l'expression d'opinions par les utilisateurs pour structurer les discussions, avant de rapidement discuter l'existant en matière d'expression d'opinions en ligne. Enfin, nous définissons un concept fortement lié à celui de débat 2.0, le concept de collectivité 2.0, et nous présentons l'ensemble des freins présentés dans la littérature concernant la mise en place de concertations en ligne.

La dernière partie de l'introduction est consacrée à la présentation de notre stratégie d'approche du problème. Nous proposons de structurer les débats 2.0 à l'aide de l'expression d'opinions par les utilisateurs. Nous utiliserons un outil d'annotation qualifié de socio-sémantique permettant aux utilisateurs d'exprimer explicitement leurs opinions et de les ancrer dans les discussions et dans des documents Web déjà existants et permettant

de construire un corpus de données interopérables en utilisant les technologies du Web sémantique. La définition des données capitales des débats 2.0 a été fortement influencée par une réflexion sur la dynamique des réseaux sociaux intrinsèque aux débats.

Background et état de l'art

La section Background et état de l'art mêle deux types d'information capitales pour bien comprendre la proposition de la thèse : le background fait référence à tout ce qu'il faut connaître, le savoir pré-requis pour pouvoir étudier le domaine et l'état de l'art fait référence à ce qui se fait de mieux concernant notre problématique.

Nous commencerons par présenter Internet, probablement la technologie la plus utilisée aujourd'hui en tant que champ expérimental pour la démocratie électronique. Bien que cette technologie soit quotidiennement utilisée pour la démocratie électronique, il nous est apparu que celle-ci restait mal connue par de nombreux utilisateurs et chercheurs. Ce n'est probablement pas un pré-requis bloquant, mais il n'empêche que revenir aux fondamentaux d'Internet permet de mieux comprendre les intérêts et enjeux d'Internet pour la démocratie.

L'interopérabilité est un enjeu clef du Web de demain et parler d'interopérabilité sans parler du Web sémantique et des technologies qu'il apporte n'aurait pas de sens. Nous présentons donc les objectifs du Web sémantique et la technologie RDF utilisée pour définir les formats de données dans ce travail. Nous en profitons pour présenter succinctement le Web Science et défendre le fait que, aujourd'hui, mener un travail de recherche dans le domaine du Web fait nécessairement intervenir des compétences pluridisciplinaires.

Bien que nous ne proposions pas de nouvelles méthodes d'analyse de réseaux sociaux et que le cœur de notre travail concerne l'annotation, le domaine de l'analyse des réseaux sociaux est omniprésent dans notre réflexion sur la conception d'outils de débats 2.0 et sur les données qu'il devient nécessaire d'échanger entre applications. Nous présenterons des éléments de théories de sciences sociales qui ont nourri notre définition de formats de données. Nous avons souhaité que les données produites par les débats en ligne soient particulièrement adaptées à l'extraction de communautés d'opinions. La bonne détection de ces communautés est fondamental pour permettre la mise en place de discussions inter-communautés.

La démocratie électronique est l'usage des technologies de l'information et de la communication en vue d'améliorer la pratique de la démocratie. C'est donc un domaine ancien mais qui semble voué à rester au niveau de l'expérimental. Nous présentons succinctement ce domaine afin de positionner nos travaux dans les champs de la concertation électronique, de la mesure d'opinions et en amont de la détection de communautés.

Nous présentons ensuite l'existant en matière de débats en ligne et d'expression d'opinions. Nous listons donc un ensemble d'outils et de pratiques bien connues ou plus innovantes sur Internet et plus particulièrement le Web. En particulier, nous présentons à la fin de cette partie un ensemble de solutions de représentation des données des débats.

Puis nous présentons la théorie des n-opposition, théorie inspirée par les travaux d'Aristote concernant l'organisation de différentes formes de négations ou contradictions. Cette théorie permet de définir des outils géométriques de représentation des oppositions entre jugements, entre opinions.

Enfin, nous finissons par une partie dédiée à la présentation de l'annotation. Nous présentons les deux pratiques courantes de l'annotation : les annotations discursives dédiées à la discussion et les annotations sémantiques dédiées à la représentation de connaissances, l'écriture de méta-données concernant des ressources. Nous présenterons également le modèle de [Agosti et Ferro, 2007] que nous critiquerons afin de proposer notre propre modèle.

Proposition

Notre proposition repose sur la construction d'un système de débats 2.0 basé sur l'usage d'annotations socio-sémantiques. L'interopérabilité des données produites par ce système en vue de les réutiliser dans d'autres applications est un point central de la conception. Les débats 2.0 sont une phase d'un processus de démocratie électronique. Les débats sont un élément d'articulation entre prise d'information et prise de décision, et ils sont une phase de construction de groupes d'opinion.

Nous présentons tout d'abord la phase de spécification du système développé que nous illustrerons par des scénarios d'usages. Nous décortiquerons ensuite plus en détail le modèle et l'usage des annotations socio-sémantiques que nous avons développé. Nous discutons ensuite les choix d'indicateurs et de données concernant les réseaux sociaux dans les débats. Tous ces points ne sont pas nécessairement traduits en fonctionnalités dans le système implémenté mais font partie des éléments d'interface entre le système et d'autres applications auxquels nous avons songé et qui ont donc en partie guidé notre conception du système. Nous présentons ensuite les formats de données pour l'interopérabilité. Puis nous détaillerons un exemple complet d'usage du système développé.

Expérimentations

Notre travail de thèse a été fortement rythmé par une série d'expérimentations. Ce rythme était partiellement imposé par l'ANR Intermed mais est aussi la conséquence d'une volonté de travailler sur le terrain, de valider l'utilisabilité dans une forme de conception Agile de notre système de débats.

Nous présentons donc dans cette partie le déroulement et les conclusions des phases d'expérimentations menées durant la thèse.

Conclusion

Nous reprendrons en conclusion les verrous abordés afin de souligner notre approche. Nous présenterons une synthèse de nos travaux et enfin proposerons des perspectives théoriques et applicatives.

CHAPITRE

1

Introduction

La dictature c'est "ferme ta gueule", la démocratie c'est "cause toujours"

NON ATTRIBUÉ

Préambule

Nous présentons notre problématique, comment améliorer la participation en ligne aux débats autour de la gestion de biens communs. Cette thèse s'inscrit dans le contexte de la démocratie électronique et des nouveaux usages du web.

Sommaire

1.1	Problématique	11
1.2	Intermed	14
1.3	Les nouveaux citoyens sur le Web	16
1.4	Débats 2.0	17
1.5	Stratégie adoptée	29

1.1 Problématique

Parmi les fondamentaux de la démocratie, on retrouve le droit d'expression, la liberté d'opinion, l'accès à tous et pour tous à l'information et aux services, l'égalité face à la parti-

cipation. Tous ces principes se retrouvent dans les idées fondatrices d'Internet. En particulier, Internet est aujourd'hui le seul dispositif technique qui permette la pleine application de l'article 11 de la déclaration des droits de l'homme et du citoyen :

La libre communication des pensées et des opinions est un des droits les plus précieux de l'homme : tout citoyen peut donc parler, écrire, imprimer librement, sauf à répondre à l'abus de cette liberté dans les cas déterminés par la loi.

Aujourd'hui, on retrouve dans les pratiques collaboratives et dans les communautés de défense du libre l'expression de ces principes. La nouvelle génération, souvent appelée génération digitale comme traduction de "digital natives", a développé un usage d'Internet et du Web qui change fortement la sociabilité et les formes de communication. Ces nouveaux utilisateurs sont déjà des acteurs de la vie politique (la génération digitale concerne les individus nés après 1985, certains d'entre eux sont donc déjà majeurs, électeurs). Il est donc impératif de prendre en considération leur façon d'interagir dans le développement de nouveaux outils pour la démocratie.

Bien que l'usage d'Internet soit avant tout concentré autour d'activités de loisirs, on observe une demande d'outils de démocratie électronique par les citoyens. En particulier concernant les outils de participation. Les citoyens, par ailleurs consommateurs de loisirs en ligne, ont pris l'habitude d'exprimer leurs avis, leurs opinions, sur tout (achats en ligne, lieux de loisirs, préférences, musique, photos ...) par le biais d'outils collaboratifs, d'outils de discussion ou encore de sondages. Le nombre de blogs, forums ou autres sites consacrés à l'information ou à la discussion politique ne cesse d'augmenter. Et ceux-ci ne sont pas toujours portés par des institutions officielles. Certains voient même Internet comme l'avenir de la démocratie directe car leurs principes fondamentaux sont proches (accès à tout librement pour tous, auto-gestion des communautés, absence de hiérarchies et de priviléges, avis du groupe avant celui de l'expert ...).

Parallèlement, on observe une appropriation d'Internet par les personnalités politiques et les institutions publiques. De nombreuses personnalités politiques ont adopté l'usage d'un blog ou d'un compte Facebook pour rallier les électeurs, les tenir au courant de leurs campagnes. De plus en plus d'administration se dotent de portails web, essentiellement pour communiquer, transmettre de l'information mais aussi de plus en plus pour récolter des retours, des avis (le terme anglais "feedback") de citoyens.

Parmi les arguments invoqués pour l'usage d'Internet dans la vie politique, l'élargissement de la concertation qui s'explique par la réduction des distances et des problèmes de temporalité, l'effet de l'anonymat qui délie les langues des timides, ou encore la simplicité d'usage, mais surtout, la réduction des coûts. Encore faut-il garantir que cette concertation reste de qualité.

Internet n'est pas le premier terrain d'expérimentation de la démocratie électronique. Toutes les innovations dans le domaine des TIC (Technologies de l'Information et de la Communication) ont été l'objet de rêves d'applications pour la démocratie. Cependant, Internet étant le premier média permettant véritablement une communication multi-canal

et active, (l'utilisateur ne fait pas que consommer de l'information, il participe à sa création, l'information n'est pas dirigé vers un seul individu mais vers une multitude d'individus) et étant aujourd'hui fortement répandu grâce à de récentes innovations technologiques (le wifi, l'ADSL ...), Internet et en particulier le Web, devient le terrain prédominant d'expérimentation de la démocratie électronique.

Cependant, malgré tous les atouts d'Internet, celui-ci ne reste qu'un terrain d'expérimentation et les applications de démocratie électronique véritablement passées en production auprès des citoyens ou collectivités restent peu nombreuses (l'usage du vote électronique par exemple est encore aujourd'hui très contesté et donc peu adopté).

Cette situation s'explique par un ensemble de verrous technologiques mais aussi sociologiques, économiques et politiques encore irrésolus. Dans notre travail, nous soulignerons essentiellement les quatre verrous suivants :

- Navigation dans une grande quantité d'information. La démocratie en ligne hérite de la problématique du Web2.0 liée à la quantité d'information. Ce verrou est essentiellement un problème de passage à l'échelle que l'on retrouve dans toutes activités collaboratives. Un grand nombre d'utilisateurs qui créent du contenu entraîne une grande quantité de contenu. Dans un outil de discussion de type forum, une trop grande quantité de contenu entraîne une impossibilité d'avoir une vision de l'ensemble des discussions.
- Se pose aussi la question de la légitimité et de la pertinence des informations. On parlera aussi de "trust" ou de confiance. Si le Web permet à de nouveaux acteurs de devenir des leaders d'opinion ou des informateurs privilégiés (comme cela peut être le cas avec des blogs de journalistes), comment s'assurer que ces acteurs ne profitent pas de leur position ? Comment évaluer leur légitimité ?
- La plupart des informations produites par l'usage d'outils de concertation ne sont finalement pas prises en compte dans les processus réels de décision. Si l'une des explications est toujours celle de la masse de données à traiter, l'autre explication avancée est celle de la représentation de ces données afin de faciliter le traitement et l'échange de données entre applications. Il n'existe pas aujourd'hui de formats de données standards dédiés à la représentation des débats en ligne.
- Enfin, la plupart des outils de discussion dans leurs usages actuels créent une sorte d'**insularisation** de la communauté qui a adopté l'outil. Lorsque un groupe crée un forum, celui-ci leur est réservé. Ils ne créent pas un forum pour partager leur opinion collective avec d'autres et s'ouvrir aux critiques. Ils créent un forum pour renforcer leur identité communautaire, leur opinion partagée.

Notre approche concerne la mise en place d'outils de concertations. Ou comment améliorer la consultation en ligne dans le cadre de la gestion de biens communs ? Nous allons donc considérer les outils actuels permettant d'exprimer son avis et proposer des solutions pour faire face aux problèmes de navigation dans les données produites, pour améliorer l'échelle ou le nombre d'utilisateurs pouvant participer, pour améliorer la détection des communautés d'opinions et leurs leaders. Augmenter la qualité des débats nécessite

de se doter d'un critère de mesure de cette qualité. Augmenter les échanges et la prise en compte des opinions des autres est un élément primordial pour améliorer sa propre argumentation. Nous basons notre mesure de qualité sur ce principe. Enfin, nous proposons des solutions pour l'interopérabilité des outils de consultation afin d'améliorer leur intégration dans des processus de prise de décision.

En particulier, nous mettons au cœur d'applications de discussion dédiées aux débats en ligne la possibilité d'émettre une opinion. Nous proposons d'adopter les technologies du Web sémantique afin de faciliter l'indexation des données et donc la navigation dans une grande quantité de données. Mais aussi pour permettre la mise en place d'autres traitements de données et l'échange de données entre applications dédiées à la démocratie participative. Aujourd'hui, de nombreuses voix citoyennes s'expriment sur Internet mais ne sont pas prises en compte dans les processus de prise de décisions. De nombreux cyber-citoyens estiment que la démocratie participative en ligne c'est "cause toujours" (nous ne reviendrons pas dans ce manuscrit sur le problème des cyber-dictatures ou des possibilités de contrôle qu'il est possible de mettre en place). Enfin, nous proposons d'adopter et d'adapter des méthodes de détection de communautés afin de détecter des groupes d'opinion et leurs nouveaux leaders tout en fournissant des indicateurs pour l'évaluation de leur légitimité.

1.2 Intermed

Notre travail de recherche a été financé par l'ANR Intermed. Intermed a à la fois défini un cadre de recherche et un cadre expérimental. Intermed est un projet financé par l'Agence Nationale de la Recherche. L'objectif du projet est de proposer des outils d'aide de gestion de zones côtières, centrés sur le débat participatif avec des outils Internet. Notre travail s'inscrit plus particulièrement dans le travail de conception et de développement d'outils d'aide à la concertation, de facilitation des débats, de facilitation d'écriture collective dans ce contexte, et permettant de recréer un lien fonctionnel entre citoyens, gestionnaires et gouvernants. La figure 1.1 récapitule les objectifs du projet.

Intermed est un projet multidisciplinaire ayant permis de mettre en place des échanges entre chercheurs de sciences humaines, des Sciences et Technologies de l'Information et des Communications (STIC) et d'informatiques, entre des laboratoires (LIRMM, Cemagref), des sociétés (Normind, Scriptal, Géomatys, Pikko) et des partenaires de terrain (sites d'application). D'autres sociétés sont venues se greffer au projet sur des missions précises comme par exemple la société Intactile Design qui nous a apporté une aide précieuse en nous proposant des designs d'interfaces innovants, et un support efficace sur les phases de coconceptions. La figure 1.2 présente les rôles des différents partenaires.

Intermed fixe un cadre expérimental. Celui-ci impose la conduite de 9 protocoles expérimentaux en 3 ans. Les expérimentations devant être avec utilisateurs, sur le terrain autour de sujets concrets de gestion de territoire ou de biens communs. Nous avons, dans

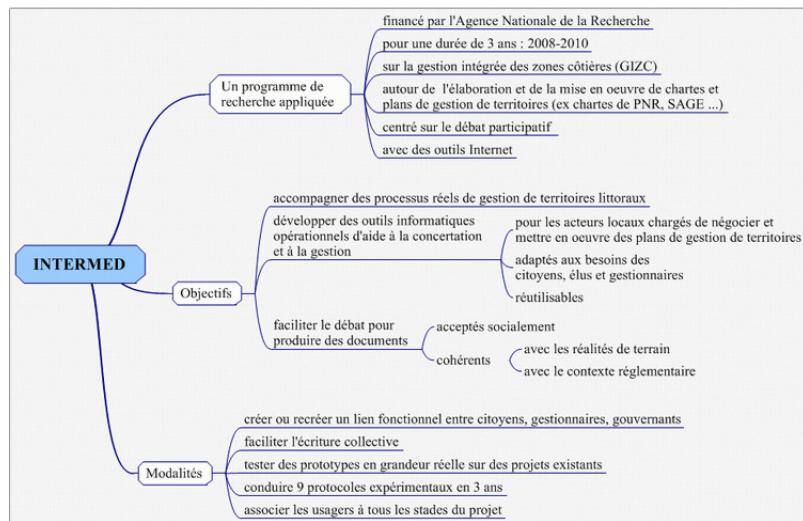


FIGURE 1.1 : Présentation du projet ANR-Intermed

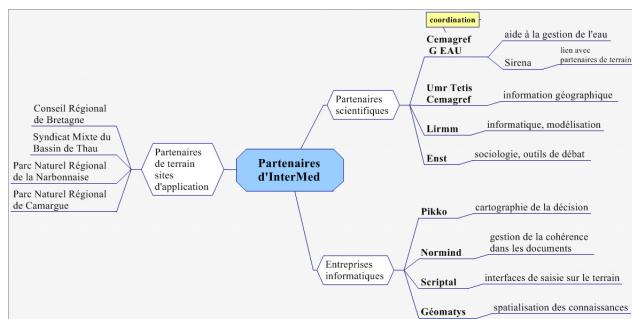


FIGURE 1.2 : Présentation du projet ANR-Intermed

le cadre de cette thèse, contribué à certaines de ces expérimentations et en avons mené d'autres, en marge du projet afin d'éprouver certaines spécificités des outils développés.

Le projet Intermed a été pour nous une première expérience de pluridisciplinarité. Il nous semble que travailler dans le domaine de la recherche autour du Web implique nécessairement cette ouverture entre Sciences. Aujourd'hui il paraît difficile d'analyser des usages sur le Web ou d'en proposer de nouveaux sans avoir à la fois des compétences en sciences humaines et sociales, en ergonomie et en informatique. Bien que notre travail soit rattaché à un laboratoire d'informatique, beaucoup de nos références bibliographiques viennent d'autres domaines comme les Sciences Humaines et Sociales. La pluridisciplinarité impose cependant un ensemble de problématiques qu'il ne faut pas sous-estimer quand on participe à un tel projet. Nous avons observé la liste de difficultés suivantes liées à la pluridisciplinarité :

- Les différents acteurs sont en général peu conscients des compétences et attentes des gens qui ne sont pas dans leur domaine.
- Les stratégies et cultures de publications sont différentes pour chaque domaine, ce qui rend complexe la valorisation équitable d'un tel projet.
- Sous-estimation du besoin de travail en présence pour limiter les clivages et incompréhensions entre les représentants des différentes disciplines. Les problèmes logistiques peuvent devenir prépondérants et nuire à la bonne conduite du projet.

Ces difficultés surmontées, la pluridisciplinarité reste un contexte très enrichissant dont les avantages sont :

- L'échange de méthodologies de conception et d'évaluation. Apport d'une culture du qualitatif qui fait souvent défaut aux informaticiens.
- La mise en place d'expérimentations fréquentes en présence d'utilisateurs réels.
- Un échange de points de vue qui permet aux différents acteurs de prendre du recul par rapport à leurs domaines d'études.
- Une dynamique de conception ou fonctionnalités et nouvelles expérimentations permettent une boucle de prototypage efficace et enrichissante.

1.3 Les nouveaux citoyens sur le Web

L'année de création d'Internet communément admise est 1989. Les "jeunes électeurs" ou électeurs de moins de 25 ont presque toujours connu Internet. En très peu de temps, la téléphonie mobile et Internet sont devenus des outils indispensables dans nos pratiques sociales, dans notre travail, dans nos loisirs. Pour la génération des moins de 25 ans, les natifs digitaux ou génération digitale, la communication et les rapports sociaux ne peuvent pas se concevoir sans ces deux technologies. Ne pas avoir de téléphone portable ou ne pas avoir d'accès Internet c'est se couper du monde. Il est important de considérer ces utilisateurs car ils sont hyper spécialisés, ce sont des hyper-communicants [[Compiègne, 2011](#)]. Leur rapport à Internet et la téléphonie mobile, leurs habitudes sont différentes et révélatrices de l'avenir de notre société. Ils sont la première génération Internet mais probablement pas la dernière.

En parallèle de cette hyper-connaissance de l'usage, se développe une hyper-méconnaissance des fondements. Les natifs digitaux sont des super-utilisateurs. Mais il a été constaté qu'ils sont peu au courant des enjeux techniques et légaux des outils qu'ils utilisent. En particulier, l'enquête [[Kvavik, 2005](#)] montre que la génération digitale se contente d'une connaissance des outils qui se limite au minimum de leurs besoins. Ils ont l'habitude des outils de communication et les considèrent comme acquis. Ils ne s'intéressent pas à leur fonctionnement et à l'aspect légal de ceux-ci. Si ils sont de grands utilisateurs des outils de communication et d'Internet, ils ne sont pas pour autant intéressés par le domaine de l'informatique, par l'apprentissage de la technique. Enfin, il faut aussi souligner que leur apprentissage de ces outils se base sur une suite d'essais et

d'erreurs créant une habitude de l'expérimentation et parfois au détriment de la réflexion. L'usage d'Internet et en particulier du Web chez les jeunes met en exergue, au-delà de la méconnaissance de la technique, l'absence de sensibilisation et de connaissance des problèmes de confidentialités, de données personnelles publics et privées, de droits de propriété intellectuelle. Il faut aussi souligner que les habitudes créent des automatismes parfois problématiques. Par exemple, la méconnaissance de la logique "Google" ou "PageRank" qui classe comme le plus pertinent ce qui est le plus cité ne favorise pas l'esprit critique et la remise en question des idées répandues. Si "la sagesse des foules" (en référence à "the wisdom of crowd" [Surowiecki, 2004]) est prépondérante sur Internet, il reste des situations, des domaines ou la majorité ne peut avoir raison (par exemple en sciences). Considérer que le plus grand nombre a toujours raison c'est ignorer beaucoup d'aspects fondamentaux dans l'évaluation de la pertinence. Concernant les résultats de Google, il faut souligner que le système PageRank a tendance à favoriser les sites bien établis, donc les plus anciens, puisqu'ils ont accumulé un nombre de liens entrants que les nouveaux sites mettront longtemps à concurrencer. Dans le contexte de l'information, au contraire, l'antériorité est rarement gage de qualité. Cette sagesse des foules prend sens dans certains contextes de démocraties si l'on accepte la règle d'un individu, une voix, un vote.

De plus, il est important, dans le cadre de processus démocratiques en ligne, de ne pas oublier les oubliés d'Internet. Ceux qui ne peuvent pas par manque de moyens, manque technologique ou encore manque de connaissance, accéder à Internet. Le terme de *fracture numérique* est employé pour souligner qu'il y a ceux qui sont à l'aise avec Internet et les autres. Considérer ce problème de fracture numérique aujourd'hui dans la conception d'applications dédiées à la démocratie en ligne permet de souligner qu'il y a encore des limites sociales et politiques à son développement. Autrement dit, nous ne pensons pas pouvoir proposer de solutions pour une démocratie exclusivement en ligne. Nous proposons des solutions pour la consultation d'opinions en ligne mais nous soulignons que ces solutions sont adaptées pour des utilisateurs habitués du Web. Adopter uniquement de telles solutions reviendrait à exclure de la vie démocratique les déjà exclus d'Internet. Il faut donc penser la consultation en ligne comme un complément d'autres formes de consultations pensées pour d'autres cibles.

1.4 Débats 2.0

Nous définissons dans cette section le concept central qui permet de préciser le contexte de notre travail. Nous avons appelé ce concept "Débat 2.0" afin de faire référence au concept de Web 2.0. Le débat 2.0 est l'usage d'outils du Web 2.0 pour les débats. Nous reviendrons donc sur l'idée qui nous semble central dans le concept de Web 2.0 : le contenu généré par l'utilisateur et l'amélioration des services par leur usage. Puis nous soulignerons ce que nous entendons par grande échelle et son importance dans les débats 2.0. Enfin, nous reviendrons sur le besoin central des débats 2.0, l'expression d'opinion. Les débats

2.0 ou débats en ligne actuels sont essentiellement dédiés à la concertation. Le but de la concertation est d'avoir un aperçu de l'opinion des citoyens sur un sujet que ce soit par la réponse à des sondages ou la participation à des débats en présence ou en ligne via forum ou autres outils. Le terme de débat n'est en général associé qu'à des concertations venant des dirigeants ou décisionnaires et rarement associé à des volontés de citoyens souhaitant participer activement par le biais d'une prise de décision collective. Alors que pourtant, il semble exister une forte demande citoyenne. Celle-ci s'observe par le biais des projets libres, des collectifs ou associations qui se fédèrent et s'emparent d'outils du Web d'aujourd'hui.

1.4.1 Contenu généré par les utilisateurs

Dans la vision Web 2.0, ce qui fait la différence avec le premier âge du Web, c'est avant tout le fait que le contenu des pages est produit par les utilisateurs. Le Web est avant tout un ensemble de pages dont le contenu est statique, rédigé par une personne et accessible par tous les autres mais en lecture uniquement. Avec l'arrivée en 1995 des premiers wikis, mais aussi les forums, il devient possible à tout le monde non plus de lire des pages Web mais de les modifier, les éditer, les commenter, les compléter...

Aujourd'hui, la plupart des applications phares du Web sont basées sur ce concept d'édition ou de création collective de contenu. Les wikis permettent à un ensemble d'utilisateurs de co-rédiger une page. Les forums permettent à un groupe d'utilisateurs de laisser des messages ou Post les uns à la suite des autres pour construire une discussion. Les blogs permettent à chacun de créer une page et aux autres de laisser des commentaires. Les sites d'écoute de musique en ligne permettent d'ajouter des morceaux, de laisser des avis sur la musique, de créer des "Playlists" ou collections de musique. Les sites de vente en ligne permettent de laisser son avis sur un produit acheté ou sur la réactivité, la qualité du service d'un vendeur. Il est possible sur n'importe quelle page de dire que l'on aime cette page via le "i like" Facebook. Certains diront que l'utilisateur n'est plus consommateur mais aussi acteur. [De Rosnay, 2006] dira même qu'il devient "consom-acteur".

Tous ces usages sont des exemples du concept de base du Web 2.0 de l'amélioration du service par l'usage. L'utilisateur d'un service crée, par l'usage, de nouvelles données qui sont au cœur de l'amélioration du service (par exemple, les systèmes de recommandation de vente en ligne analysent l'historique des achats, les traces de l'usage pour recommander d'autres achats pertinents). L'utilisateur par l'usage contribue à l'amélioration du service. On parle aussi d'utilisateur comme ressource computationnelle quand le système délègue certaines tâches aux utilisateurs. Ainsi, le programme et l'utilisateur créent une sorte de symbiose un peu comme l'envisageait "ce bon vieux Lick" [Licklider, 1960]. Par exemple, afin de vérifier lors d'une inscription que la création de compte est bien réalisée par un humain et non par un programme, les applications web demandent souvent aux utilisateurs de saisir les caractères qu'ils lisent dans une image présentant des caractères alphanumériques déformés. Ce système est appelé *Captcha*. Ces images sont souvent des bords de

pages de livres scannées et les caractères sont déformés par la courbure de la page. Il est impossible, ou très difficile actuellement pour un programme de reconnaître ces caractères. Du coup, le système de Captcha a deux intérêts :

- puisque seuls des humains peuvent reconnaître les signes déformés, le programme peut trier lors de l'inscription les humains des programmes malveillants.
- lorsque le programme demande à un humain de s'inscrire, il présente aussi dans le Captcha des images de caractères déformés que le programme ne sait pas reconnaître. Ainsi, petit à petit, la somme des humains s'inscrivant vont lui permettre de déchiffrer les images de bords de livres.

A chaque inscription, le programme vérifie que l'utilisateur est bien un humain avec une partie du Captcha que le programme connaît bien et il lui demande la signification d'une partie que le programme ne connaît pas déjà, augmentant ainsi sa base de Captcha connus. Donc, lorsque vous saisissez à l'inscription les caractères que vous reconnaissiez dans l'image, vous contribuez à renseigner un système de reconnaissance de caractères déformés. Ceci illustre bien le concept d'utilisateur comme ressource computationnelle. L'utilisateur n'est pas toujours conscient de la portée réelle de la tâche qu'il accomplit. Et il peut contribuer à enrichir un service sans même s'en rendre compte. Le tagging est un autre exemple. Finalement, l'utilisateur est souvent plus efficace pour produire une catégorisation de textes que les méthodes actuelles de fouille et analyse de texte.

Un autre effet de la possibilité de créer et d'édition des pages est de permettre une décentralisation du savoir. Les ressources du Web sont pratiquement accessibles à tous. Et si elles ne le sont pas, elles le sont en tout cas pour un groupe de personnes qui peut s'approprier le contenu de ces ressources et le reproduire, créant d'autres ressources qui elles, peut-être, seront accessibles à tous. Ainsi, le fait que les utilisateurs puissent être producteurs de ressources permet un essaimage du savoir et empêche une forme centralisation et de propriété des données. Autrement dit, empêche une forme de contrôle de l'information. Dans le contexte des débats 2.0, ce point est important car il permet aux participants de s'informer convenablement. Ils peuvent se transmettre du savoir en recommandant la consultation de ressources en ligne. Ce point est tout de même à nuancer en prenant en compte les problèmes de fracture numérique et de barrière de la langue. L'information reste accessible pour ceux qui peuvent et savent utiliser Internet.

1.4.2 Débat à grande échelle

Le débat est étudié et pratiqué depuis l'émergence de la démocratie en Grèce au Vème siècle avant notre ère. Le débat est un moyen adopté pour permettre à chacun de participer dans les processus de décision qui le concernent. L'une des difficultés est de satisfaire "le plus grand nombre". La palabre [Bidima, 1997] est un exemple de processus bien rodé permettant à plusieurs individus d'une même société de prendre la parole, d'exprimer son opinion puis de trouver un consensus. Il existe des solutions pour prendre des décisions collectives dans un petit groupe. Mais il est techniquement très difficile, voir impossible de

mettre en œuvre des solutions pour la concertation d'un grand nombre d'acteur. Le problème principal est celui du temps que nécessiterait d'accorder à chacun un temps de parole. L'autre problème est la disponibilité des acteurs durant la concertation, notamment si les acteurs ne se trouvent pas dans une même zone géographique. Finalement, le coût de la concertation est souvent évoqué comme frein à sa mise en place.

Actuellement, les démocraties occidentales ont recours à la représentativité pour pallier ces problèmes. Un groupe délègue à une personne le soin de défendre les intérêts du groupe et lui confère un pouvoir d'action, de prise de décision. La position de représentant est la plupart du temps légitimée par un vote. Le représentant est alors dit "élu". Le groupe délègue une fonction à l'élu qui est censé agir dans l'intérêt du groupe. Cependant, les formes de contrôle à disposition du groupe reposent la plupart du temps sur une simple relation de confiance et le lien entre une décision prise par un élu et les idées ou opinions partagées par le groupe est rarement explicite.

L'une des attentes de la démocratie électronique est d'augmenter le champ de la concertation et de garantir la transparence des systèmes administratifs et politiques en rendant au public sa place dans les débats (notamment en lui donnant la possibilité de questionner, de demander des débats). Internet devient le premier canal de communication et vient naturellement remplacer ses prédécesseurs (le télégraphe, la radio, la télévision, le téléphone) au statut de champ d'expérimentation de la démocratie électronique. Le Web est parfois cité comme étant le nouvel espace public, "un des concepts fondamentaux de la démocratie. Il désigne le lieu symbolique où peuvent s'exprimer toutes les opinions qui structurent le jeu politique, où l'on traite des questions relevant de la collectivité" [Testard-Vaillant, 2008].

Cependant, pour garantir la démocratie, il ne suffit pas d'un espace public. Il faut aussi que le citoyen puisse être "éclairé", correctement informé. Il faut donc lui garantir l'accès à "la bonne information". Hors, à l'heure du Web, trouver la bonne information n'est pas une mince affaire. Les capacités cognitives de l'individu restent limitées alors que la quantité de données sur le Web ne cesse d'augmenter. Évaluée à 281 exaoctets en 2007 (soit 281 milliards de gigaoctets), la taille du nombre de ressources disponibles en 2011 devrait atteindre les 1,8 zettaoctet (soit 1800 exaoctets) [Gantz *et al.*, 2007]. Il faut donc faciliter la tâche de tri et d'évaluation de la qualité de l'information pour le citoyen, car celui-ci, "(...) pour être parfaitement éclairé, (...) *a besoin d'information mais aussi d'information sur l'information*" [Vedel, 2003].

Un débat à grande échelle est donc pour nous un débat permettant à un grand nombre d'acteurs de participer à la concertation. Un débat à l'échelle d'une collectivité territoriale est un débat à grande échelle. Un débat à grande échelle fait nécessairement intervenir un grand nombre d'information produit pour et par les participants du débat. Les débats à grande échelle se heurtent à des problématiques de coûts et de logistiques (spatialité, temporalité). Ils doivent donner l'opportunité aux citoyens de s'exprimer sur de nombreux sujets mais aussi de construire des jugements collectifs. De par la représentativité des ju-

gements produits par des débats à grande échelle, ceux-ci devraient exercer une véritable influence sur les décisions politiques.

1.4.3 Exprimer et représenter des opinions

L'un des points d'entrée fondamental de notre approche est celui de l'expression et de la représentation des opinions. Par opinion, nous entendons un acte discursif ou l'énonciateur "modalise" explicitement l'objet de son énonciation avec les dimensions "possible/impossible", "souhaitable/non souhaitable", "beau/laid", "agréable/désagréable", etc. Nous adoptons la définition de [Chabrol et Bromberg, 1999] à ceci près que nos utilisateurs modalisent toujours explicitement l'objet de leur énonciation, et non parfois implicitement. Dans le cadre d'un outil informatique, adopter une forme de modalisation implicite crée de nouvelles difficultés techniques qui sont en dehors du cadre de ce travail. Modaliser implicitement l'objet des énonciations revient à exprimer un jugement par la saisie d'un texte libre, non pré-formaté. Il devient alors nécessaire de se doter de mécanismes d'extraction de la modalité du discours. Des techniques d'extraction de polarités dans des textes existent et relèvent du domaine du Traitement Automatique de la Langue Naturelle (TALN). Nous avons opté pour une approche plus simple et bien adoptée par les utilisateurs : exprimer les opinions à l'aide d'un formulaire de saisie. Concrètement, nos utilisateurs sont encouragés à préciser la modalité de leurs opinions par le choix entre des modalités fixées telles que "je suis d'accord", "je ne suis pas d'accord", "j'aime", "je n'aime pas", "je comprends", "je ne comprends pas" ... Cette pratique est proche de celle du "i like" de Facebook, pratique bien adoptée par les utilisateurs, comme nous le discuterons plus tard dans ce document.

La discussion dialectique Aristotélicienne définie par [Vignaux, 1988] souligne l'importance de l'opposition. L'opposition est exprimée entre des termes traduisant l'affirmation et la négation, la victoire et la défaite. Une opposition marque un désaccord entre deux parties. Ces oppositions sont exprimées par le biais de termes particuliers permettant aux utilisateurs de préciser leurs opinions. Nous avons adopté l'expression d'opinion comme étant la fonctionnalité centrale d'un outil de concertation. Nous défendons que cette information est centrale pour déduire des relations entre des arguments dans le débat, entre des individus et de faciliter la navigation. Pour nous, les oppositions sont ce qui structure un débat 2.0.

Les outils classiques de discussions permettent de structurer les discussions sous la forme de fils de discussions ou d'arbres de discussions. En règle générale, les "Posts" ou commentaires saisis par les utilisateurs dans les discussions s'enchainent avec un statut de "réponse". Dans un Forum, un premier utilisateur crée une discussion en lui donnant un titre puis en laissant un premier Post. Ensuite, les autres utilisateurs peuvent cliquer sur "répondre" pour laisser un nouveau Post. Si les Posts se suivent tous, on parle de fil de discussion. Si les Posts créent des branches à chaque nouvelle réponse, on parle d'arbre de discussion. La figure 2.4 récapitule les deux types de structure possibles pour les forums. La

figure 1.3(a) présente une capture d'écran d'un fil de discussion sur le site Désirs d'avenir (<http://www.desirsdavenir.org>). La figure 1.3(b) présente une capture d'écran d'un arbre de discussion sur le site FreeRepublic (<http://www.freerepublic.com>).

Dans un Blog, le point de départ de la discussion est un article. L'auteur de l'article a

(a) Un fil de discussion sur Désirs d'avenir

(b) Un arbre de discussion sur FreeRepublic

FIGURE 1.3 : Les deux types de structure dans les forums

donc un rôle différent que les autres utilisateurs du Blog. Il est auteur dans le blog et le blog lui est en général associé. Les autres utilisateurs peuvent lui laisser des commentaires ou Posts en les associant à son article. Puis ils peuvent associer des Posts réponses aux autres Posts exactement comme sur un Forum. La figure 1.4 présente la fin d'un article et la discussion qui lui est associée. Dans ce blog, les discussions peuvent être structurées en arbre, ce qui se traduit par un décalage des réponses vers la droite.

A la différence des Blogs ou Forums, les outils de messagerie instantanée permettent de discuter de façon synchrone. C'est à dire que les utilisateurs prenant part à une même conversation sont connectés en même temps à Internet. Pour les Blogs ou Forums, la discussion est qualifiée d'asynchrone puisque chacun peut répondre à l'autre lors de sa connexion à Internet. Ils ne sont pas censés être connectés simultanément pour discuter. Dans une conversation via messagerie instantanée, les messages s'enchaînent par ordre

1.4. DÉBATS 2.0

23

Du coup en modifiant mon rdf on peut dire que Julien Cotret est un collègue de travail.

```

<rdf:Person rdf:about="http://www.natoline.fr/#Julien_Cotret">
  <rdf:name xml:lang="fr">#Julien Cotret</rdf:name>
  <rdf:accountname>JulienCotret</rdf:accountname>
  <rdf:friendname>Julien</rdf:friendname>
  <rdf:title>rgbd Shudeurs</rdf:title>
  <rdf:accountname>rgbd</rdf:accountname>
  <rdf:currentProject rdf:resource="https://intermeds.coreplay.org/tiki-index.php?"/>
</rdf:Person>

```

Mes comptes sur le web

Il reste un dernier point, l'une des problématiques du web actuel avec la prolifération des applications de social networking est l'identité fragmentée des utilisateurs. Etant utilisateur de Facebook, de twitter, d'un blog, de plusieurs forums, je me retrouve avec autant de comptes utilisateurs différent que d'applications web que j'utilise. Du coup, j'en ai envie de dire que tous ces comptes me sont associés. Il est possible de le faire grâce à la propriété `foaf:account`.

Par exemple je déclare mon compte twitter :

```

<rdf:Person rdf:about="http://www.natoline.fr/#Julien_Cotret">
  <foaf:accountname>JulienCotret</foaf:accountname>
  <foaf:accountname>rgbd</foaf:accountname>
</rdf:Person>

```

Pour conduire ce tutoriel, vous trouverez mon FOAF complet en ligne. Vous pouvez l'explorer avec FOAF explorer qui permet d'explorer de proches en proches dans le réseau. Et essentiel pour tester votre RDF/XML, il existe un validateur W3C.

Mots-clés : FOAF, ontologies, RDF, réseaux sociaux, semantic web, web social

Cet article a été publié le Lundi 28 décembre 2009 à 14 h 42 min et est classé dans [RDF](#), [Tutoriels](#). Vous pouvez suivre les commentaires par le biais du flux [RSS 2.0](#). Vous pouvez laisser un commentaire, ou faire un [trackback](#) depuis votre propre site.

5 commentaires sur "utilisation d'un vocabulaire RDF - le cas de FOAF"

natoline dit :
28 juin 2010 à 17 h 11 min
Un générateur de FOAF: <http://natoline.fr/>

Répondre

ROF et les débats en ligne « natoline dit :
28 juin 2010 à 18 h 52 min
[...] Je présente dans un article précédent de ce blog un cas d'utilisation du vocabulaire FOAF. [...]

Répondre

Catherine Merlet-Poirier dit :
28 juin 2010 à 22 h 40 min
Bravo pour ce tutoriel lumineux, je la fois très progressif et complet !
Je trouve aussi qu'il manque de documents en Français pour bien comprendre et user pratiquer, ou à recommander en formation pour ceux qui veulent comprendre et se lancer.
J'attends les suivants avec impatience ...

Répondre

admin dit :
28 juin 2010 à 22 h 45 min
Merci pour l'encouragement à continuer. Bientôt la suite avec SIOC.
Répondre

Cactus Acidé » = L'observatoire du neuromançonn 01/02/2010 dit :
28 juin 2010 à 23 h 35 min
[...] utilisation d'un vocabulaire RDF - le cas de FOAF + natoline [...]

Répondre

FIGURE 1.4 : Discussion sur un article dans un blog

chronologique. La plupart des outils de messagerie instantanée permettent de stocker un historique des conversations. Celles-ci ne sont pas structurées, comme les fils de discussion, que par l'ordre chronologique des différents messages qui les composent.

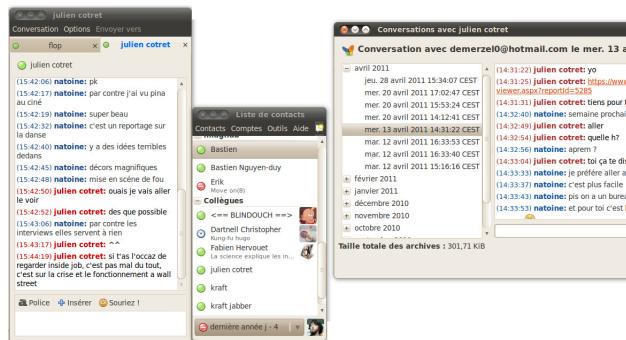


FIGURE 1.5 : Discussion, listing des contacts et historique dans un outil de messagerie instantanée

1.4.4 L'existant

Avec Internet, et plus particulièrement le Web, les utilisateurs ont aujourd'hui l'habitude d'exprimer leurs opinions sur tout. En particulier dans le domaine de la vente en ligne. Prenons l'exemple d'Amazon, premier site de vente en ligne, après avoir effectué un achat, l'utilisateur reçoit un mail lui demandant de revenir sur le site commenter la qualité du service. L'utilisateur de retour sur le site peut alors associer une note à son vendeur. Cette note servira ensuite, additionnée aux autres notes d'utilisateur, à évaluer le vendeur et le classer par rapport aux autres. Permettant ainsi de lui associer un indicateur de confiance pour les prochains acheteurs. L'expression de l'opinion de l'acheteur sur la qualité de service rentre ici parfaitement dans ce que nous avons présenté précédemment, l'amélioration du service par son usage.

Maintenant, il faut souligner que les systèmes de vente en ligne sont probablement les systèmes les plus en avance sur ce point. Cette pratique tend à se diffuser, à devenir une bonne pratique pour tous les services en ligne. Cependant, on trouve d'autres solutions plus classiques, sur le Web pour exprimer des opinions. De nombreux sites proposent aux utilisateurs de laisser leurs "feedbacks", leurs retours ou impressions, dans des forums. L'utilisateur est moins guidé que dans les systèmes d'évaluation de service. Il peut ici rédiger un commentaire à sa convenance. Cette solution est très adaptée pour récupérer de nouvelles idées, des opinions que l'on aurait difficilement pu anticiper. Pour la qualité d'un service il suffit de dire que le service a été bien rendu ou mal rendu. Mais dans de nombreux contextes, la question est plutôt : "que faire pour améliorer le service ?". Et dans ce cas, un outil de type Forum semble plus approprié. La même logique vaut aussi pour les blogs, ou à chaque article, les lecteurs peuvent laisser des remarques sous la forme d'un commentaire libre. Il ne s'agit pas là d'évaluer uniquement si le lecteur a aimé ou non l'article, si il est d'accord avec le contenu ou non, il s'agit d'instaurer un dialogue, de faire évoluer son opinion et celle de l'auteur de l'article. Cependant, que ce soit pour les blogs ou forums, ces solutions impliquent que le traitement des commentaires soit fait par un agent humain. Un robot pourra difficilement analyser le contenu d'un commentaire et proposer une réponse. Les systèmes de consultation d'opinions à partir de commentaires libres impliquent donc un coût de traitement humain non négligeable.

L'autre solution, la plus classique dans les enquêtes d'opinion est celle du sondage. De nombreux outils de sondage en ligne existent, proposant tous à peu près les mêmes fonctionnalités : rédaction des questions et des choix de réponse, envoi de mail d'invitation à participer, récupération des données dans un format type tableur, traitement des données ou représentation sous forme de graphiques de certaines des données. Ce système implique un faible coût de traitement après sondage. Par contre, il nécessite de bien travailler ses questions pour influencer le moins possible les réponses et il fixe généralement les réponses possibles pour les sondés. On a donc les problèmes totalement inverses des outils de consultation à partir de commentaires libres.

Enfin, il reste que certaines tendances, ou opinions collectives peuvent être mesurées

à partir d'analyse de réseaux. Les communautés en ligne, concept établi dès les premières heures d'Internet sont des groupes de personnes partageant des intérêts communs. En règle générale, les membres d'une communauté partagent donc une même opinion. Déetecter ces communautés est donc une bonne solution pour évaluer l'importance d'une opinion partagée. De nos jours, la plupart des outils de réseaux sociaux encouragent les utilisateurs à préciser leurs centres d'intérêts. Ainsi, il est possible de cibler ces individus. Par exemple sur Facebook, il existe un système permettant aux utilisateurs de créer des groupes et d'en devenir membre. Il est aussi possible de créer une page est d'en devenir "fan", marquant ainsi son soutien à l'objet de la page. De nombreux partis politiques et associations ont adopté ce système pour rassembler des militants.

Tous les outils précédemment présentés succinctement, permettent de consulter l'opinion d'un ensemble d'individus. Cependant, seuls ceux dédiés à la discussion, comme les forums ou les blogs permettent véritablement la mise en place de dialogues. Ceci est tout de même à nuancer puisque aujourd'hui, la plupart des outils de réseaux deviennent de nouveaux outils de communication. Facebook n'est pas juste un carnet d'adresse, il permet de discuter par messagerie instantanée, de laisser des messages aux autres et d'y répondre. Si notre objectif est de proposer de nouvelles solutions de débats en ligne à grande échelle, il nous faut porter une attention particulière aux méthodes d'évaluation d'opinions collectives à moindre coût mais aussi aux outils de discussion.

Il nous semble important aussi de souligner un autre problème des outils de consultation en ligne actuels. Ces outils ne sont généralement utilisés que dans deux cas : soit une entité (institut, collectivité, société ...) est demandeur d'une évaluation de l'opinion dans un but bien précis (cibler un nouveau marché, améliorer la qualité de son service, améliorer un indice ...) et ne se soucie donc pas des idées mais juste d'évaluer l'impact de sa politique, soit un groupe déjà existant a décidé d'utiliser Internet pour améliorer sa communication interne et externe. Dans le premier cas, l'opinion est vue comme un indicateur de réussite et sa mesure est figée dans le temps. Dans le deuxième cas, l'opinion collective est déjà bien établie et le but n'est pas de la remettre en question mais de la renforcer. On observe sur le Web que la plupart des sites communautaires tendent à renforcer, à radicaliser l'opinion partagée de la communauté. La communauté ne crée un site que pour elle et s'isole, on parlera d'insularisation. Autre phénomène fortement lié, on observe que les groupes d'opinions les plus marginaux sont les plus représentés sur le Web. Le Web est un moyen pour eux de communiquer sans la censure traditionnelle des médias. Enfin, dernier point noir de cette insularisation, les sites communautaires d'opinion sont en règle générale non liés à une institution. Comme nous l'avons déjà dit, ils ne sont pas créés et donc pas soutenus par des institutions, des décideurs. Mais en plus, la plupart du temps, l'information générée dans ces sites ne remonte même pas jusqu'aux institutions et décideurs.

1.4.5 Collectivités 2.0

Nous souhaitons dans cette partie souligner l'importance des institutions que sont les collectivités territoriales. Dans la même logique que pour les débats 2.0, les collectivités 2.0 sont les outils et moyens du Web mis au service des collectivités pour améliorer leur qualité de service. Si l'on se réfère à l'article 72 de la constitution Française pour définir une collectivité territoriale :

Les collectivités territoriales de la République sont les communes, les départements, les régions, les collectivités à statut particulier et les collectivités d'outre-mer régies par l'article 74. Toute autre collectivité territoriale est créée par la loi, le cas échéant en lieu et place d'une ou de plusieurs collectivités mentionnées au présent alinéa.

Les collectivités territoriales ont vocation à prendre les décisions pour l'ensemble des compétences qui peuvent le mieux être mises en œuvre à leur échelon.

Dans les conditions prévues par la loi, ces collectivités s'administrent librement par des conseils élus et disposent d'un pouvoir réglementaire pour l'exercice de leurs compétences. [...]

Le rôle d'une collectivité territoriale est d'améliorer les conditions de vie des citoyens d'un territoire. Cela passe donc par l'aménagement du territoire, par une mise en place de services aux citoyens. Les secteurs d'intervention sont :

- le développement économique
- les politiques sociales spatialisées
- les politiques du logement
- le développement des infrastructures
- la disponibilité des ressources en eau et leur gestion intégrée afin d'assurer leur durabilité
- la préservation et la mise en valeur de l'environnement comme on la conçoit par exemple dans la gestion intégrée des zones côtières

Une collectivité territoriale est ancrée sur un territoire. Elle est donc liée à un espace, une zone géographique. Ceci se traduit par une ontologie du territoire, un ensemble de cartes et de plans, une histoire locale ... Autant d'éléments qui auront un impact sur les outils mis en place en ligne.

Une collectivité territoriale est toujours dirigée par un ou plusieurs élu(s). Il ne faut donc pas perdre de vue que les actions de la collectivité sont toujours liées aux décisions politiques de ses élus. Et les moyens d'action de la collectivité sont : un pouvoir réglementaire au niveau de l'urbanisme (elles construisent et subventionnent en fixant des règles). Elles font, construisent, mettent en place des services en utilisant les moyens financiers que leurs procurent la fiscalité locale et les dotations financières de l'État. Elles disposent de pouvoirs de police en matière de gestion d'espaces publics, de droit social, et de réglementation sur l'urbanisme (gestion de l'espace, environnement, ou de prestations privées ou publiques à certains publics tels les transports, les tarifs de maisons de retraite ou autres

services publics).

Dans l'ensemble de ces actes, la communication avec le public est forte, allant dans certains cas de la simple information à de la concertation-participation voire des démarches de sensibilisation et jusqu'à de la formation ou de l'éducation. Ce qu'il faut retenir donc c'est que l'institution collectivité territoriale (une mairie, un conseil général ou régional,...) est d'un côté à l'écoute des citoyens et de l'autre fortement guidée par la politique d'un ou plusieurs élu(s). Donc son véritable rôle est de faire le lien entre les citoyens et la politique pratiquée. Il y a donc un travail de concertation et aussi un travail de communication et d'information (de sensibilisation). Il faut savoir ce qu'attendent les citoyens et leur communiquer ce que fait la politique et en quoi cela les concerne. Les sensibiliser, c'est aussi leur demander d'adhérer à une action collective qui va ne porter ses fruits que par l'addition des bons comportements individuels (exemple du tri collectif des déchets). Il nous semble alors que les collectivités territoriales sont les premières à avoir besoin d'outils de débats 2.0. Collectivités 2.0 et débats 2.0 sont donc deux concepts très liés. Si aujourd'hui la plupart des collectivités ont bien adopté Internet et le Web pour leur communication, elles sont encore, dans l'ensemble, en retard sur tout ce qui concerne la concertation et la transparence entre politique de l'élu et attentes citoyennes.

1.4.6 Verrous des débats 2.0

Nous revenons ici sur un ensemble de verrous associés aux outils de concertation en ligne que nous avons relevé dans la littérature.

- Le problème de la quantité d'information et de la navigation dans cette quantité d'information. Le citoyen a besoin d'information et de solutions pour filtrer cette information. Si trop d'information lui est proposé, il ne fera pas l'effort de s'informer [[Vedel, 2003](#)].
- L'information n'est pas suffisante, il faut de l'interaction et de la participation. L'interaction permet de manipuler l'information et donc une appropriation plus active du savoir. Trop de sites soi-disant de consultation sont en réalité de très bons sites d'information ne prenant que très peu en compte la participation, les remarques, les avis des citoyens [[Confino, 2009](#)].
- Il ne faut pas oublier que Internet n'est toujours pas accessible à tous pour des raisons techniques, économiques et sociales. Il y a ceux qui peuvent utiliser Internet et les autres. Il ne faut pas exclure des débats ceux qui n'accèdent pas à Internet [[Compiègne, 2011](#)].
- Il ne faut pas que les citoyens débattent juste pour le plaisir de débattre. Si les citoyens ne voient pas de réelles retombées, de réelles prises en compte de leurs avis, ils vont vite abandonner les outils de concertation [[Vedel, 2003](#)].
- Les groupes d'opinion ont tendance à s'insulariser. Les sites de débats sont souvent réservés à une communauté qui n'entretient pas ou très peu d'échange avec d'autres communautés. Cette tendance à se replier sur soi radicalise les opinions des

membres de la communauté. De plus, cette tendance ne permet pas de faire remonter les opinions aux décideurs [Price, 2006].

- Les petites communautés ont tendance à être sur-représentées dans les sites de débat. Elles voient en Internet une solution pour s'exprimer, une alternative aux autres médias qui ne leur donnent pas la parole [Breindl et Francq, 2008].
- Internet court-circuite les leaders de l'information et en crée de nouveaux mais se pose alors la question de leur légitimité [Breindl et Francq, 2008].
- Internet court-circuite aussi les processus traditionnels de publication faisant tomber les barrières de contrôle de la qualité du contenu. Ceci peut supprimer une censure trop contraignante mais supprime aussi tous les gardes fous permettant de vérifier le bienfondé des dires de chacun. Comment alors empêcher quelqu'un de populaire de profiter de sa popularité pour colporter des mensonges? Les règles qui définissent les jeux de popularité ne sont pas claires et le lien entre popularité et qualité de l'information n'est pas assuré [Breindl et Francq, 2008].
- L'effet "*PageRank*" tend à donner raison aux sites les mieux référencés. Les sites les plus anciens, les premiers arrivés ont tendance à être mieux référencé que les derniers arrivés. Ce qui n'est pas forcément un gage de qualité du contenu. La tendance est de donner raison à celui qui a le plus de supporters [Flichy, 2008].
- L'hypermnésie du Web (le Web n'oublie rien, ce qui est dit sur le Web est copié et reproduit) et la possibilité de recouper les informations concernant les individus fait craindre aux utilisateurs la mise en place d'un système de fichage très performant des individus.
- L'anonymat ou pseudonymat (usage d'un pseudonyme) garantit une certaine liberté d'expression sans fichage, permet aux timides de s'exprimer mais déresponsabilise par rapport aux propos émis (problème de qualité de l'information et de confiance en l'auteur) [Revillard, 2000]. Ce qui se traduit souvent par une hausse du taux d'insultes sur les sites avec anonymat par rapport à ceux sans anonymat.
- Le manque de confiance de la part des citoyens envers certaines institutions porteuses des consultations en ligne est un facteur de frein de la participation.
- Le débat en face à face entretient une certaine forme de rapports sociaux alors que Internet et les nouveaux outils de communication auraient tendance à désindividualiser, déshumaniser. Les débats 2.0 risquent de participer un peu plus à la médiation des relations sociales par écrans interposés.

Nous n'avons pas choisi dans ce travail de traiter tous ces verrous, cependant, une bonne connaissance des verrous existants permet de faire des choix stratégiques dans notre approche. Certains verrous peuvent être contournés en choisissant la population ou le sujet de nos expérimentations par exemple.

1.5 Stratégie adoptée

Nous revenons sur les verrous que nous allons aborder et sur la façon dont nous allons les aborder. En particulier, nous mettrons l'accent sur l'interopérabilité entre les outils de concertation et les autres. L'interopérabilité joue sur l'accessibilité [Compiègne, 2011] mais aussi sur la réutilisabilité des données issues de la concertation dans d'autres contextes.

Nous avons décidé dans cette thèse de nous intéresser essentiellement aux verrous concernant les problèmes d'échelle dans les débats 2.0. Nous verrons que deux axes sont prépondérants dans notre solution pour aborder ces verrous : l'interopérabilité et le réseau social sous-jacent aux débats 2.0. Nous reformulons ici les verrous qui nous intéressent en prenant en compte notre stratégie :

1. Un grand nombre de participants débattant sur un grand nombre de sujets entraîne la production d'un grand nombre de données. Ce problème est intrinsèque au Web 2.0. Les outils de discussion pour passer à la grande échelle doivent proposer de nouvelles solutions pour structurer les discussions et pour faciliter le traitement du contenu des discussions. En particulier, elles doivent faciliter l'extraction des opinions collectives exprimées dans les débats.
2. Dans un débat 2.0 se constituent des groupes d'opinion autour d'idées portées par un grand nombre et autour de personnalités influentes. Comment mesurer la légitimité de ces opinions et la légitimité de ces leaders d'opinion ? Les outils de débats 2.0, toujours pour améliorer la navigation dans la grande quantité de données mais aussi pour améliorer la qualité des débats doivent permettre aux participants de se doter de critères d'évaluation de la pertinence des informations et des compétences des individus.
3. Un débat ne sert à rien s'il n'aboutit pas sur une prise de décision. Au-delà de cette considération, la non prise en compte de l'avis des citoyens dans la prise de décision influe sur leur participation. Si ils pensent que leurs avis ne seront pas pris en compte, ils ne les expriment plus. Les outils de débats 2.0 doivent donc favoriser le traitement et la réutilisabilité des données qu'ils produisent concernant les opinions des citoyens. Ce n'est pas à un outil de débat de proposer les décisions à prendre.
4. Les groupes d'opinion ont plutôt tendance à se fermer sur eux-mêmes, à discuter au sein du groupe plutôt qu'entre groupes d'opinions divergentes. Ainsi est observé une insularisation des groupes d'opinion sur Internet et avec celle-ci une radicalisation des opinions. Cette dynamique ne favorise pas l'information des individus, la naissance de nouvelles opinions alternatives et donc n'améliore pas la qualité des débats et ne permet pas non plus de trouver de solutions consensuelles. Les outils de débats 2.0 doivent donc favoriser l'ouverture d'esprit et les discussions entre les groupes plutôt que de favoriser l'insularisation.

Notre démarche s'est essentiellement concentrée autour de deux points : l'interopérabilité et la dynamique du réseau social intrinsèque aux débats. Nous avons donc proposé

un outil de débats permettant une structuration facilitant l'indexation des contenus mais aussi facilitant la définition de relations d'opposition et de soutien entre les participants aux débats. Cet outil, même s'il reste un prototype expérimental, nous a permis de valider des formats de représentation des données des débats afin de favoriser la réutilisation des données de la concertation dans d'autres outils.

CHAPITRE

2

Background

Préambule

Ce chapitre présente tous les pré-requis nécessaires pour pouvoir aborder notre problématique. Nous commencerons par introduire les notions intrinsèques aux débats. Ensuite nous présentons Internet et ses usages. En particulier, nous détaillerons l'importance grandissante de l'application phare d'Internet, le World Wide Web comme plateforme pour des débats 2.0. Nous présenterons ensuite quelques fondamentaux de l'analyse des réseaux sociaux.

Sommaire

2.1	Notions intrinsèques aux débats	31
2.2	Les usages d'Internet	34
2.3	L'interopérabilité	55
2.4	Les réseaux sociaux	61

2.1 Notions intrinsèques aux débats

Dans cette partie, nous allons définir simplement quelques notions fondamentales pour bien apprécier ce qu'est un débat. Ces notions seront notamment essentielles lorsqu'il s'agira de l'implémentation de notre application de débats en ligne.

2.1.1 Visées énonciatives et Actes de paroles

Dans son manuscrit de thèse [[Desquinabò, 2005](#)], Nicolas Desquinabò distingue les "visées énonciatives" des "actes de paroles". Les "visées", définition proposée par [[Charraudéau, 1995](#)], sont l'intention d'influencer l'autre par le moyen d'une action communicative qui est composée d'un ensemble plus ou moins important d'énonciations. Nicolas défend par la suite qu'une visée peut-être partagée par plusieurs prises de paroles ou discours. Plusieurs interventions d'une même personne dans un débat peuvent servir la même visée. Mais également qu'une prise de paroles ou discours peut servir plusieurs visées. Nicolas va par la suite décomposer chaque unité communicationnelle, ou prise de parole, en plusieurs unités plus réduites : des actes de parole. Un acte de parole correspondant de ce point de vue à l'acte de langage de la pragmatique conventionnelle [[Austin, 1970](#)], [[Searle, 1982](#)]. Un acte de parole a une visée illocutoire, le but, l'intention consciente de celui qui énonce. Et l'acte de parole a également un effet perlocutoire, un impact, la réaction, le changement produit sur celui à qui l'on s'adresse.

Nous n'irons pas plus loin dans la définition des actes de paroles et des visées énonciatives. Pour notre travail, il nous semble uniquement judicieux de souligner que toute prise de parole peut-être décomposée en un ensemble d'unités de langages ayant une intention ou visée, et un effet. Le travail de thèse de Nicolas Desquinabò a fortement consisté en une analyse de ces unités de langages dans le cadre de débats télévisés. Nous allons dans notre travail tenter de proposer un cadre, une application, permettant de mettre en exergue ces unités de langages dans des prises de paroles en ligne. L'annotation socio-sémantique va permettre à chacun de décomposer chaque texte. Le travail de Nicolas Desquinabò aurait pu être fait en utilisant notre outil à partir des retranscriptions textuelles des débats télévisés qu'il a analysé.

2.1.2 Argumentation et Arguments

Dans notre travail, nous utilisons souvent les termes "Argument" et "Argumentation". Nous nous référons au sens commun de ces notions. Un Argument est un élément de la discussion venant soutenir ou contrer une position. Une Argumentation est l'ensemble des arguments soutenant ou attaquant une position.

Par exemple, dans un débat concernant les nanotechnologies, un ensemble de positions peuvent être adoptées par les participants. Chacun peut être "pour" ou "contre" les nanotechnologies mais également adopter une position plus modérée comme "être pour si des dispositifs de contrôle de l'impact des nanotechnologies sur la santé sont mis en place". Une personne adoptant la position "pour" les nanotechnologies peut alors dé-

fendre sa position. Elle va devoir apporter des éléments justifiant sa position et pouvant éventuellement convaincre d'autres de la justesse de sa position. Par exemple, une personne "pour les nanotechnologies" pourrait invoquer le fait que : "finalement, il y a déjà des nanotechnologies dans les produits de consommation courantes depuis des années". Cette sentence, "finalement, ... des années", est un argument venant soutenir la position de l'individu. Bien entendu, le fait que cet argument soutienne effectivement ou non la position de l'individu est discutable. Chacun peut alors juger cet argument et estimer qu'il soutient effectivement la position de l'individu ou au contraire, que l'argument est discutable, non fondé, fallacieux ... et qu'il dessert le propos de l'individu.

Le jugement porté par un utilisateur sur un argument est l'expression d'une opinion au sens de [Chabrol et Bromberg, 1999]. Par opinion, nous entendons un acte discursif ou l'énonciateur "modalise" explicitement l'objet de son énonciation avec les dimensions "possible/impossible", "souhaitable/non souhaitable", "beau/laid", "agréable/désagréable", etc.

2.1.3 Parties Prenantes

La dernière notion que nous introduisons ici est celle de "parties prenantes". Les parties prenantes sont les acteurs concernés par une prise de décision. Lorsque l'on considère des débats publics, l'objectif du débat est de permettre aux décideurs de prendre une décision qui prenne en compte l'avis du public. Les parties prenantes sont les individus ou groupes d'individus qui seront impactés par la décision.

Lors des débats nationaux de 2003 sur l'école, les parties prenantes facilement identifiables sont les parents d'élèves et les instituteurs. Le dossier de synthèse¹ fournit la liste des parties prenantes considérées pour ce débat. Cette liste est donnée en fin de dossier dans la partie "Liste des organisations, experts et personnalités entendus". Les parties prenantes suivantes y sont notamment listés (nous ne reprenons pas l'ensemble de la liste) :

- Les anciens ministres de l'éducation nationale.
- Les parties politiques représentées au Parlement.
- Les fédérations de parents d'élèves.
- Les organisations patronales.
- Les confédérations syndicales.
- Les associations d'enseignants.
- L'association des maires de France.
- Les associations de Français à l'étranger.
- Les associations de personnes atteintes de handicap et de parents d'enfants atteints de handicap.
- Les CCI et Chambres de métiers.
- Les IUFM.

1. <http://www.education.gouv.fr/archives/2003/debatnational/upload/static/lerapport/pourlareussite.pdf>

2.2 Les usages d'Internet

Quid d'Internet, qui devait nous apporter la Bibliothèque nationale gratuitement dans nos campagnes reculées (...), faciliter la vie de l'intelligence en mettant le savoir digne de ce nom à disposition de tout le monde ?
Tout passe par le Net et quiconque ne dispose pas d'un ordinateur est un citoyen de seconde zone. Les traces laissées par l'usage de nos ordinateurs servent aux marchands, aux publicitaires, aux polices diverses et multiples. Ne parlons pas de la possibilité pour chacun de dire tout et n'importe quoi, de montrer son indigence sans vergogne pour en informer la planète en temps réel.

Le catéchisme postmoderne, Le Monde 9-10 janvier 2011
MICHEL ONFRAY

2.2.1 Origine d'Internet

Internet est un réseau d'ordinateurs à l'échelle mondiale. Aujourd'hui, Internet permet à des individus du monde entier de communiquer à partir d'ordinateurs, en particulier, à partir d'ordinateurs personnels. Le fonctionnement d'Internet repose sur l'usage de protocoles de communication permettant d'envoyer des paquets de données d'un point du réseau à un autre. Ces protocoles assurent que quelque soit le constructeur de la machine et quelque soit le système d'exploitation installé sur cette machine, celle-ci pourra faire partie du réseau, celle-ci pourra envoyer et recevoir des paquets de données.

La date de création d'Internet est estimée vers la fin des années 1970. En réalité, Internet est un nom donné à un réseau de machines déjà interconnectées au moment du choix du nom. Avant de s'étendre au monde, le réseau était réservé à quelques organismes participant à une expérimentation. Internet regroupait le réseau ARPANET, projet de la DARPA, et d'autres machines, essentiellement dans des universités américaines.

Les idées fondatrices d'Internet sont plus anciennes. En 1968, Taylor et Licklider imaginaient déjà que l'ordinateur en réseau nous permettrait de communiquer [Licklider et Taylor, 1968]. En 1945, Vannevar Bush imaginait pour les scientifiques un outil de communication et de partage d'informations multimédia, le Memex [Bush, 1945]. On retrouve l'idée de créer une Bibliothèque de tous les savoirs accessibles à tous dans des mythes et utopies plus anciens comme le Mundaneum de Paul Otlet [Levie, 2006], le Xanadu de Ted Nelson [Nelson, 1982], ou encore le mythe de la tour de Babel. La liste des œuvres de fictions ayant inspirées des usages d'Internet aujourd'hui et la liste des œuvres de fictions imaginant des usages de demain est longue. Internet est nourri d'un imaginaire collectif ou se dessinent des visions utopistes et dystopiques [Compiègne, 2011]. Les utopistes promettent l'ouverture du savoir à tous, l'échange et le partage libre et sans limite, un monde plus égalitaire et plus coopératif. Et en particulier une démocratie plus participative et interactive. Les dystopiques eux, imaginent une société du contrôle où tout le monde est surveillé par tout le monde [Lavelle, 2009], Internet est suspecté créer des addictions, la médiation technique déshumaniserait le tissu relationnel, le contenu n'est pas fiable ... L'article de Michel Onfray dans le Monde de janvier 2011 résume assez bien cet ensemble de craintes.

Aujourd'hui, les usages montrent que les idées fondatrices d'Internet sont réalistes mais qu'elles nécessitent des progrès autres que techniques. Le Web originellement conçu pour la communication et le partage de connaissances, à des fins égalitaires, pour faire fi de la censure, libérer l'expression, est essentiellement utilisé dans le cadre d'une activité de loisirs. Les enjeux financiers et politiques d'Internet renforcent les visions dystopiques. Le rêve de la grande Bibliothèque de tous les savoirs sera probablement l'œuvre d'une société privée. Les données même les plus personnelles sont finalement des valeurs marchandes. Et les gouvernements n'hésitent pas à adopter des stratégies de filtrage d'accès à certaines ressources.

Si dans cette thèse, nous nous attaquons aux solutions techniques nécessaires pour la mise en place de nouveaux outils de concertation, il nous semblait devoir signaler que pour qu'Internet devienne un véritable outil démocratique, il faut une réelle volonté politique et une réelle appropriation de l'outil comme tel par les citoyens. Comme le souligne Benjamin Bayart dans son exposé sur ce qu'est Internet à Sciences Politique², nous pensons que pour parler des impacts politiques et sociaux d'Internet, il faut une certaine connaissance d'Internet (enlever tout ce qui est magique).

Nous énumérons la liste suivante de concepts fondateurs ou intrinsèques à Internet et au Web qui sont des éléments porteurs de démocratie :

- Dès les origines d'Internet, dans les mythes fondateurs comme le Mémex [Bush, 1945], ou dans la vision de Licklider et Taylor [Licklider et Taylor, 1968], on retrouve cette idée de créer du lien social par le biais d'échanges de données. Vannevar Bush pense que l'avenir de la recherche repose sur des outils permettant l'échange de don-

2. <http://www.libertesnumeriques.net/archives/394>

nées et d'opinions sur des sources de savoir. Licklider et Taylor parlent d'un outil de communication pour des communautés d'intérêt faisant fi des contraintes spatiales.

- Le tout numérique ou la possibilité de tout enregistrer et transmettre sous format numérique encourage l'idée de pouvoir, un jour, donner accès à tout le monde à toute source de savoir, de culture ... Aujourd'hui on parle de bibliothèque universelle. Google et l'Unesco, chacun de leur côté, tentent de concrétiser ce rêve de plus en plus abordable.
- La décentralisation des services et des données est l'une des propriétés originelles d'Internet aujourd'hui menacée. L'idée originelle est que les données sur Internet transitent d'une machine à une autre sans avoir à passer par un point central du réseau. Ceci permet de proposer des applications distribuées dans lesquelles aucune machine n'est indispensable. Chacun est libre de choisir ou il va enregistrer des données le concernant et même en rester propriétaire s'il décide d'héberger lui-même le service nécessaire. L'utilisateur est libre du choix des services qui l'intéressent et du choix d'implémentation de ce service. Par exemple, chacun peut choisir le serveur de mail qu'il souhaite et l'installer soi-même, ceci ne limite pas l'envoi de mails à un groupe restreint aux utilisateurs du même serveur (quelqu'un ayant une adresse chez *Gmail* peut très bien écrire à quelqu'un ayant une adresse chez *Free*). Aujourd'hui, des services très populaires comme *MSN* ou *Facebook* ont tendance à recentraliser les données sur Internet. Il est impossible de communiquer avec le réseau d'utilisateurs de *MSN* ou *Facebook* sans être soi-même membre de ces réseaux.
- Internet et le Web sont des plateformes. L'un comme l'autre permettent de déployer des services se basant sur leurs protocoles (respectivement ip et http). Ainsi, il est possible de proposer une multitude de services servant des usages différents. Par exemple, il existe de nombreuses solutions de communications synchrones ou asynchrones, comme le mail, la messagerie instantanée, les blogs, les forums, les outils de réseaux sociaux numériques ... Tout le monde peut techniquement développer et déployer les services qu'il souhaite.
- Le réseau Internet n'est pas mono-directionnel. Tout point du réseau, toute machine, peut transmettre des données à toute autre machine du réseau. C'est en cela qu'Internet est un outil de communication $n * n$ (de tout le monde vers tout le monde) à l'opposé des outils de communication $1 * n$ (un seul émetteur pour n récepteurs) comme la radio ou la télévision.
- Il est très facile de copier une donnée sur le Web et la durée de vie d'une donnée devient quasi éternelle puisque elle va dépendre de la durée de vie de chacune des copies. Ainsi, si une donnée est publiée sur le Web, elle peut rester accessible très longtemps.
- Encore actuellement, la censure est peu présente sur le Web car elle est difficile à mettre en place (cependant, ce point est régulièrement attaqué par des lois visant à rendre responsable les hébergeurs de contenu). Il est facile à un fournisseur d'accès Internet (FAI) de supprimer l'accès à une ressource Web que vous publiez sur l'un

des serveurs qu'il contrôle. Cependant, il faut encore que ce fournisseur d'accès ait conscience de l'existence de cette ressource.

- Encore aujourd'hui, le réseau est censé être neutre. C'est à dire qu'aucune discrimination ne peut être faite envers l'une des machines du réseau en terme de débit et de droit d'accès. Quelque soit le point du réseau d'où part la donnée et quelque soit le point où arrivent les données, les données transitent sans privilège ou restriction particulière (pas d'augmentation ou de diminution du débit par exemple en fonction de la source ou de la destination).
- Le droit à l'expression anonyme est fortement assuré par de nombreux outils du Web 2.0. Par exemple, le premier wiki³ [Buffa, 2008], celui de Ward Cunningham permettait à tout le monde de contribuer anonymement. Ce principe est encore valable dans des forums, des blogs, des wikis ...

Nous venons de lister un ensemble de concepts fondamentaux d'Internet qui devraient être à la base d'outils dédiés à la démocratie électronique : le libre accès pour tous à tout un ensemble de données conservées très longtemps garantissant la transparence, la possibilité d'échanger et partager entre individus par le biais d'un réseau (un réseau d'ordinateur mais aussi un réseau d'utilisateurs, un réseau social), la neutralité face au service ou l'égalité d'accès pour tous, la liberté d'expression et de publication, le droit à l'anonymat, le droit de choisir entre plusieurs services concurrents ...

Nous nous intéressons essentiellement à l'apport d'Internet pour la démocratie électronique, et essentiellement pour l'amélioration des processus de concertations. Cependant, on ne peut pas aborder ce point sans un minimum de connaissances du fonctionnement d'Internet et de son application principale, le World Wide Web. Nous mettrons l'accent sur l'usage d'Internet dans les tâches collectives, qu'elles soient coopératives ou collaboratives, en soulignant les usages dédiés à la communication et au partage de ressources. Nous présenterons ensuite le concept de Web2.0 et ce qu'il a changé dans la vision du Web. Et enfin, nous présenterons quelques usages du Web qui nous semblent avoir un fort potentiel pour la démocratie électronique.

2.2.2 Le fonctionnement d'Internet et du Web

Dans cette partie, nous allons présenter les mécanismes fondamentaux qui permettent de comprendre le fonctionnement d'Internet et de son application principale, le World Wide Web. Nous présenterons donc le protocole IP et expliquerons le fonctionnement du routage et des envois de paquets qui sont la base du fonctionnement d'Internet. Puis nous présenterons le protocole HTTP qui a permis de bâtir le World Wide Web.

3. <http://c2.com/cgi/wiki>

Internet

Internet est avant tout un réseau informatique, fait de machines et de *tuyaux* (à l'origine que des câbles, aujourd'hui avec les technologies sans fil, on peut se passer de certains câbles, donc on adoptera le terme de tuyau qui induit l'idée de flots entre les machines). Un réseau informatique est un ensemble de machines connectées les unes aux autres par des tuyaux sur lesquels peuvent transiter des données. Les machines peuvent donc communiquer les unes avec les autres en envoyant des données. Cependant, toutes les machines du réseau ne sont pas connectées directement les unes aux autres. Chaque machine, que l'on considérera désormais comme un nœud du réseau, n'est connectée qu'à un certain nombre d'autres nœuds du réseau. Et du coup, pour qu'un nœud A envoie des données à un nœud B auquel il n'est pas directement connecté, les données devront passer par un ou plusieurs nœuds intermédiaires. Toutes les machines du réseau sont reliées mais pas directement. Entre A et B, il y a des intermédiaires, ce qu'il illustre la figure 2.1.

Lorsqu'un nœud du réseau veut envoyer un message à un autre nœud du réseau, il découpe la donnée du message en plusieurs paquets de données et adresse cet ensemble de paquets à l'autre nœud. L'autre nœud, celui qui réceptionne les paquets, vérifie qu'il a bien récupéré tous les paquets et agrège leurs contenus pour retrouver la donnée qui lui est adressée. Les paquets étant numérotés, si il advenait qu'un paquet se perde, qu'il n'arrive pas à destination, le nœud à la réception pourrait adresser un nouveau message au noeud chargé de l'envoi afin de lui demander le paquet manquant. Il faut bien souligner ici que le système d'envoi de paquets est symétrique. Il n'y a pas de nœuds dans le réseau qui sont chargés d'envoyer des paquets et d'autres de les recevoir. Il y a bien communication dans les deux sens. Tout nœud du réseau peut envoyer un message, donc un ensemble de paquets et en recevoir. Voici un exemple de communication entre deux nœuds A et B d'un réseau :

1. A envoie un message à B. Donc un ensemble de paquets de données est envoyé de A à B pour coder le message.
2. B reçoit les paquets de A. Il agrège les paquets et découvre l'intégralité du message de A.
3. B renvoie une réponse à A. Donc un ensemble de paquets de données est envoyé de B à A pour coder la réponse.
4. A reçoit les paquets de B mais il en manque. A renvoie un message à B lui demandant les paquets manquants.
5. ainsi de suite ...

La commutation par paquets est une des nouveautés d'Internet. Ses ancêtres, le téléphone ou le minitel, fonctionnaient sur le principe de la commutation par circuits. Avec le téléphone par exemple, lorsque une communication entre deux nœuds (deux combinés) est établie, le circuit ou chemin reliant les deux nœuds est réservé pour la communication

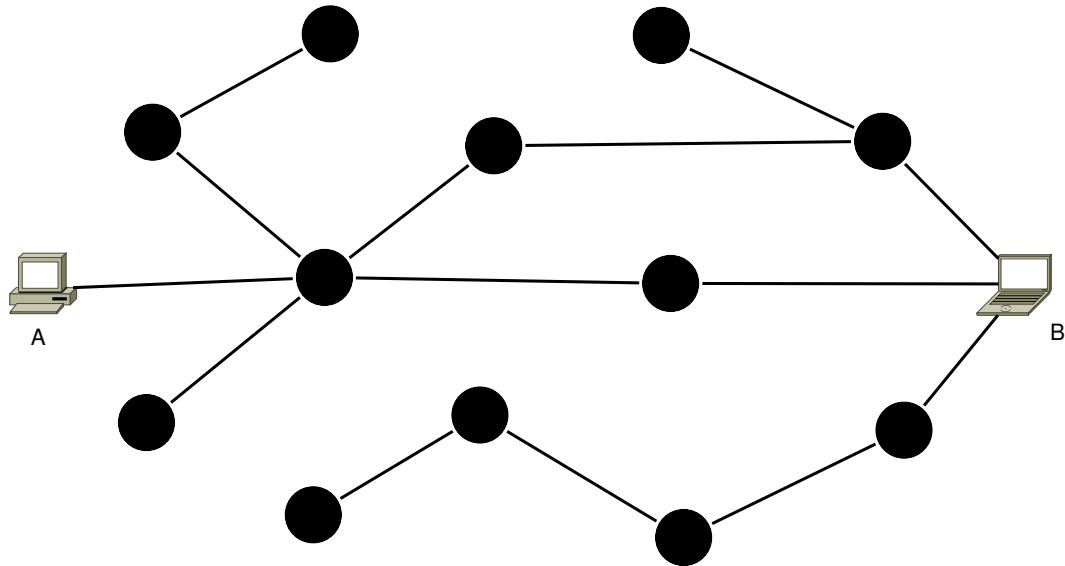


FIGURE 2.1 : Un réseau de machines reliant A et B. Il existe plusieurs chemins pour relier A et B. Les paquets transitant de A à B ne passeront pas tous nécessairement par le même chemin.

du début jusqu'à la fin (le circuit est en fait constitué de trois noeuds, le premier combiné, l'opérateur central et le deuxième combiné. C'est l'opérateur central qui se charge de mettre les deux combinés en communication, c'est lui qui réserve le circuit). Avec la commutation par paquets, le chemin que va emprunter chaque paquet est potentiellement différent. Le choix du chemin est fait pour chaque paquet. Et le chemin n'est occupé que lorsqu'il y a un paquet à faire transiter. La commutation par paquets apporte plusieurs avantages :

- Il est possible pour un noeud A d'établir des communications en parallèle avec plusieurs autres noeuds.
- Les risques de congestion sont diminués puisque la communication n'est plus vue comme continue mais comme une somme de paquets.
- Avec le téléphone, il n'est possible de n'appeler qu'un seul correspondant à la fois. Avec Internet, il est possible de recevoir des mails et de consulter plusieurs sites Web en parallèle, les paquets de données pouvant arriver par plusieurs chemins et s'intercaler à la suite les uns des autres même en provenant de plusieurs sources.

Pour cibler précisément un noeud du réseau, des adresses permettent d'identifier chaque noeud. Ces adresses sont définies par un protocole, le protocole IP. Si une adresse cible précisément un noeud dans le réseau, elle ne définit pas le chemin que vont devoir suivre les paquets. Pour définir ce chemin, il faut établir ce que l'on appelle une stratégie de routage. Le routage étant le mécanisme par lequel sont sélectionnés les chemins, aussi appe-

lés *Routes*. Une première stratégie est celle du routage statique. Chaque nœud du réseau connaît ses voisins et sait par quels nœuds précisément il doit passer pour atteindre n'importe quel autre nœud. Cette solution nécessite une mise à jour à chaque modification du réseau. L'objectif du routage dynamique est justement de proposer une automatisation de ces mises à jour. Les nœuds peuvent échanger leurs connaissances du réseau. Certains noeuds, appelés routeurs, passent leur temps à mettre à jour leurs routes et les échanger avec les autres. Ainsi, de proche en proche, ils remettent à jour leur connaissance des chemins, que l'on appelle *tables de routage*. L'information des routes est alors copiée en plusieurs machines et sans cesse remise à jour. Quand une machine A voudra communiquer avec une machine B, plutôt que de se baser sur sa connaissance a priori du réseau, elle va commencer par demander sa route à son routeur favori disponible.

Internet fonctionne donc sur ce principe d'envoi de paquets et de routage des paquets. Les nœuds du réseau Internet ont tous une adresse IP. La différence entre un réseau local et le réseau Internet, c'est que les adresses IP dans Internet sont *publiques*. Il est important dans un réseau que les adresses IP soient uniques. Sinon elles n'assurerait pas l'identification d'un seul noeud à la fois sur le réseau et la communication serait plus complexe. Internet est donc constitué d'un ensemble d'adresses IP uniques dites publiques (car accessibles par tous via le réseau). Internet est qualifié de réseau des réseaux car en réalité, derrière les adresses IP uniques et publiques peuvent se cacher d'autres réseaux d'adresses IP, cette fois-ci privées (car non accessibles par tous via le réseau mais que par un réseau local). Quand un utilisateur se connecte à Internet, il connecte son réseau local (l'ensemble de ses machines connectées entre elles chez lui avec des adresses privées) à un noeud particulier d'Internet, un nœud fourni par son fournisseur d'accès que l'on appelle une *passerelle*. Une passerelle a une adresse IP publique sur Internet et une adresse IP privée chez l'utilisateur. La passerelle (le plus souvent appelée une Box, exemple la FreeBox, la 9box, la liveBox ... d'où le terme générique de BiduleBox de Benjamin Bayart) permet de servir d'interface au réseau d'adresses IP privées de l'utilisateur. Cette passerelle permet de faire transiter des paquets de données du réseau privé vers le réseau public et vice versa. La passerelle est en réalité connectée à un noeud du réseau public privilégié, le routeur du fournisseur d'accès.

Le système d'adressage du protocole IP dans sa version 4 permet de créer un numéro d'identification pour chaque machine d'un réseau. Le numéro est constitué de 4 nombres séparés par des points. Chaque nombre est compris entre 0 et 255. Ce qui permet de créer des adresses allant de 0.0.0.0 à 255.255.255.255. On parle d'adresses publiques et privées pour parler réciproquement des adresses IP utilisées par des machines du réseau Internet et des machines d'un réseau privé (comme un réseau à domicile). Pour identifier chaque machine, il faut garantir que son adresse est unique dans le réseau. Il n'y a donc pas deux machines sur Internet ayant la même adresse IP. Dans un réseau privé également, chaque machine a sa propre adresse IP. Cependant, certaines adresses IP sont réservées à un usage particulier. Par exemple, toutes les adresses IP commençant par 192.168 sont réservées aux réseaux privés et donc ne peuvent pas identifier une machine du réseau Internet. Au-

jourd'hui on parle de la version 6 du protocole IP. L'une des nouveautés de la version 6 est d'étendre considérablement le nombre d'adresses. Actuellement, avec IP version 4, le nombre d'adresses est trop limité pour que chaque machine soit directement identifiée sur le réseau et c'est pourquoi on passe par le système de passerelles. Une passerelle est une machine dans un réseau privé mais également identifiée et connectée sur Internet (donc ayant également une adresse IP publique). C'est du coup par celle-ci que toutes les machines du réseau privé peuvent accéder aux autres machines d'Internet. Avec IP v6, qui étend considérablement le nombre d'adresses, on pourra augmenter le nombre d'adresses IP publiques attribuées aux utilisateurs d'Internet par les FAI (aujourd'hui, un abonnement Internet dans un foyer, c'est une seule IP publique, celle de la passerelle).

On a dit qu'un nœud peut envoyer des paquets à plusieurs autres nœuds en même temps. Il en va de même pour la réception. Un nœud peut recevoir des paquets provenant de plusieurs autres nœuds en même temps. Dans une communication client/serveur, il est important de souligner que le serveur peut avoir plusieurs clients en même temps, et il faut aussi souligner que le serveur peut rendre plusieurs services différents. Par exemple, un serveur peut être à la fois serveur Web et serveur de mail. Il devient alors nécessaire de préciser pour quel service le client s'adresse au serveur. Pour ce faire, les nœuds sont équipés de "ports". Les ports sont en quelque sorte des portes d'entrée pour la communication de paquets. Si le serveur reçoit des paquets via son port 80 par exemple, il sait que c'est son service Web qui est demandé. Si c'est son port 110, c'est pour son service de mails...

Toute information numérisée peut transiter sur le réseau (de la voix, des images, de la vidéo, du texte ...). Elle est découpée en paquets et peut être envoyée à plusieurs nœuds du réseau en parallèle. Quelque-soit la qualité de l'accès à Internet (wifi, adsl ou plus petites connections), on peut accéder à toutes les données présentes sur le réseau. Internet permet un accès universel à toutes sortes de médias. Le principe de base d'Internet est très simple, beaucoup plus simple que les réseaux précédent comme la téléphonie ou le minitel. Les routeurs pour Internet sont beaucoup moins chers à produire que des routeurs pour la téléphonie. C'est cette simplicité et ce moindre coût qui a permis l'expansion d'Internet.

La commutation des paquets est symétrique, il n'y a pas des nœuds dédiés à l'envoi de message et d'autres à la réception. Il n'y a pas de différence de nature entre les différents nœuds du réseau. Toute machine connectée à Internet peut-être à la fois client et serveur. Les communications entre les machines peuvent se faire en parallèle. Et donc, une machine peut être cliente de plusieurs serveurs en même temps et un même serveur peut avoir plusieurs clients en même temps et proposer plusieurs services. Le réseau est dynamique. En permanence, des nœuds apparaissent ou disparaissent et des chemins sont modifiés. Comme le routage d'Internet est dynamique, le réseau reste robuste (si une machine tombe en panne, le reste du réseau fonctionne). Internet interconnecte des réseaux privés par le biais d'interface, des passerelles, dont les adresses sont publiques.

Internet est devenu l'outil de communication privilégié pour bien des usages et en particulier pour la consultation. Cependant, si la technique est simple et peu coûteuse, il ne

faut pas oublier que ces solutions techniques d'accès à Internet impliquent aujourd'hui des intérêts commerciaux. Les passerelles permettant à un utilisateur lambda de connecter son réseau privé au réseau public qu'est Internet sont gérés par des Fournisseurs d'Accès Internet (FAI). Ces fournisseurs d'accès ont des intérêts commerciaux à défendre et donc appliquent les politiques qui défendent leurs intérêts économiques. En particulier, les FAI ont la possibilité aujourd'hui d'interdire la mise en place de certains services chez leurs clients. Par exemple, il peut être interdit à un client d'héberger un serveur de mail chez lui. Les FAI peuvent filtrer les adresses IP et interdire l'accès à certaines ressources. Ils peuvent aussi filtrer certains paquets. Il reste que les FAI peuvent tracer les communications, conserver les historiques des communications et des paquets qui transitent. Donc l'avenir des usages d'Internet n'est pas uniquement technique, il est aussi fortement juridique, économique et politique.

Le world wide web

Le World Wide Web, plus communément appelé Web existe depuis la fin des années 80 [[Connolly, 2000](#)]. Le Web est l'une des applications d'Internet, avec les mails, la messagerie instantanée (IRC notamment), le partage de fichiers par FTP ... Le Web permet d'accéder à des ressources hébergées sur des serveurs du réseau Internet selon un protocole, le protocole HTTP et via une application, le plus souvent, un navigateur. Chaque ressource pouvant rendre un service différent, le Web est vu comme une plateforme permettant de proposer une multitude de services aux utilisateurs. Et c'est ce qui explique en partie son succès. Le Web est probablement l'application la plus célèbre d'Internet, à tel point que, beaucoup ne font pas la distinction entre le Web et Internet. Nous allons dans cette section revenir sur les fondamentaux du Web, à savoir, le protocole HTTP, le concept d'hypertexte et le langage HTML.

Le principe de base du fonctionnement du Web est assez simple, les pages Web sont en réalité soit des fichiers texte que le serveur envoie tels quels au client (le navigateur) via le réseau, soit des pages "calculées" par des programmes. Dans son document fondateur [[Berners-Lee, 1989](#)] présente une architecture client-serveur de stockage de documents hypertextes. Au moyen de requêtes, le navigateur (jouant le rôle du client) demande l'accès à une ressource stockée sur un serveur de fichier. Le serveur répond à la requête en renvoyant un fichier (s'il existe ou s'il est généré par un programme) ou un message d'erreur s'il n'existe pas. Les bases du protocole HTTP étaient posées.

Une requête HTTP est constituée d'un mot-clef désignant une méthode ou type d'action spécifique au protocole, d'une URL (Unified Resource Locator) permettant de référencer une ressource précise et d'un corps pouvant être vide ou contenir des données. Une URL est constituée du nom d'une machine et du chemin permettant de retrouver la ressource sur cette machine. Le nom de la machine est répertorié dans un annuaire de nom de domaine (DNS pour Domain Name Server) qui lui associe une adresse ip. Le protocole HTTP est construit au dessus du protocole ip. Une fois que l'on a résolu l'URL, donc re-

trouvé la machine, on transmet une requête à cette machine, requête HTTP transmise par paquets de données selon le protocole ip. La machine répond en renvoyant des paquets de données qui, une fois rassemblés constituent la réponse à la requête HTTP. Le protocole HTTP définit un ensemble de méthodes [Berners-Lee, 1992] permettant de demander au serveur d'effectuer des opérations diverses sur la ressource adressée. Nous ne présentons ici que les 4 méthodes centrales permettant de faire du Web une plateforme participative :

- GET : permet de demander l'accès à la ressource adressée par l'URL. La requête que renvoie votre navigateur pour accéder à une ressource.
- PUT : permet de stocker les données passées dans le corps de la requête à l'URL. L'URL doit déjà exister.
- DELETE : supprime la ressource adressée par l'URL. L'URL est invalide pour de futures requêtes.
- POST : crée une nouvelle ressource constituée des données passées dans le corps de la requête. La nouvelle ressource se voit affecter une URL générée par le serveur et renvoyée en réponse au client.

Ces quatre méthodes définissent les opérations de base pour pouvoir créer, accéder et modifier des ressources. Ce sont des opérations indispensables pour faire du Web une plate-forme participative. S'il n'était possible que de consulter des ressources, les utilisateurs ne pourraient pas ajouter de nouveaux contenus et ne pourraient pas créer de liens entre les ressources. C'est l'autre point central du Web, l'hypertexte.

Le terme d'hypertexte est attribué à Ted Nelson en 1965, il le décrit en détail dans [Nelson, 1974]. Cependant, le concept est plus ancien. Dans [Bush, 1945], le Memex permet de créer un index de documents multimédia que l'on peut échanger entre personnes (notamment entre enseignant et étudiant ou entre chercheurs). La navigation ne se fait pas uniquement par des méta-données ou informations sur le document telles que la date, l'auteur, le titre, le type de document, mais aussi par association entre documents. Vannevar Bush imaginait une navigation permettant d'aller d'un document à un autre par le biais de liens entre documents, par des références ou associations. L'hypertexte est ce concept de navigation par liens. Jean Clément décrit l'hypertexte comme étant un ensemble de documents reliés entre eux par des liens permettant de naviguer rapidement dans l'ensemble des documents [Clément, 1994]. Le Web est conçu autour de ce concept d'hypertexte. Le protocole HTTP signifie littéralement protocole de transfert d'hypertextes. Le Web permet donc de naviguer de ressources en ressources par le biais de liens entre les documents et ancrés dans les documents. Il ne manque plus qu'une solution pour représenter le contenu de ces documents afin que tout navigateur puisse les afficher correctement.

Le langage HTML pour *HyperText Markup Langage*, est un langage à balises dédié à la représentation des documents, des pages pour le Web. Les langages à balises fonctionnent sur l'utilisation de marqueurs de début et de fin (les balises) permettant d'encadrer des données et d'en préciser la nature. Les balises servent donc à préciser le sens, la sémantique, des données qu'elles encapsulent. Par exemple, en HTML la balise de début `<title>` et la balise de fin correspondante `</title>` signifient que ce qui est contenu entre ces deux

balises est le titre du document. Ce système de balisage permet de structurer un document et d'écrire des programmes qui interprètent, ou comprennent, ce balisage. Un navigateur Web est un programme comprenant le HTML et proposant une représentation graphique d'un document HTML. Le HTML est donc un langage standard créé pour garantir l'interopérabilité entre les documents Web et les navigateurs. Tout navigateur est censé comprendre le HTML pour permettre d'afficher toutes les pages Web écrites en HTML. En particulier, HTML permet de définir dans un document une ancre, un élément qui permettra de référencer un autre document, une autre ressource Web. Concrètement, la balise `<a>` (et sa balise fermante ``) permet d'encadrer un élément (du texte, une image ...). Cette balise prend un attribut `href` dont la valeur peut être une URL. Dans ce cas, le navigateur rend cliquable l'élément encadré par la balise `<a>` afin que lorsque l'utilisateur clique sur cet élément, il accède à la ressource dont l'URL est la valeur de l'attribut `href`. L'utilisateur va ainsi de clics en clics accéder à plusieurs pages du Web. Il navigue dans l'hypertexte du Web.

Nous venons donc de voir que le Web est une application d'Internet permettant à des machines d'échanger des pages écrites dans un langage à balise, le langage HTML. Ces pages peuvent contenir des liens facilitant le navigation entre les pages. Un ensemble de pages reliées constituant un document hypertexte. Le protocole HTTP garantit la possibilité d'accéder aux pages mais aussi d'en créer et d'en modifier. Le Web permet donc aux utilisateurs de créer et de partager des documents. Ce dispositif technique vient donc considérablement améliorer les possibilités de communications offertes par Internet.

Le Web apporte encore bien plus, le Web n'est pas juste un ensemble de pages reliées entre elles. Le Web est un ensemble de Ressources reliées entre elles. Et si toute page Web est une ressource, il est des ressources qui ne sont pas des pages Web. Se contenter de considérer les Ressources pages Web revient à ne considérer qu'une sous partie du Web, la partie statique du Web. Les pages Web sont stockées telles quelles sur des serveurs. Elles ont, comme toutes ressources, une URL. Par cette URL, des clients peuvent demander et accéder aux ressources page Web. On parle de Web statique car on accède à des ressources statiques. Par statiques il faut comprendre que le serveur renvoie une page Web telle quelle est stockée. Il n'y a pas de traitement supplémentaire et deux clients accédant simultanément à une même page Web auront exactement le même résultat.

Le concept de *Ressource* est finalement assez difficile à définir tant il recouvre de nombreuses entités différentes. Les URI, dont les URL sont un sous ensemble, sont dédiées à nommer toute chose, des pages Web, des objets réels, des concepts ... [Halpin et Presutti, 2009]. Une *ressource* peut-être n'importe quoi qui ait une identité. Par exemple, un service, une image, une page Web, une collection de ressources ... Toutes les ressources ne sont pas accessibles par le biais du réseau Internet (par exemple, les bâtiments, les personnes ...) [Berners-Lee et al., 1998]. Il est impossible d'accéder à la Tour Eiffel via le Web. Cependant la Tour Eiffel a une URI, c'est une *ressource*. Et par le Web, il est possible d'accéder à des représentations de cette *ressource*, comme une image, une page de texte décrivant l'histoire de la tour Eiffel ...

Certaines URL permettent d'accéder non pas à des pages Web statiques mais à des pages Web dynamiques, des pages Web résultant de l'exécution d'un programme dont la réponse peut être différente d'une fois sur l'autre. Par exemple, une requête sur Google renvoie une page Web dans laquelle sont listés les résultats de la requête. Cette page Web est donc le résultat de l'exécution d'un programme.

Cette partie du Web, le Web dynamique, permet de concevoir de nombreux usages du Web. Ce Web n'est plus une simple application d'accès à des données statiques mais est une plateforme sur laquelle peuvent être déployés de nombreux services, des applications ou programmes renvoyant des pages Web résultants de calculs.

2.2.3 Internet et le travail collectif

Internet permet de communiquer et de stocker des données. Le Web permet à ses utilisateurs de créer des ensembles de ressources et de lier les unes aux autres. Certaines de ces ressources sont des pages Web statiques, des représentations de ressources existantes, ou bien encore des services renvoyant des pages Web dynamiques, résultant de l'exécution d'un programme, d'un calcul.

Internet permet de s'affranchir des distances pour la communication et le partage de ressources. Internet permet de s'affranchir de certaines contraintes temporelles. En particulier, il est possible de réaliser plusieurs tâches en parallèle. Il est possible grâce à la téléphonie mobile, au wifi, au satellite de se connecter de n'importe où et à n'importe quel moment. Internet permet d'optimiser et d'intensifier son temps. Il est naturel que ce nouveau rapport au temps et à l'espace ait un impact sur la façon dont l'homme vit en société et en particulier sur son rapport au travail.

Nous allons dans cette partie présenter les approches du travail dit coopératif et du travail dit collaboratif sur Internet, toujours dans l'optique de souligner les usages qui peuvent servir à la démocratie électronique et en particulier à la concertation en ligne.

La distinction entre travail coopératif et travail collaboratif est fine et floue. La plupart du temps, ces termes sont employés comme des synonymes. Dans les deux cas, travail coopératif ou collaboratif, sont concernés des individus interagissant dans un but commun. On parlera en règle général d'outils collaboratifs ou du Web collaboratif sans distinguer coopératif et collaboratif. Le principe du travail collaboratif repose sur une vision utopiste du Web ou la somme de petites participations peu organisées permet de construire un tout utile, solide, concret. L'encyclopédie collaborative en ligne Wikipédia en est certainement le meilleur exemple. Cet exemple permet aussi d'illustrer que même si le travail collaboratif repose essentiellement sur des participations spontanées et non organisées, le travail collaboratif nécessite tout de même un minimum d'organisation. Jérôme Delacroix rappelle qu'il faut bien "*jardiner*" son wiki [Delacroix, 2005]. Qu'il faut qu'un individu (ou un groupe d'individus) soit responsable du "*nettoyage*" du contenu. L'encyclopédie Wikipédia fonctionne notamment grâce à la fondation Wikipédia qui est structurée, comme toute association. L'association a un Bureau qui fixe la politique de l'encyclopédie. Elle

"emploie" aussi une armée de bénévoles qui sont spécialisés dans le *jardinage* du wiki. Le jardinage regroupant toutes les opérations de modération (créer une dynamique, gérer les contenus malveillants ...) nécessaires au bon fonctionnement d'un wiki. Donc Wikipédia est un projet mêlant l'intervention des membres de l'association, membres bien organisés dont les tâches sont clairement définies et réparties pour nettoyer l'encyclopédie, et des auteurs produisant spontanément et sans aucun engagement du contenu. Pour Jean-Michel Cornu, [Cornu, 2004], la coopération est une nouvelle forme de gestion de projet émergeant de l'arrivée de l'information immatérielle. La production d'informations immatérielles implique des fonctionnements différents que la production de biens matériels. Le travail coopératif est une pratique courante dans le monde du libre et des associations.

Dans son histoire des outils collaboratifs [Buffa, 2008], Michel Buffa souligne que le travail collaboratif (mais en réalité tout travail collectif) repose sur deux types de pratiques utilisées séparément ou conjointement. Le premier type de pratique est la communication. Le deuxième est le partage de ressources. Nous allons présenter ici un ensemble d'usages et d'outils pour ces deux types de pratiques.

La communication et le partage d'informations

[Licklider et Taylor, 1968] définit la communication comme étant plus que le transfert d'une information d'un point A à un point B. Il s'agit de mettre en interaction deux individus qui ont chacun un modèle mental différent et qui vont tenter de rapprocher leurs modèles. Chaque individu transmet de l'information à l'autre qui va à son tour modifier et enrichir cette information. Cette activité permet une appropriation active de l'information (contrairement à la lecture par exemple où le lecteur ne peut pas modifier l'information). Agissant conjointement à enrichir cette information, les individus communicants construisent ensemble quelque chose, une représentation commune. La communication permet de partager une connaissance commune. On retrouve cette idée au cœur de la gestion des connaissances (*knowledge management*) dont l'objectif est de partager et mieux organiser la connaissance d'un groupe, d'une entreprise, afin de pérenniser le savoir et de créer de nouvelles idées [Nonaka et Takeuchi, 1995]. Ce que l'on retrouve dans l'utilisation d'un Wiki par exemple.

Un premier point important est donc de bien faire la distinction entre communication et discussion. La discussion permet uniquement à des individus d'échanger. La communication ajoute l'alignement de modèles, la création d'une connaissance commune. La communication se base sur la discussion mais apporte un plus. Concrètement, la plupart des outils de discussion ne font que garder des traces des échanges. Si l'on peut retrouver l'intégralité des interventions, on n'a pas une vision de l'évolution des modèles des participants. Un outil de communication devrait donner avant tout une visibilité de ce modèle. Nous utiliserons cette distinction pour parler d'outils de discussion ou d'outils de communication.

Dans les pratiques de communication ou de discussion, il existe une distinction entre les outils permettant d'être en co-présence simultanée, on parle alors d'outils synchrones, et les outils ne permettant pas la co-présence, on parle alors d'outils asynchrones.

L'autre point essentiel du travail collectif est le partage de ressources [Buffa, 2008]. Dès le début d'Internet des solutions de partage ont été proposées. FTP (*File Transfer Protocol*) est le premier protocole de transfert de fichiers permettant à un client de déposer ("upload" en anglais) ou de télécharger ("download" en anglais) un fichier sur un serveur. Le Web statique en soi est une solution pour accéder à des fichiers partagés. Les problèmes liés au partage de ressources sont celles de la gestion de version et de l'édition concurrente de ces ressources. Dans le cadre des débats 2.0, le partage de ressources est important pour permettre aux participants d'accéder à l'information de base nécessaire à la compréhension du débat. La plupart des sites institutionnels proposent d'accéder à des documents supports de communication. Parmi les outils cités précédemment, les outils permettant de partager des ressources sont les Blogs et les wikis. Les autres outils permettent de transmettre des liens pour accéder à des ressources mais ne sont pas en soi des solutions d'accès à des ressources. Dans un Blog ou un Wiki, un article est une source d'information. Ces sources d'information sont éditables dans un Wiki et on peut les commenter dans un Blog. Le rapport à l'information dans ces outils est interactif, ce qui permet une appropriation active de l'information. Dans le contexte de débats 2.0, cette fonctionnalité est importante.

L'information et la communication sont des facteurs de pouvoir. Hors, Internet, le Web et plus globalement les technologies numériques impliquent un changement d'échelle et un rapport très relationnel à l'information. Les citoyens d'aujourd'hui s'informent sur des sites de discussion, auprès des gens de leur communauté. Des sites comme Agora Vox⁴ sont des exemples de nouvelles approches de l'information citoyenne. Chaque citoyen peut être un capteur d'information. On parle du Web participatif qui incite à une démocratisation de l'expression publique et à une intensification des échanges [Compiègne, 2011]. Dans ces espaces, la hiérarchisation se fait par la qualité des échanges, par le suivi par un plus grand nombre. L'individu devient important non pas par son statut a priori (études, emploi ...) mais par l'adhésion qu'il suscite auprès des autres dans le site de la communauté. On est dans la logique de "l'intelligence des foules" ("the wisdom of crowds") [Sutcliffe, 2004]. Les indicateurs de pertinence fournis par les systèmes en ligne sont le nombre d'auteurs, le nombre de personnes qui soutiennent l'idée exprimée, le nombre de corrections effectuées, le nombre de lectures ... Comment repérer dans ce contexte ce qui est vrai ? ce qui est original ? ce qui est l'idée de demain ? Celles-ci ne créent pas nécessairement l'adhésion du plus grand nombre.

4. <http://www.agoravox.fr/>

2.2.4 Les nouveaux usages du Web

Dans le champ de la démocratie électronique, il existe bien d'autres applications Web. Nous allons souligner ici quelques pratiques qui peuvent être liées à des outils de discussion pour augmenter la participation en ligne.

Ces dernières années, avec l'avènement du Web 2.0, l'utilisateur est de plus en plus sollicité et amené à participer dans tous les aspects de la vie en société. Il y a des applications pour répondre à tous les besoins de l'individu ou presque. Internet est un formidable outil pour recruter et mobiliser par le biais des réseaux sociaux notamment. L'information se diffuse de proche en proche et peut créer l'adhésion. Les systèmes de co-présence permettent d'interagir avec n'importe qui à partir du moment que celui-ci est consentant. La messagerie instantanée mais aussi les outils du type de Aka Aki⁵ couplant la disponibilité temporelle et la localisation spatiale permettent de consulter la disponibilité et la proximité d'interlocuteurs potentiels. L'utilisateur peut aussi devenir producteur. Des applications comme mymajorcompany⁶ invitent chacun à investir sur un artiste et devenir son producteur. Le cloud funding est aujourd'hui décliné à tous les produits (films, jeux vidéos, ...). Il peut même servir à lever des fonds pour une cause.

Durant la campagne présidentielle américaine de 2008, le candidat Barack Obama s'est illustré notamment de par l'usage d'Internet dans sa campagne [TerraNova, 2009]. Le site de campagne du candidat est un véritable outil de réseau social, un *SocNet* à part entière où chacun peut devenir directeur de campagne, gérer ses contacts, rédiger des articles de type Blogs (400 000 articles rédigés pendant la campagne) et laisser des commentaires sur tout, partout dans le site. Le site fournit un ensemble d'outils de géolocalisation et de calendriers partagés pour organiser les mobilisations. Il fournit des kits de sondage à imprimer et des outils de saisies des résultats en ligne. Il permet à chacun de faire des dons en ligne et de suivre la levée de fonds qu'il a réalisé (combien j'ai donné, combien les gens que j'ai contacté ont donné). Pour la campagne, il y a eu 3 millions de donateurs qui ont apporté 500 millions de dollars.

Un compte pour le candidat a été ouvert sur chacun des SocNets déjà existants (MySpace, Facebook, mais aussi des SocNets ciblant des communautés comme BlackPlanet⁷ ou Disaboom⁸). Ces comptes ont permis de recruter sur des réseaux en ligne et rapatrier les utilisateurs vers le site du candidat. Le nombre de militants recrutés via ces SocNets est estimé à 5 millions durant la campagne.

Le candidat a aussi largement utilisé la vidéo en ligne pour délivrer un message de campagne non déformé par les représentants de la presse traditionnelle. Une chaîne youtube dédiée à la communication de la campagne a été créée et a totalisé plus de 1000 vidéos le temps de la campagne.

5. <http://www.aka-aki.com/>

6. <http://www.mymajorcompany.com>

7. <http://www.blackplanet.com/>

8. <http://www.disaboom.com/>

Terra Nova dresse une liste de recommandations pour les partis politiques et la juridiction française concernant les campagnes électorales à partir d'observations de la campagne de Barack Obama [TerraNova, 2009]. Parmi la dizaine de recommandations de Terra Nova, la numéro quatre encourage les partis politiques à investir dans une épine dorsale numérique.

1. Penser Internet comme outil de "*back office*" (c'est à dire comme un espace de gestion, l'arrière boutique) de l'organisation de campagne.
2. Réorganiser les partis autour d'un système d'information essentiellement "*web services*" pour le réseau social des militants.
3. Développer la présence sur les réseaux sociaux externes.
4. Réaliser et maintenir une veille technologique sur les nouveaux usages.

L'idée centrale est qu'il faut développer une campagne "de masse" portée par les militants et les sympathisants plus que par les acteurs des partis politiques. On est en plein dans une vision de la démocratie participative ou la place de la représentativité, en tout cas au cours des campagnes, est fortement remise en question. La campagne de Barack Obama a notamment montré que le recrutement par les proches (techniques de porte à porte ciblées) est la forme de recrutement la plus efficace. Il s'agit donc de bien déterminer le profil des électeurs et de les convaincre sur le terrain par la mise en relation de militants à même de les convaincre (parce que partageant les mêmes intérêts, habitant le même quartier ...). Dans cette optique, l'usage d'Internet et des réseaux sociaux est un atout incontesté qui permet de mieux préparer les actions de terrain, de cibler et de mettre en relation.

Les usages des différents outils listés précédemment répondent à des besoins différents et parfois complémentaires. En particulier, il est possible d'utiliser un Blog pour faire partager ses idées et de notifier via un SocNet ou par mail la publication d'un article. La rédaction d'une page de Wiki est souvent le résultat de nombreux échanges par mails, Forums ou messagerie instantanée. L'internaute aujourd'hui, et en particulier l'internaute de la génération digitale, est habitué à passer d'un outil à un autre, à les utiliser de façon complémentaire. Cependant, cet usage complémentaire est rarement prévu techniquement. Il existe des solutions techniques où plusieurs fonctionnalités sont regroupées, par exemple certains wikis permettent d'associer des fils de discussion, comme dans des Forums, à des pages, Google a proposé l'application google-wave⁹ (qui n'est aujourd'hui plus maintenu) permettant de regrouper les fonctionnalités des mails, de google-docs¹⁰ (google-docs est une application développée par Google permettant de rédiger un document à plusieurs de façon synchrone ou asynchrone), de la messagerie instantanée et d'autres. Aujourd'hui Google a même créé son propre SocNet, google+¹¹. Les *GroupWare*, ces outils dédiés à l'Entreprise sont en général une agglomération de fonctionnalités, d'outils de discussions,

9. <http://wave.google.com>

10. <http://docs.google.com>

11. <http://plus.google.com/>

de communication et d'outils permettant de gérer les processus de production de l'Entreprise. Le travail collectif et particulièrement les entreprises agrègent les applications pour répondre à leurs besoins. La figure 2.2, tirée de l'article de Fred Cavazza [Cavazza, 2007] illustre un cas d'Entreprise 2.0 [McAfee, 2006c] [McAfee, 2006b]. Une Entreprise 2.0 utilise les applications sociales du Web 2.0. Le but étant d'améliorer la communication interne et externe, avec les partenaires, clients ...

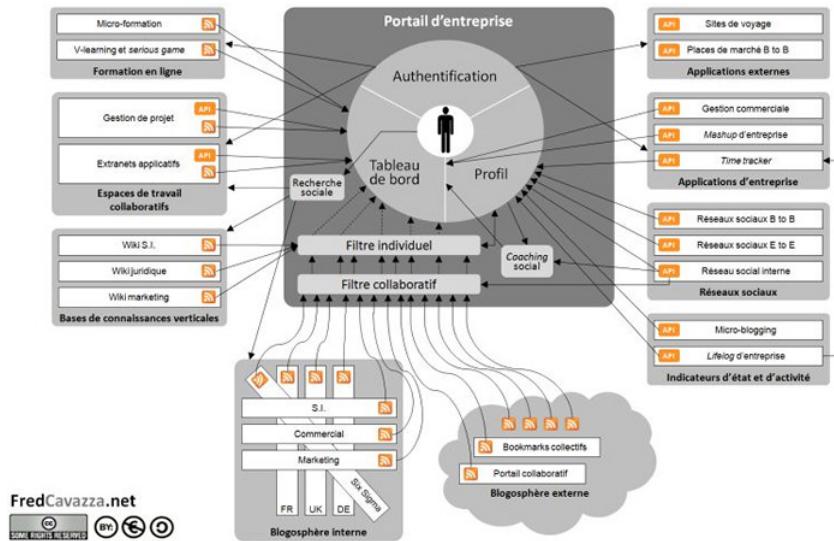


FIGURE 2.2 : Un exemple d'Entreprise 2.0

Dans son travail de thèse, Alexandre Passant propose des solutions pour l'Entreprise 2.0 se basant sur la mise en place de Mashup [Passant, 2009]. Réaliser un Mashup applicatif nécessite de jongler entre des formats de données différents d'une application à une autre. Un Mashup étant le terme désignant la création d'applications composites réutilisant des données issues d'autres applications. Typiquement, réaliser un flux RSS qui permettrait d'afficher l'ensemble des activités en ligne d'une Entreprise nécessite de pouvoir récupérer et traiter des données provenant d'un SocNet, de plusieurs wikis, de plusieurs Blogs, de plusieurs Forums. Bien que souvent, cette tâche se résume à agréger des flux RSS existants, à chaque outil, le format de données peut être différent. Il faut parfois faire des requêtes SQL, parfois traiter du XML ou directement du HTML, parfois des format textuels ad hoc ... Alexandre Passant propose d'utiliser les technologies du web sémantique pour représenter les données issues de ces différents outils et ainsi faciliter la mise en place de Mashups.

De plus, l'utilisateur est, en règle générale, identifié différemment dans chaque outil. Pour chaque outil, l'utilisateur a dû créer un nouveau compte avec parfois un pseudonyme différent. Ce problème est celui que OpenID essaye de résoudre en centralisant les données de tous les comptes utilisateurs et en fournissant un identifiant unique à chacun qui lui évite de recréer un compte à chaque nouvelle application. Cependant, OpenID reste

marginalement utilisé. Le fait que les données des différents comptes soient centralisées sur un serveur OpenID est peut-être un frein créant des problèmes de confiance auprès des utilisateurs. Ce problème de gestion des comptes utilisateurs et de leur identité numérique éclatée entre les applications n'est pas encore résolu.

Nous considérons que les applications listées précédemment sont réutilisables dans le contexte des débats 2.0 et plus largement dans le contexte de la démocratie électronique. Cependant, adopter ces technologies pose exactement le même problème que dans le contexte de l'Entreprise 2.0 : Comment recouper les informations provenant de sites différents ? Comment relier les identités et les actions d'un utilisateur sur ces différents sites ?

2.2.5 Web 2.0

Le concept marketing de Web 2.0 est proposé en 2005 par Tim O'Reilly, Dale Dougherty et Craig Line [O'Reilly, 2005]. Au premier Web essentiellement statique est opposée la vision d'un nouveau Web plus dynamique ou les idées de participation, d'interaction et de partage deviennent fondamentales. Les consom-acteurs ne sont plus uniquement consommateurs de pages Web mais ils deviennent aussi producteurs de contenus. Parfois même sans en être conscients. On parlera aussi du Web en tant que plateforme. Dans cette vision, les applications de bureau sont déportées vers le Web (comme des clients mails accessibles par navigateurs ou des outils d'édition de documents par exemple). Le Web propose désormais du contenu dynamique et des applications.

Il faut bien voir que le concept de Web 2.0 n'a pas amené de révolution technique au Web. Le concept de Web 2.0 est, au contraire, né de l'observation de nouveaux services déployés sur le Web grâce notamment à des technologies comme AJAX (*Asynchronous Javascript And XML*). Le Web 2.0 est une nouvelle vision du Web. Une vision définie après le constat que le Web a évolué depuis ses débuts et que aujourd'hui, il est fondamentalement participatif (ce qui était finalement l'une des intentions originelles du Web prévue dans le protocole HTTP). Certains acteurs importants au temps de l'origine du Web ont fini par s'effondrer devant de nouveaux arrivants très vite devenus des acteurs incontournables. Les points qui ont fait la force de ces nouveaux acteurs sont les suivants :

- Le Web est une plateforme. Les applications proposées sur le Web sont de plus en plus complètes, comme n'importe quelle application qu'on installerait sur un système d'exploitation. Elles présentent en outre les avantages de la zéro-installation et de la transparence des mises à jour pour l'utilisateur. Utiliser une application Web ne nécessite pas d'installation ni de mises à jour par l'utilisateur.
- Une application Web 2.0 rend un ou plusieurs services mais ne rend pas tous les services. Une application Web 2.0 peut être pensée en complémentarité d'autres applications déjà existantes. Les applications échangent des données entre elles et permettent de créer facilement de nouveaux services complémentaires. Par exemple, Facebook a ouvert une partie de ses données et de ses apis pour faciliter le développement par des tiers d'applications complémentaires (ce qui a notamment permis

la prolifération de jeux dits sociaux exploitant le réseau social des utilisateurs Facebook).

- Le ou les services proposées par l'application sont performants grâce aux données que créent les utilisateurs. Par l'utilisation du service, l'utilisateur crée de nouvelles données qui vont améliorer d'autant la qualité du service. Par exemple, plus il y a de gens qui participent à la rédaction d'un article de Wiki plus la qualité et la pertinence de l'article sont censées augmenter.

Ces points impliquent que lors de la création d'un service, l'aspect social et le partage des données doivent être pris en compte. Les créateurs de nouveaux services Web sont amenés à se poser les questions suivantes :

- De quelles données le service a-t-il besoin ?
- Comment obtenir ces données sans alourdir l'utilisation ?
- Est-ce que ces données n'existent pas déjà ailleurs ?
- Quelles données peuvent être partagées sans risque ?
- Comment profiter du réseau des utilisateurs pour propager l'utilisation du service ?

La place de l'utilisateur et ses interactions avec les autres deviennent prépondérantes. Plus il y a d'utilisateurs plus le service s'améliore. On parle du Web Social pour signifier le sous-ensemble d'interactions sur le Web qui sont hautement sociales [Breslin *et al.*, 2009].

Les dérivés 2.0

Suite à l'invention du concept 2.0, une mode du 2.0 s'est opérée. Tout domaine pouvant être impacté par l'usage des technologies du Web 2.0 a eu droit à sa déclinaison en 2.0. L'idée est à peu près toujours la même : comment améliorer la pratique dans tel domaine en s'appuyant sur des technologies du Web ? Si la plupart de ces concepts 2.0 n'apportent pas grand chose à la réflexion dans notre travail, il n'empêche que de chercher à décliner en 2.0 notre problématique était essentiel pour bien cerner les enjeux et l'état de l'art.

Nous avons déjà proposé une définition du concept de débat 2.0. Il nous semble que parmi les concepts 2.0 importants, celui de collectivité 2.0 mérite d'être souligné. Les collectivités territoriales sont les premiers concernés par la concertation des citoyens, leur passage aux technologies du Web représente de réels enjeux techniques, économiques et politiques. Le concept de collectivité 2.0 renvoie donc à l'usage de technologies du Web 2.0 pour les collectivités territoriales. Comme nous l'avons déjà énoncé, les collectivités ont un rôle de mise en place de services aux citoyens, elles sont responsables de la mise en œuvre sur le terrain des politiques d'élus locaux et donc elles sont directement impactées par les décisions de gestion de territoires et de biens communs. Les collectivités territoriales doivent informer les citoyens et ont besoin de récolter les feedbacks des citoyens sur toutes leurs actions. Les collectivités territoriales sont donc les premières concernées par les outils de débats 2.0.

Aujourd'hui, les collectivités adoptent les outils du Web 2.0. L'Auvergne a ouvert le site

auwwwergne¹², portail d'information sur l'activité de la région mais aussi un SocNet à part entière. Les utilisateurs du site peuvent créer des articles de Blogs, gérer leurs contacts, partager des photos ... La ville de Brest a créé le projet wiki-brest¹³ mettant un Wiki à disposition des citoyens. La ville de Rennes a adopté un SocNet de proximité utilisant la géolocalisation, la Ruche¹⁴. Petit à petit, les régions adoptent des outils Web 2.0 pour élargir le champ de la concertation. Le site Blog Territorial¹⁵ proposait en 2009 un panorama de l'usage des Blogs par les collectivités territoriales [Confino, 2009]. Plus largement, cet ouvrage est aussi l'occasion de lister les pratiques 2.0 des collectivités territoriales. De nombreux sites d'institutions se contentent de diffuser de l'information. Le portail est un moyen d'accès à des documents mis à disposition par la collectivité mais il n'est pas toujours possible de laisser un avis sur ces documents. Pire que ça, il n'est même pas toujours possible pour la collectivité de savoir qui a eu recours à quels documents. L'ouvrage souligne bien que si toutes les collectivités ne sont pas encore passées aux outils Web 2.0, ce n'est pas, la plupart du temps, par manque de compétences. Bien souvent, la volonté est politique avant tout. Ouvrir un site de collectivité 2.0, c'est accepter que les citoyens puissent exprimer leurs critiques et que celles-ci soient visibles. Si certaines personnalités politiques ont tout à fait accepté de courir ce risque, même à titre personnel, il reste que certains craignent que le site ne devienne un lieu de dénonciation et d'opposition permanente. En même temps, de nombreuses collectivités ont fait le dur constat que la critique sur le Web était déjà existante ailleurs. La collectivité n'ayant pas ouvert de site 2.0, les citoyens ont créé leurs propres Blogs, Forums ou wikis et l'institution peine à accéder au contenu des débats qui y ont lieu. Il y a un double risque pour les collectivités de ne pas s'ouvrir à la participation en ligne, le premier est d'être vue comme fermée à la concertation, aux débats, le deuxième est d'arriver en retard, de passer à côté d'avis de citoyens déjà exprimés ailleurs.

L'un des concepts 2.0 le plus en vogue est celui d'Entreprise 2.0. Nous l'avons rapidement évoqué dans notre partie consacrée au travail collectif. L'Entreprise 2.0 est probablement après le concept de Web 2.0 lui-même, le concept ayant suscité le plus de réflexions et d'innovations. Nous revenons ici sur des principes de l'Entreprise 2.0 qui nous semblent pertinents pour les débats 2.0. Le concept d'Entreprise est proposé par Tim McAfee dans deux articles sur son Blog personnel [McAfee, 2006c] et [McAfee, 2006b]. McAfee introduit les notions d'émergence et d'horizontalité dans l'organisation. Les outils adoptés par l'Entreprise doivent avant tout être adoptés par les employés de l'Entreprise. Et autant que possible, ils doivent permettre la participation de tous dans l'Entreprise, indifféremment de leur statut. L'usage présenté par McAfee reste fortement cantonné à un contexte de veille et de capitalisation de connaissances. Pour McAfee, il faut avant tout agréger tout le savoir de l'Entreprise et tout le savoir sur l'Entreprise (ce qui se dit de l'Entreprise ailleurs que

12. <http://auwwwergne.com/>

13. <http://www.wiki-brest.net/index.php>

14. <http://beta.ruche.org/>

15. <http://www.blog-territorial.com>

dans l'Entreprise). Il s'agit donc de récupérer et de structurer des données par le biais des activités des employés et partenaires, clients ... McAfee présente 6 composantes qui sont pour lui incontournables dans l'Entreprise 2.0 et qu'il regroupe sous l'acronyme *SLATES* [McAfee, 2006a] :

1. **Search** : les utilisateurs doivent pouvoir chercher dans les données ce qui les intéresse.
2. **Links** : les utilisateurs doivent pouvoir créer des liens entre les données.
3. **Authoring** : les utilisateurs doivent pouvoir créer des données.
4. **Tags** : les utilisateurs doivent pouvoir tagger, catégoriser, les données. Ceci permet de créer une indexation sociale (appelée aussi *Folksonomie*) et permet de garder une trace de la lecture des données par les utilisateurs.
5. **Extension** : le système doit proposer des formes d'assistance ou recommandations aux utilisateurs.
6. **Signals** : le système doit fournir des indicateurs de l'activité. Tout ce qui permet à l'utilisateur de rester au courant des modifications des données.

Cet acronyme est repris dans la thèse d'Alexandre Passant [Passant, 2009] qui en propose une version étendue par les technologies du web sémantique. Passant souligne un ensemble de limites à l'approche SLATES que les technologies du web sémantique peuvent aider à résoudre :

- La recherche dans un ensemble de données se résume souvent à vérifier la présence ou non d'un terme.
- Les liens pouvant être créés entre les données ne présentent pas de réelles sémantiques.
- Les tags ne sont que des termes dont le sens n'est pas précisé.
- Les approches des systèmes de recommandation ne reposent que sur des statistiques ou des co-occurrences. On ne prend pas en compte par exemple, les centres d'intérêts des utilisateurs puisque il n'y a pas de concepts pour les représenter et les lier aux données.
- Les systèmes de notification de l'activité créent des surcharges de notification. Il n'y a pas de filtrage de l'activité et tout est notifié à tout le monde.

Les réflexions autour du Web 2.0 et de ses domaines d'application apportent des enseignements dont il faut tenir compte dans notre approche des débats 2.0. En particulier, elles soulèvent un ensemble de problématiques auxquelles les débats 2.0 seront confrontés.

Problématiques du Web2.0

Nous revenons dans cette partie sur les problématiques du Web 2.0 que l'on retrouve dans les débats 2.0.

Le premier problème du Web 2.0 est la masse d'information. Tout le monde produit du

contenu et la taille du Web devient astronomique, de l'ordre aujourd'hui de plusieurs zettaoctets ... Outre les difficultés de stockage que cela crée, il y a avant tout un problème de navigation. Il n'est pas anodin que l'une des sociétés les plus importante du Web aujourd'hui soit Google, une société dont le premier produit est un moteur de recherches, une solution pour trouver ce que l'on cherche sur le Web.

Le deuxième problème, fortement lié au premier, est celui de la pertinence des informations. Aujourd'hui, les informations pertinentes sont celles remontées en premiers résultats par les moteurs de recherche. Ce qui signifie donc que ce sont ces moteurs de recherche qui fixent les règles d'évaluation de la pertinence des informations. Une règle qui prévaut dans le calcul par les moteurs de recherche, est celle du nombre de liens. L'information est donc classée selon un critère de référencement. Plus une ressource est liée à d'autres ressources, plus celle-ci est pertinente. Il est important de souligner que ces liens n'ont pas de sémantique. La raison de l'existence d'un lien entre deux ressources n'est pas explicite et encore moins compréhensible pour un programme.

Le troisième problème est celui de la séparation des données. Chaque utilisateur a un compte différent dans chaque applications. Il n'est pas aisément de synthétiser son activité globale. Plus largement, les différents services Web ont tendance à créer des silos de données, à préserver leurs données. L'ouverture de ces silos de données est l'un des enjeux du Web de demain. Une personne ayant un compte Facebook ne peut pas communiquer avec les utilisateurs de Google+ par exemple.

Un autre problème du Web 2.0 qui a un impact fort sur les débats 2.0 est celui de la dé-responsabilisation vis à vis des avis exprimés. Retranché derrière un pseudonyme ou l'anonymat, l'utilisateur a tendance à libérer sa parole et même plus, à considérer qu'il peut tout dire. Cette tendance à jouer des rôles sur le Web pose la question de la sincérité des avis exprimés et particulièrement dans les débats en ligne.

2.3 L'interopérabilité

Nous allons dans cette partie présenter la problématique générale de l'interopérabilité entre applications sur le Web. Nous reviendrons sur ce projet que l'on appelle le Web des données [Heath et al., 2011]. Puis nous présenterons le Web sémantique et la technologie RDF qui sont l'un des éléments clefs pour la réussite du Web des données. Enfin, nous présenterons une nouvelle approche du Web qui tend de définir le Web comme étant une Science à part entière alliant des compétences pluridisciplinaires pour étudier le Web d'un point de vue à la fois micro et macro.

Le Web des données

Qu'appelle-t-on le Web des données ? Aujourd'hui sur le Web on manipule énormément de données quotidiennement. On peut par exemple consulter son compte en

banque, partager des photos de famille, gérer son agenda ... Toutes ces données sont gérées par des applications qui les gardent pour elles mêmes, dans leurs propres bases de données, créant des silos de données. Ceci crée des limites dans l'usage du Web. Par exemple, on aimeraient pouvoir afficher nos mouvements bancaires dans un agenda pour pouvoir coupler plus facilement l'information de nos dépenses et de notre activité ce jour là. Ceci est impossible si les applications gérant l'agenda et les coordonnées bancaires ne se mettent pas d'accord sur un partage de données, sur l'adoption d'un format commun et sur un protocole de communication de ces données. Cette situation impose à l'utilisateur de devoir bien souvent recouper les informations lui-même mais cela crée aussi d'autres problèmes. Par exemple, il va être obligé de créer un compte pour chaque application. Avec un système comme les mails, les utilisateurs peuvent communiquer les uns avec les autres quelque-soit le serveur mail qu'ils utilisent (quelqu'un ayant un compte mail chez Google peut très bien écrire avec quelqu'un ayant un compte chez Free). Mais avec le Web dans son état actuel, pour pouvoir communiquer avec des gens qui sont sur Facebook ou Myspace, il faut avoir un compte sur Facebook et un compte sur Myspace.

Dans le Web actuel, la plupart du temps, les données sont dans des formats propres à chaque application. Une grande quantité de données (celles concernant le profil de l'utilisateur par exemple) est reproduite d'une application à l'autre dans des formats différents. L'utilisateur est sans cesse en train de jongler d'une application à une autre avec un compte différent. C'est à lui, ou à des programmes développés spécialement pour cette tâche, de recouper les informations d'une application à une autre. Peu d'applications encore sont pensées pour être interopérables.

Pour s'attaquer à cette problématique, le Web sémantique défend le projet du Web des données. Le Web des données vise à lier toutes les données, ou en tout cas un maximum de données possibles, aujourd'hui présentes sur le Web. Ceci passe nécessairement par le développement de nouveaux standards de représentation des données permettant d'augmenter l'interopérabilité. Les origines du Web des Données reposent sur un effort de la communauté de chercheurs du Web sémantique et particulièrement sur les activités du projet *Linking Open Data* (LOD)¹⁶ du W3C, effort initiateur démarré en Janvier 2007. Le Web des données s'appuie sur les technologies de base du Web, le protocole HTTP, et sur les technologies du Web sémantique, RDF.

Nous reprenons ici un passage du livre [Heath *et al.*, 2011] permettant de synthétiser ce qu'est le Web des données. Le Web des données peut être vu comme une couche additionnelle du Web fortement entrelacée avec les documents classiques du Web :

- Le Web des données est générique et peut contenir n'importe quel type de données ;
- Tout le monde peut y publier des données ;
- Le Web des données est capable de représenter des informations contradictoires au sujet d'une entité ;
- Les entités sont connectées via des liens RDF, créant un graphe global de don-

16. <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

nées avec des ponts entre plusieurs sources de données, ce qui permet de découvrir de nouvelles sources de données. Cela signifie que les applications ne sont pas contraintes d'exploiter une seule source de données, mais peuvent découvrir à la volée de nouvelles sources de données en suivant des liens RDF ;

- Les éditeurs de données ne sont pas contraints dans leurs choix de vocabulaire pour représenter leurs données (un vocabulaire étant un format de description de données écrit en RDF, ce point est expliqué plus en détail plus loin dans ce document) ;
- La donnée se décrit elle-même. Si une application basée sur les données liées rencontre des données décrites dans un vocabulaire inconnu, l'application peut utiliser les URI qui identifient les termes du vocabulaire afin de trouver les descriptions qui permettent de définir les termes (on appelle ce procédé le *déréférencement*) ;
- L'utilisation d'HTTP comme un mécanisme standardisé d'accès aux données et RDF comme modèle standard de données simplifie l'accès aux données en comparaison aux API Web qui reposent sur des modèles de données et des interfaces hétérogènes.

Le Web des données consiste à publier des données en ligne dans des formats standards et à les lier à d'autres données permettant de les ancrer dans le Web. Ce mécanisme de permet de naviguer de sources de données en sources de données et de créer un très grand graphe de données bien que chaque source puisse ne contenir finalement que peu de données. Aujourd'hui le graphe de données du Web des données est constitué de plusieurs milliards de triplets RDF et traite de données couvrant de très nombreux domaines : les sciences, les personnes, les données géographiques, les livres, les films ... L'image 2.3 correspond à l'ensemble des sources de données connues reposant sur le principe des données liées et donc faisant partie du Web des données à la date du 19 septembre 2011. Cette image est tirée du site dédié au projet Linked Data¹⁷ et est régulièrement mise à jour¹⁸.

2.3.1 Web Sémantique

L'un des objectifs du Web sémantique est de résoudre les problèmes d'interopérabilité en proposant des technologies permettant de représenter les données dans un format standardisé afin que des programmes mais aussi des humains puissent facilement comprendre la sémantique d'une ressource et des données qui lui sont associées, sa nature, son contexte de production, ce qu'elle représente, les liens qu'elle entretient avec d'autres ressources...

L'une des propositions du Web sémantique, est le langage RDF (*Resource Description Format*). Le langage RDF permet d'écrire des triplets de type sujet-prédicat-objet. Ces triplets peuvent être vus comme les arcs d'un graphe sommet-arc-sommet. Le langage RDF est un méta-langage permettant de créer un vocabulaire dédié à la description d'un type

17. <http://linkeddata.org/>

18. <http://lod-cloud.net/state>

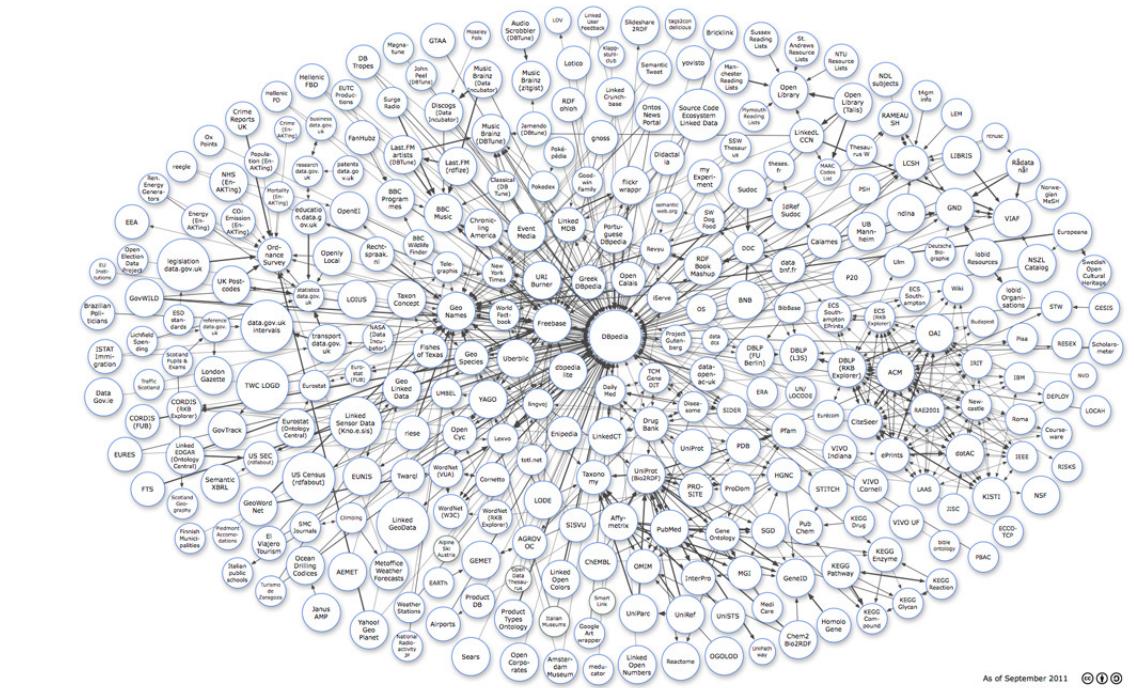


FIGURE 2.3 : Le Web des données au 19 septembre 2011

de données (par exemple, le vocabulaire FOAF est dédié aux données d'un réseau social). Le vocabulaire créé et l'utilisation de ce vocabulaire pour représenter un ensemble de données obéissent aux mêmes règles syntaxiques, celles définies par RDF.

Un vocabulaire RDF est souvent appelé une ontologie. Le terme d'*Ontologie* dans le contexte du Web sémantique renvoie au terme philosophique d'*ontologie* [Moeschler et Reboul, 1994] p438 :

Le terme *ontologie* est un terme philosophique qui désigne, pour une théorie donnée, l'ensemble des objets physiques ou mentaux que cette théorie admet.

En réalité, on distingue souvent les ontologies ne décrivant que les objets mentaux de la théorie, appelés concept, et les ontologies qui décrivent les instances de ces concepts, qui peuvent être des objets physiques. Avec RDF un vocabulaire est une description des concepts et leurs propriétés. Le vocabulaire est ensuite utilisé pour décrire les objets concrets, la base de faits, les instances des concepts. [Moeschler et Reboul, 1994] p444 :

Ces concepts sont des adresses mémoires sous lesquelles sont classées un certain nombre d'informations de natures diverses : logiques, encyclopédiques, lexicales.

Les informations logiques, encyclopédiques et lexicales se différencient de la façon suivante :

- Les premières correspondent aux informations sur les relations logiques (implication, contradiction, etc.) que le concept peut entretenir avec d'autres concepts.
- Les secondes correspondent à toutes les informations de nature non logique que l'on a sur le concept et qui permettent de lui attribuer une extension s'il en a une.
- Les troisièmes, enfin, correspondent à la contrepartie du concept en langue naturelle, au(x) mot(s) qui lui correspond(ent).

[Gruber, 1993] définit une ontologie comme étant une spécification d'une conceptualisation. Une ontologie est une description de concepts et de relations qui peuvent exister pour un agent ou une communauté d'agents. Ce qui est important ici, c'est ce pour quoi est faite l'ontologie. Dans le domaine du Web sémantique, les ontologies sont faites pour faciliter l'échange de connaissances et la réutilisation de connaissances. Cette définition est revue dans [Gruber, 2009], une ontologie définit un ensemble de primitives avec lesquelles est modelé un domaine de connaissance ou de discours. Les primitives sont classiquement les suivantes :

- Des classes
- Des attributs ou propriétés
- Des relations entre membres des classes

Le problème de la notion philosophique d'ontologie est qu'elle renvoie souvent à la modélisation du monde dans son intégralité et qu'elle présuppose qu'il n'y a qu'une seule façon de voir le monde. La tâche de modéliser le monde dans son intégralité et de façon universelle, unique est une tâche impossible. Chaque individu peut avoir une vision différente du monde. Concrètement, avec le Web sémantique, il s'agit plutôt de définir des ontologies dédiées à un contexte particulier. Par exemple l'ontologie Dublin Core est dédiée à la représentation des documents et des informations les concernant (l'auteur, la date, la licence d'exploitation ...). Le but n'est pas de créer une ontologie unique mais plutôt un ensemble d'ontologies convenant à des usages ou domaines particuliers.

RDF permet de lier des ressources en précisant la sémantique du lien et de naviguer par les liens de ressources en ressources. Ces liens peuvent être définis entre des ressources étant décrites dans des vocabulaires différents. C'est l'un des grands avantages de RDF, il permet de réutiliser des vocabulaires existants mais aussi de créer des liens entre des données écrites à l'aide de vocabulaires différents. Il est donc possible de composer, selon les applications, avec plusieurs vocabulaires pour décrire les données manipulées par l'application. RDF est parfois présenté comme le langage pensé pour permettre de faire des Mashups exactement pour les raisons précitées (croisement des données provenant de plusieurs sources).

Souvent, la définition et l'utilisation de vocabulaires est considérée comme une tâche

d'expert. Pour créer un vocabulaire, il faut avoir une connaissance du modèle, une connaissance métier liée à la manipulation des données qui sont produites. Et bien souvent, la tâche de création des données est aussi une tâche d'expert. Décrire l'arbre des espèces en RDF (c'est à dire définir un vocabulaire dédié à la représentation des espèces et leurs liens puis effectivement, écrire l'arbre des espèces) est un travail d'expert. Il faut une connaissance de l'arbre des espèces pour pouvoir expliciter le langage permettant de l'écrire et pour pouvoir écrire l'arbre. Mais aujourd'hui, avec les outils du Web 2.0, les données sont produites par les utilisateurs. Il faut donc, dans ce contexte, que l'utilisateur devienne lui-même producteur de la base de données RDF mais aussi, pourquoi pas, du vocabulaire. On parle alors d'ontologies à l'état sauvage [[Giboin et Gandon, 2008](#)].

2.3.2 Web Science

La dimension sociale dans les applications Web est avérée au moins depuis l'apparition du terme de Web 2.0. Cet aspect social crée de nouvelles problématiques. Le Web d'aujourd'hui a un impact sur les modes de communication de l'individu et sur la société dans sa globalité. Les phénomènes observés sur le Web font intervenir de très grandes échelles. Pour Tim Berners-Lee, à qui l'invention du Web est attribuée, étudier l'évolution du Web d'un point de vue technique mais aussi sociologique nécessite aujourd'hui de créer un nouveau champ pluridisciplinaire [[Berners-Lee et al., 2006](#)].

Les objectifs de la science du Web sont nombreux. Il s'agit de considérer l'évolution du Web et de l'échelle (du nombre d'utilisateurs mais aussi de la quantité de données produites) tout en se préoccupant de problèmes de droits de propriété des données, de confidentialité, de pertinence des données ... Le Web est à la fois créé par un ensemble de technologies et par l'usage qu'en font les utilisateurs. Les interactions utilisateurs créent des modèles, des structures qui émergent et qu'il convient d'étudier. Les usages et la technologie ont toujours avancé conjointement. De nouvelles technologies permettant de nouveaux usages et ces nouveaux usages créant le besoin de nouvelles technologies. Il faut donc créer des approches qui permettent d'améliorer cette synergie.

On retrouve dans le *Web science* les mêmes positions éthiques que dans le Web sémantique : la décentralisation de l'information pour éviter les goulots d'étranglement sociaux et techniques, l'ouverture des données pour permettre la réutilisation de l'information dans des contextes inattendus. Cette science est nécessairement pluridisciplinaire car elle repose sur l'emploi d'outils provenant de plusieurs domaines comme les mathématiques et les sciences sociales, permettant d'observer à une échelle microscopique et macroscopique les phénomènes du Web, afin de créer notamment de nouveaux langages, de nouveaux algorithmes, de nouvelles solutions informatiques pour répondre à de nouveaux besoins d'usage.

2.4 Les réseaux sociaux

L'analyse des réseaux sociaux est un domaine des sciences sociales alliant sociologie et théorie des graphes dont le Web, avec sa vague 2.0, est devenu le terrain de prédilection. Nous allons présenter dans cette partie une rapide histoire introductory du domaine de l'analyse des réseaux sociaux, puis nous présenterons les éléments de théorie des graphes constituant la boîte à outils du domaine. Ensuite, nous présenterons des théories du domaine qui ont attiré notre attention pour notre contexte des débats 2.0.

2.4.1 Histoire

Le terme de "réseau" n'est pas propre à la sociologie. A travers les siècles, il a été utilisé dans des contextes variables faisant évoluer progressivement sa sémantique. [Mercklé, 2004] présente une histoire des réseaux sociaux dans laquelle il rappelle que le terme de réseau fût utilisé successivement par les tisserands pour parler de l'entrecroisement des fibres, par la médecine pour signifier l'appareil sanguin puis pour désigner le système de routes et chemins pour circuler. Dans la partie précédente, nous parlons d'Internet en tant que réseau de machines. Le concept de réseau est aujourd'hui une notion abstraite pouvant être déclinée dans des contextes différents. La sémantique du terme de réseau couvre des idées d'entrecroisement, de circulation, de flux, de contrôle, de cohésion, de structure, de représentation topologique. En sciences sociales, le terme de réseau social est venu remplacer les termes de structure, système, cercle, groupe pour signifier un ensemble d'individus et de relations entre ces individus.

Si le terme de "*réseau social*" est invoqué pour la première fois par A. Barnes en 1958 [Barnes, 1958] (*social network* en anglais), il n'empêche que l'origine des réseaux sociaux est plus ancienne. Pierre Mercklé souligne que la recherche d'une paternité stricte au domaine des réseaux sociaux est une tâche pratiquement sans fin tant le domaine s'est nourri d'idées et de concepts présents dans d'autres disciplines comme la philosophie, l'anthropologie ou encore les mathématiques appliquées. Il souligne aussi le fait que l'on retrouve à travers les siècles certaines idées récurrentes qui confirme l'un des postulats de départ des réseaux sociaux qui est que, s'il est possible d'étudier des formes de relations sociales, c'est parce qu'elles présentent une certaine régularité et une certaine stabilité dans des domaines et des époques variables.

L'un des travaux fondateurs cité est celui de Georg Simmel dans lequel il s'efforce de montrer que la catégorie des gens pauvres n'est pas unifiée par les relations entre les individus appartenant au groupe en question mais par les relations qu'entretiennent la société dans son ensemble avec chacun des membres du groupe. [Forsé, 2002] présente une analyse systématique de ce que doit l'analyse des réseaux sociaux à Simmel. En particulier, Georg Simmel est cité comme père fondateur car il a en premier formulé cette idée selon laquelle les structures émergent des interactions entre individus et non des propriétés ou attributs des individus.

Les premières formes de représentation graphiques de réseaux sociaux apparaissent avec la sociométrie dont Jacob L Moreno est le fondateur incontesté et dont il expose les enjeux, méthodes et principes dans [Moreno, 1953]. La sociométrie définit une procédure de recueil de données relationnelles qui consiste à faire déclarer par des individus d'un groupe des relations d'attraction et de répulsion avec les autres membres du groupe. La sociométrie définit aussi une forme de représentation graphique de ces données sous la forme d'un graphe où les individus sont représentés par des sommets dans un plan et les relations par des flèches entre ces sommets. Ce type de représentation est appelé un socio-gramme et correspond à la représentation nœuds-liens en théorie des graphes.

La théorie des graphes est une branche des mathématiques redécouverte dans les années 50 dont les travaux pionniers remontent pourtant à Euler. La théorie des graphes apporte un ensemble d'outils à l'analyse des réseaux sociaux. Elle permet de proposer des représentations graphiques et des analyses par le biais de théorèmes, d'algorithmes et de métriques qui vont permettre de qualifier, distinguer et classer des structures relationnelles en fonction de propriétés de distribution et d'agencements des relations entre individus.

2.4.2 Théorie des graphes

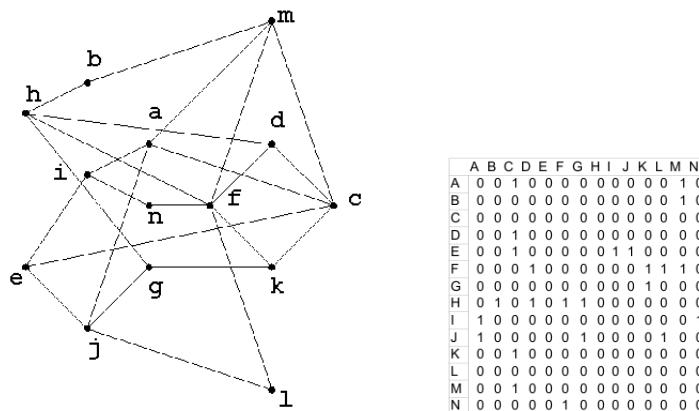
La boîte à outils de l'analyse des réseaux sociaux est essentiellement composée d'éléments de théorie des graphes. Un graphe est un ensemble de sommets reliés par un ensemble d'arêtes qui peuvent être orientées. En analyse des réseaux sociaux, on définit un réseau comme étant un graphe où les sommets sont les entités du réseau, le plus souvent, des individus, et les arêtes les relations entre les individus. En règle générale, le graphe n'est composé que d'un type de lien correspondant à la relation que l'on considère dans le réseau. Par exemple, le réseau social des utilisateurs de Facebook peut être représenté comme étant un graphe d'individus liés par la relation d'amitié Facebook. Un chemin dans un graphe est défini comme étant un ensemble de sommets reliés les uns aux autres un à un par des arêtes.

Les deux représentations d'un graphe les plus utilisées en analyse des réseaux sociaux sont :

- la représentation nœud-lien, représentation intuitive d'un graphe où chaque sommet est représenté par un point dans l'espace et chaque lien entre deux sommets est représenté par un trait entre les points représentant les sommets considérés.
- la représentation matricielle qui permet de représenter dans une matrice d'adjacence l'existence d'un lien entre deux sommets du graphe.

Un ensemble de mesures permet ensuite de caractériser le graphe.

- **L'ordre d'un graphe** est le nombre de sommets dans le graphe.
- **Le voisinage ou degré d'un sommet** est le nombre de sommets liés au sommet considéré. Si les arêtes sont orientées dans le graphe, on parlera de degré entrant et de degré sortant pour respectivement les sommets reliés par une arête pointant vers



(a) Un graphe en représentation noeud-lien

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
A	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0
F	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0
G	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
H	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0
I	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
J	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0
K	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

(b) Le même graphe en représentation matricielle. Les 1 signifient la présence d'un lien entre les 2 sommets, les 0 représentent l'absence de lien.

FIGURE 2.4 : Les deux représentations les plus courantes d'un même graphe

le sommet considéré et pour les sommets reliés par une arête partant du sommet considéré.

- **La densité d'un graphe** est le rapport entre le nombre de liens dans le graphe et le nombre de liens qu'il pourrait y avoir dans le graphe. S'il s'agit d'un graphe d'ordre n , le nombre maximum d'arêtes possibles dans le graphe est de $n(n - 1)$.
- **La distance** entre un sommet et un autre dans le graphe est le nombre de sommets reliés qu'il faut parcourir pour atteindre un sommet partant de l'autre sommet. La distance moyenne est la moyenne des distances dans le graphe. La distance géodésique est la plus courte distance entre deux sommets.
- **Le diamètre d'un graphe** est la plus grande distance dans le graphe.

La centralité est définie pour indiquer l'importance d'un sommet dans le graphe. Il existe plusieurs mesures de centralité :

- **la centralité de degré** (degree centrality) est calculée à partir des degrés des sommets et définit le nœud ayant le plus haut degré comme étant le plus important.
- **la centralité de proximité** (closeness centrality) est calculée à partir des distances dans le graphe, elle considère comme étant le nœud le plus important celui ayant la distance moyenne la plus courte pour couvrir tous les autres sommets du graphe.
- **la centralité d'intermédiaire** (betweenness centrality) souligne la capacité d'un sommet à être un intermédiaire entre deux sommets. Plus cette centralité est élevée plus

le chemin pour aller d'un sommet à un autre dans le graphe a de chance de contenir le sommet considéré.

L'un des problèmes récurrents en analyse des réseaux sociaux est **la détection de communautés**. En règle générale, cette détection de communautés se fait par une méthode de *clustering*. Un **cluster** dans un graphe est défini comme étant un ensemble de sommets. Selon les méthodes de clustering, la densité d'un cluster est plus ou moins élevée. Mais là encore, en général, une communauté est définie comme étant un cluster dense de sommets. On cherche en général des ensembles de sommets fortement liés les uns aux autres. On parlera aussi de connexité pour désigner le fait que tous les sommets d'un graphe sont liés entre eux par des chemins ou si au contraire, il existe des sommets isolés ou plusieurs composantes dans le graphe n'étant pas reliées les unes aux autres. Par exemple, un graphe est dit connexe si il existe entre tous points du graphe un chemin permettant d'aller d'un point à l'autre. En règle générale, une communauté dans un réseau est assimilée à un cluster dans le graphe correspondant, c'est à dire un sous-graphe connexe du graphe représentant le réseau considéré.

L'une des méthodes pour détecter l'existence de sous-graphes connexes est la méthode de bloc-modeling. Dans la représentation matricielle d'un graphe, l'ordre des lignes et des colonnes n'est pas important. Il est donc possible de permuter les lignes (et réciproquement les colonnes) sans modifier le graphe. La technique de bloc-modeling consiste à réaliser des permutations afin d'obtenir des sous-matrices dont toutes les valeurs sont à 1. Ces sous-matrices correspondent à des sous-graphes connexes. Cette méthode est en règle générale attribuée à Harrisson C. White qui dans les années 70 a proposé de nombreuses analyses systématiques à partir de matrices d'adjacence. De nombreuses méthodes basées sur la représentation et le calcul matriciel sont venues améliorer l'efficacité de l'analyse des réseaux sociaux. Le calcul des degrés ou le test de l'existence de chemins entre des sommets sont par exemple des opérations beaucoup plus simples à partir d'une représentation matricielle (il suffit d'additionner toutes les valeurs d'une ligne dans le premier cas et d'élever la matrice à la dimension k pour tester l'existence d'un chemin de longueur k entre deux sommets).

La plupart du temps, les analyses de réseaux sociaux portent sur des graphes dont les sommets et les arêtes ne sont que d'un seul type. L'exemple précédemment cité du graphe des relations Facebook ne considère que des individus (on ne fait pas de distinction entre les hommes et les femmes par exemple) et un seul type de relation, la seule existante dans Facebook, la relation d'amitié. Ce type de réseau est appelé **réseau homogène et mono-relationnel**. Pourtant, nous allons voir des théories considérant des **réseaux hétérogènes** (faisant intervenir des entités de plusieurs natures) et **multi-relationnels** (faisant intervenir des relations de sémantiques variées). Dans nos travaux, nous nous intéresserons à des réseaux hétérogènes (constitués d'individus et de ressources communicationnelles) ainsi que multi-relationnels (notamment considérant des relations d'accords et de désaccords).

2.4.3 Caractéristiques des graphes de réseaux sociaux

Classiquement, on considère que les graphes permettant de représenter des réseaux sociaux font partie de la catégorie des graphes appelés "réseaux réels" ou encore "grands graphes de terrain" [Pons, 2007]. Ceci est valable en tout cas pour des réseaux mono-relationnels et homogènes. Ces graphes possèdent un ensemble de caractéristiques structurelles que nous allons ici résumer :

- Le nombre de sommets est généralement élevé, allant de plusieurs centaines à plusieurs milliers.
- Le nombre moyen de voisins de chaque sommet est faible. Dans l'ensemble, chaque sommet est lié à peu de voisins.
- La distribution des degrés est hétérogène. Il existe un nombre faible mais non négligeable de sommets possédant un très fort degré par rapport à une majorité de sommets possédant un très faible degré. Cette distribution est souvent bien approximée par une loi de puissance. On parle de "graphe sans-échelle". Ceci implique qu'il existe quelques rares sommets de très fort degré qui jouent nécessairement des rôles particuliers par rapport aux autres sommets.
- Il existe des densités locales très fortes par rapport à une faible densité globale du graphe. Ces fortes densités traduisent le fait que les liens entre les sommets qui sont proches sont beaucoup plus probables que les liens entre des sommets éloignés.
- On qualifie les graphes de réseaux sociaux de "graphes petit-monde" (en référence à l'expérience de Travers et Milgram [Travers et Milgram, 1969]). Ceci traduit le fait qu'il existe des très courts chemins pour relier chaque paire de sommets dans le graphe. D'après l'expérience de Travers et Milgram, certains concluent que nous ne sommes qu'à 6 poignées de mains les uns des autres.

2.4.4 Détection de communautés

L'un de nos objectifs est de détecter les groupes d'opinion dans les débats en ligne. La détection de communauté est souvent utilisée pour nourrir des systèmes de recommandation [Iskold, 2007]. Nous souhaitons proposer des recommandations permettant l'ouverture des discussions à des avis nouveaux, différents, afin de lutter contre l'insularisation des groupes d'opinion. Dans le domaine de l'analyse des réseaux sociaux, la détection de communautés est souvent ramenée à un problème de clustering (recherche de clusters dans un graphe). Nous allons dans cette partie faire un résumé des méthodes de clustering existantes en nous focalisant essentiellement sur celles utilisées en analyse de réseaux sociaux. Le but de cette partie est de montrer qu'il existe de nombreuses solutions basées soit sur l'analyse structurelle du réseau soit sur des mesures de similarité (ou ressemblance) entre les sommets des graphes. Nous défendrons qu'il faut trouver une solution mixant ces deux approches et que, de plus, il faut adapter ces approches à des graphes hétérogènes et multi-relationnels.

Pour [Fortunato, 2010], la détection de communauté est un problème de clustering. Pour [Schaeffer, 2007], un problème de clustering est la recherche de clusters dans un graphe ou dans un ensemble de données. Un cluster est défini dans un graphe comme étant le regroupement d'un ensemble de sommets en prenant en considération la structure du graphe, les relations entre ces sommets. Un cluster dans un ensemble de données est le regroupement d'un ensemble de données similaires, partageant des propriétés communes. On parle aussi parfois de classification de données et le domaine est très proche de l'apprentissage non supervisé dans la reconnaissance de "patterns" (un pattern est un schéma de structuration que l'on va pouvoir retrouver). Les articles [Fortunato, 2010] et [Schaeffer, 2007] soulignent qu'il n'existe pas de définition stricte et universelle de ce qu'est un cluster ou une communauté. Les définitions changent en fonction des domaines, de l'application, du contexte dans lequel on cherche les communautés.

Clustering par similarité

Le principe du clustering par similarité repose sur la sélection de propriétés de références dans les éléments traités. Par exemple, si les éléments considérés représentent des documents, une mesure de similarité pourrait être définie en fonction du contenu des documents. On utilise en fouille de données notamment des vecteurs de termes représentant pour un document la fréquence à laquelle apparaissent dans ce document les termes considérés. On peut alors définir une distance entre les documents à partir de ces vecteurs. Plus la distance est réduite, plus les documents sont considérés similaires.

Il existe beaucoup de mesures de distance possibles et celles-ci dépendent du contexte. Une mesure de distance $\text{dist}(d_i, d_j)$ devrait remplir les conditions suivantes [Schaeffer, 2007] :

- la distance entre une donnée et elle-même est nulle. $\text{dist}(d_i, d_i) = 0$
- les distances sont symétriques. $\text{dist}(d_i, d_j) = \text{dist}(d_j, d_i)$
- l'inégalité triangulaire est respectée. $\text{dist}(d_i, d_j) \leq \text{dist}(d_i, d_k) + \text{dist}(d_k, d_j)$

Clustering d'un graphe

Dans le cas du clustering d'un graphe, les méthodes proposées reposent sur la structure du graphe. L'objectif est d'obtenir un partitionnement du graphe. Les méthodes prennent en compte des informations comme la densité des liens entre des sous ensemble de sommets (une clique a tendance à définir un cluster), l'existence de chemins entre des sommets (les sommets d'un cluster sont tous reliés les uns aux autres par des chemins), ou encore le voisinage de sommets (deux sommets d'un graphe ayant les mêmes voisins ont tendance à être dans un même cluster).

Là encore, les propriétés considérées dépendent du contexte, de ce que l'on veut détecter comme communautés. Par exemple, imposer que seules des cliques font des clusters (chaque sommet d'une clique est le voisin de tous les autres sommets de la clique) revient

à supposer que tous les membres d'une même communauté entretiennent les mêmes relations les uns avec les autres, qu'ils ont tous le même rôle dans la communauté. On pourrait au contraire considérer que dans une communauté, certains membres sont centraux et connectés à l'ensemble des autres membres, tandis que d'autres membres sont à la périphérie, qu'ils entretiennent des liens qu'avec certains membres de la communauté et aussi avec des individus hors de la communauté.

Même si la méthode de calcul d'un cluster de graphe peut varier d'un contexte à un autre, il reste que certaines propriétés sont en générales adoptées dans tous les cas comme indicateur de ce qu'est un bon clustering [Fortunato, 2010] :

- Il y a plus de liens au sein du cluster, entre des sommets du cluster, que entre des sommets du cluster et le reste des sommets du graphe.
- La densité du cluster est plus élevée que la densité du graphe d'origine.
- Le cluster est connexe, il existe un chemin entre chaque sommet du cluster et, de plus, tous les sommets sur le chemin sont aussi dans le cluster.

La définition de communautés proposée par [Pons, 2007] résume assez bien ces propriétés :

une communauté est un groupe d'acteurs fortement liés entre eux et faiblement liés aux autres.

Clustering hiérarchique

Enfin, que l'on soit dans le cadre d'un clustering de graphe ou d'un clustering de données, on séparera les clusterings cherchant à proposer un ensemble de clusters à plat, de ceux cherchant à proposer une hiérarchie de clusters. Dans un clustering hiérarchique, on obtient une structure particulière, appelée dendrogramme dans laquelle les feuilles sont des sommets du graphe et les noeuds sont des clusters. La figure 2.5 présente un dendrogramme. Le noeud de départ du dendrogramme, ou noeud le plus haut dans la hiérarchie est le graphe entier. Puis chaque niveau dans la hiérarchie définit un ensemble de clusters contenant des clusters plus bas dans la hiérarchie. Ce type d'approche est intéressant en analyse des réseaux sociaux car l'on peut s'attendre à ce que les communautés soient organisées en hiérarchie (par exemple dans le cas de communautés géocentriques, on peut avoir une communauté à l'échelle nationale regroupant des communautés à l'échelle régionales ...).

Clustering dynamique

Soulignons enfin, que la plupart des travaux de clustering se consacrent à l'analyse de données statiques. Peu de travaux concernent l'étude de la dynamique des clusters, ou communautés. Nous citons le travail de [Palla *et al.*, 2007] dans lequel sont étudiés l'évolution de réseaux et la dynamique de clusters. Les auteurs y décrivent des augmentations de clusters, des diminutions de clusters mais aussi l'apparition et la disparition de clusters, la

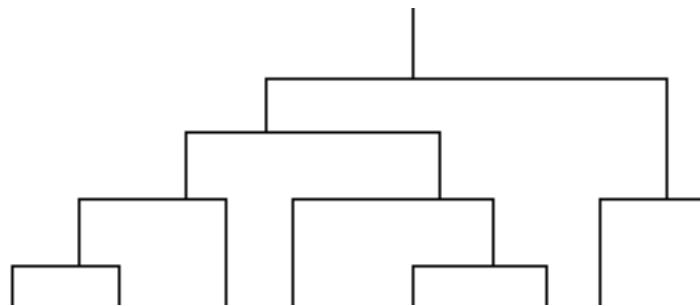


FIGURE 2.5 : Un dendrogramme

fusion et la séparation de clusters. Autant d'événements qui marquent la dynamique d'un réseau social et qui sont d'un fort intérêt dans le contexte de débats en ligne.

2.4.5 Éléments de sciences sociales

Dans cette section, nous allons détailler quelques éléments de théories de sciences sociales qui ont éclairé notre approche de l'analyse des réseaux sociaux et de l'activité en ligne.

Force des liens

Nous présentons ici la théorie de Marc Granovetter sur la force des liens [Granovetter, 1973] [Granovetter, 1982]. Pour Granovetter, les individus ont tendance à se lier à des individus leur ressemblant et du coup à partager les mêmes relations les uns avec les autres (ceci peut se traduire par "les amis de mes amis sont mes amis"). Un réseau social est donc fait de plusieurs groupes fortement connectés ou tout le monde connaît tout le monde. Les liens à l'intérieur de ces groupes sont en règle générale des "liens forts". Les liens extérieurs au groupe sont en général des "liens faibles". Granovetter évalue la force des liens en fonction de la fréquence d'échange entre les individus, l'intensité de l'échange et la part d'intimité dans l'échange (l'intensité peut être définie par la durée de l'échange ou par le contenu de l'échange, l'intimité peut être définie par le contenu et par le cadre de l'échange, ces deux notions sont mal définies dans [Granovetter, 1973], et sont présentées dans [Granovetter, 1982] comme dépendant du contexte expérimental). La théorie de Granovetter défend le fait que les individus interagissent essentiellement avec un groupe restreint d'individus leurs ressemblants. Les échanges entre des individus plus différents, appartenant à d'autres groupes, sont alors marginaux.

L'impact de la distribution polynomiale des degrés dans un réseau social implique qu'il y a quelques individus qui entretiennent beaucoup de liens avec beaucoup d'individus et donc, ces individus relient de nombreux groupes par des liens faibles notamment. La figure

2.6 représente un graphe petit-monde sans échelle dans lequel des groupes partageant des liens forts sont colorés en rouge et les liens faibles sont colorés en vert.

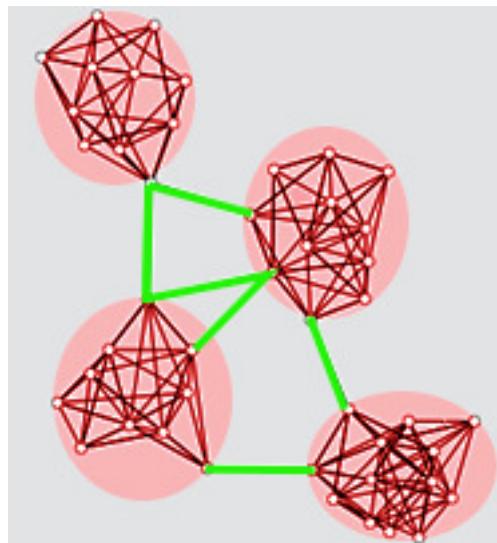


FIGURE 2.6 : Un graphe petit monde sans échelle faisant apparaître la force des liens

Netdoms

Harrison C White de son côté, propose la notion de NetDoms dans laquelle un réseau social est fortement lié à un domaine. Un individu va donc appartenir à des réseaux différents et entretenir des relations différentes selon les domaines. L'individu changeant alors régulièrement de réseau et de domaine [White, 2008]. La figure 2.7 présente une façon de représenter ce concept de Netdoms. L'illustration proposée est inspirée de celle tirée de [White, 2008] en page 8. Les individus sur chaque plan sont les mêmes. Chaque plan représente un domaine différent.

Réseaux centrés objets

Latour met l'accent sur les objets d'interaction qui permettent de construire les relations [Latour, 2007]. Des individus sont en relation parce qu'ils travaillent au même endroit, parce qu'ils manipulent les mêmes outils ... Cette vision des réseaux sociaux est centrée sur les objets d'interaction.

Cette vision semble très adaptée aux pratiques du Web 2.0 ou l'on considérera l'existence d'un lien entre deux individus parce qu'ils écoutent la même musique, achètent le même livre ou participent à une même conversation. Certains parlent d'*object centered sociality* [Engeström, 2005].

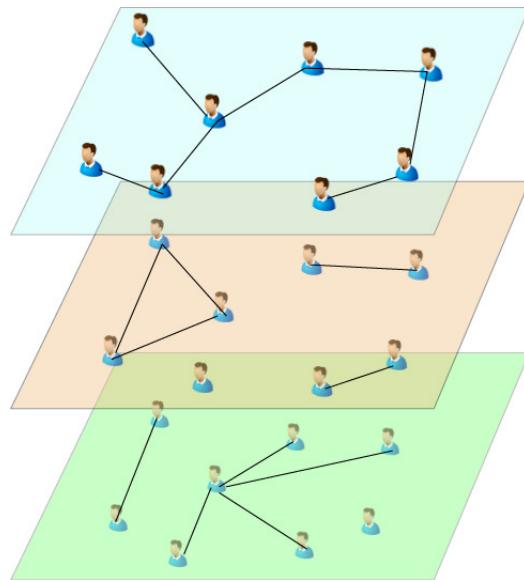


FIGURE 2.7 : Une représentation de NetDoms

Rôles et interactions

Dans la théorie de la structure sociale [Nadel, 1957], Nadel définit une relation entre deux individus comme étant une somme d'interactions. Pour lui, les relations sont aussi structurantes que l'absence de relations entre individus. La présence ou l'absence d'une relation est la conséquence d'une norme sociale qui tend à définir des rôles dans une société. Les individus jouent alors des rôles ce qui impose qu'ils se comportent comme le rôle l'indique. Le rôle indique les relations que l'individu doit entretenir ou le contraire, celles qui lui sont interdites.

CHAPITRE

3

Etat de l'art

Préambule

L'état de l'art constitue ce qui se fait de mieux par rapport à notre problématique. Nous commençons celui-ci par un bref historique de la démocratie électronique afin d'aborder les applications qui nous intéressent : les débats en ligne. Nous présenterons ensuite l'annotation par une approche informelle, considérant cette fonctionnalité comme la plus adaptée à une pratique web2.0 pour les débats en ligne.

Sommaire

3.1	La démocratie électronique	71
3.2	Les débats en ligne	76
3.3	L'annotation	95
3.4	L'analyse des SocNets	112
3.5	Limites de l'état de l'art	116

3.1 La démocratie électronique

Democracy is the government of
the people, by the people, for the
people

ABRAHAM LINCOLN

Dans cette partie nous allons présenter le domaine de la démocratie électronique afin de mieux situer notre proposition. La démocratie électronique est l'usage des sciences et technologies de l'information et de la communication en vue d'améliorer la pratique de la démocratie. Les applications de la démocratie électronique couvrent toutes les étapes de la prise de décision, de la concertation au vote. La démocratie électronique est un domaine plus ancien qu'Internet, on en parle depuis les premiers outils de communication comme le télégraphe. Cependant, aujourd'hui, Internet et plus particulièrement le Web est devenu le terrain de prédilection des expérimentations du domaine. La démocratie électronique se décompose en plusieurs domaines. Les débats 2.0 font partie du champ de la concertation électronique. Cependant, bien qu'un découpage de la démocratie électronique en plusieurs champs soit possible, ces différents champs sont en interactions. En particulier, la concertation électronique si elle n'est pas liée à un processus de prise de décision et à un processus d'information perd de son intérêt et de son efficacité.

3.1.1 Histoire

La démocratie émerge en Grèce au Vème siècle. Afin de satisfaire aux envies, nécessités et besoins du plus grand nombre, l'homme a cherché des moyens de faire participer chacun dans les processus de décision qui le concernent. La palabre dans certaines tribus africaines en est un très bon exemple. Chaque membre de la communauté prend la parole et exprime son opinion jusqu'à ce qu'un consensus soit trouvé [Bidima, 1997].

S'il est possible de prendre des décisions collectives à l'issue de débats en petits groupes (débat en présence, suivi d'un vote, débat jusqu'à consensus, vote simple etc.), la concertation collective devient techniquement impossible dans des groupes dont le nombre d'acteurs est plus important. Il faudrait que chacun puisse prendre la parole et être écouté par l'ensemble des participants, et qu'une négociation collective ait lieu suite à ces prises de parole. Un tel dispositif ne permet pas de débat faisant intervenir un grand nombre d'acteurs en raison du temps important qu'il faudrait accorder à la prise de parole individuelle et aux négociations collectives.

Pour pallier cette difficulté, les démocraties occidentales actuelles ont recours à la représentativité : un groupe délègue à une personne le soin de défendre les intérêts du groupe et lui confère un pouvoir d'action. La position de représentant est la plupart du temps légitimée par un vote. Le représentant est alors dit « élu ». Un problème de confiance est inhérent à la représentativité. Le groupe a délégué un pouvoir d'action à son élu et espère que celui-ci l'utilisera judicieusement. Cependant, le lien entre une décision prise par un élu et les idées ou opinions partagées par le groupe est rarement explicite.

L'une des attentes de la démocratie électronique est d'augmenter le champ de la concertation et garantir la transparence des systèmes administratifs et politiques. Internet vient supplanter ses prédecesseurs que sont le télégraphe, la téléphonie et la télévision. Le Web devient le premier canal de communication et devient donc le nouveau champ d'expérimentation de la démocratie électronique, le nouvel espace public. Cet espace « *est*

un des concepts fondamentaux de la démocratie. Il désigne le lieu symbolique où peuvent s'exprimer toutes les opinions qui structurent le jeu politique, où l'on traite des questions relevant de la collectivité » [Testard-Vaillant, 2008].

Cependant, il ne suffit pas d'un espace public pour garantir la démocratie. Il faut aussi pouvoir accéder à l'information. Les capacités cognitives de l'individu restant limitées, il faut faciliter la tâche de tri et d'évaluation de l'information : « (...) pour être parfaitement éclairé, le citoyen a besoin d'information mais aussi d'information sur l'information » [Vedel, 2003].

Si le Web a évolué de médium réservé à des experts vers une place virtuelle ou la démocratie par les utilisateurs devient une réalité [Breindl et Francq, 2008], le Web 2.0 a aussi apporté un lot de problématiques qui ne sont toujours pas résolues. La démocratie est d'après Abraham Lincoln : *le gouvernement du peuple, par le peuple, pour le peuple*. La place du citoyen est donc centrale dans la démocratie et face au constat d'une baisse de la participation dans les processus démocratiques traditionnels, Internet amène beaucoup d'espoirs. Internet par rapport à ses prédecesseurs apporte une forte interactivité, un nouveau modèle de communication où les utilisateurs ne sont pas uniquement consommateurs mais aussi producteurs de contenu. De plus, les concepts fondamentaux du Web 2.0 de participation et de décentralisation des savoirs sont semblables à ceux de la démocratie électronique. Pourtant, la démocratie électronique reste un champ d'expérimentation et n'est pas encore une pratique largement établie.

3.1.2 Les domaines de la démocratie électronique

La démocratie électronique pour [Breindl et Francq, 2008] est à distinguer de la gouvernance électronique et de l'administration électronique. La gouvernance électronique est l'utilisation de technologies de l'information et de la communication dans le but d'informer, solliciter et augmenter la participation des citoyens dans la gouvernance des affaires publiques par le biais du processus politique. La gouvernance électronique et l'administration électronique font référence à des tâches administratives ou publiques dont la mise en ligne est portée par les gouvernements tandis que la démocratie électronique fait plutôt référence à la mise en ligne, le développement et l'usage par les citoyens d'outils électroniques pour la démocratie. La gouvernance électronique est une approche Top-down alors que la démocratie électronique est une approche Bottom-up.

La démocratie électronique est séparée en plusieurs champs d'application. Ces champs couvrent toutes les dimensions de la démocratie aujourd'hui, de la prise d'information en ligne aux processus participatifs et à la prise de décision. Nous dressons une liste de ces différents champs ici et des applications phares des différents champs :

- **L'information.** L'un des premiers champs couverts par le Web. Parmi les missions des institutions, il y a la diffusion de l'information concernant les politiques publiques dans un souci de transparence.
- **La construction de communauté et le travail en communauté.** On retrouve dans

les domaines de la **démocratie électronique** le besoin d'outils de consolidation et d'organisation de groupes d'opinions et de groupes de pressions. On retrouve aussi ce besoin comme transversal à l'ensemble des parties prenantes lors de gestion de biens communs.

- **La consultation.** Il s'agit de récolter des avis sur une politique publique. La différence primordiale avec une mesure d'opinion est que l'on attend pas une information quantitative sur les choix possibles mais plutôt la proposition de nouveaux choix.
- **La défense d'une cause.** De nombreuses associations ou collectifs supportant une cause ont adopté l'usage du Web pour gagner en visibilité, récupérer des fonds par des dons en ligne.
- **Le support de campagne d'un candidat.** Plusieurs hommes politiques ont adopté Internet comme une solution pour communiquer sans passer par les intermédiaires classiques. Cette solution leur permet d'adopter des positions plus subjectives et de garder la maîtrise de leurs discours.
- **La délibération,** la prise de décision, le vote. Les outils dédiés à la prise de décision se restreignent en général à des dispositifs de vote en ligne.
- **L'analyse de discours.** Permettant plus d'interactivité avec le média qu'il soit vidéo, son ou texte, le Web offre l'opportunité à tous d'analyser le contenu des discours politiques, de revenir sur des passages et de confronter les discours.
- **Les plans d'aménagement de territoire.** En se basant sur les outils de cartographies géographiques ou de représentation d'espace en 3D, il est devenu plus facile de faire des propositions "jetables" de plans d'aménagements.
- **La mesure de l'opinion.** En règle général à base de sondages ou d'analyse de discours dans des espaces de discussion, les outils de mesure d'opinion sont très prisés pour permettre essentiellement de faire des pronostiques.

Les débats 2.0 se positionnent selon nous sur les axes de la construction de communauté, de la consultation et de mesure d'opinion. Cependant, comme défendu par [Vedel, 2003], la démocratie électronique est un processus faisant intervenir 3 axes fortement liés :

1. L'axe de l'information
2. L'axe de la discussion
3. L'axe de la prise de décision

Dans cette vision, les débats 2.0 se situent donc à l'interface des deux axes de l'information et de la prise de décision.

3.1.3 Enjeux et objectifs

Quels sont les enjeux listés par les théoriciens ? Quels sont les objectifs des gouvernements ? Quelles sont les attentes des citoyens ? Selon les acteurs, les attentes sont différentes.

Les partisans de la démocratie électronique via Internet s'appuient sur des observations et des évidences pour défendre le bien fondé de l'usage d'Internet [Price, 2006] :

- Le nombre de débats en ligne augmente bien qu'il reste peu important, signe d'une demande citoyenne.
- Les débats en ligne permettent des interactions de groupe au delà des contraintes d'espace et de la diversité des participants. Les catégories socio-professionnels ne sont pas en général des critères pour participer.
- Les coûts de mise en place sont réduits.
- Ils augmentent la visibilité.
- L'anonymat et la nature textuelle des participations réduit la dominance sociale. Internet n'est plus réservé à une élite.
- Internet débride les introvertis, les participants se sentent plus libre d'exprimer leurs opinions.

Mais les solutions actuelles sont encore limitées, elles ne sont pas adaptées à la participation. En particulier, elles ne permettent pas vraiment aux citoyens de s'éduquer, s'entraîner et véritablement avoir un impact sur les décisions [Fung, 2006]. Les débats publics devraient permettre de rendre légitime les politiques publiques en évaluant le taux de support de celles-ci. Ils devraient aussi produire des décisions plus directement applicables en impliquant les citoyens dans leurs réalisations.

Pour les théoriciens, il s'agit donc de débrider la parole publique afin de réintroduire le citoyen dans la politique, de retourner à la vraie démocratie par le peuple et pour le peuple. Cela passe notamment par l'entraînement aux débats, l'amélioration des compétences nécessaires au débat.

Pour les citoyens, il s'agit de pouvoir participer et suivre les débats qui les concernent. C'est à dire qu'ils puissent s'informer et émettre leurs avis dans des discussions relatives à des décisions qui les intéressent, les impliquent directement. Il s'agit aussi pour le citoyen d'être sûr que sa participation servira à quelque chose, qu'elle sera au moins lue par les autres et non perdue dans la masse.

Enfin, pour les décideurs, le gouvernement ou les gouvernants, il s'agit de mesurer les tendances, d'évaluer l'éventail des possibles face à une situation et donc d'éviter les crises. Si une décision est prise suite à une concertation efficace, on espère que la décision, plus légitime et mieux comprise par les citoyens sera plus facilement acceptée.

3.1.4 Limites de la démocratie électronique

Considérée dans son ensemble et non domaine par domaine, les verrous exprimées sous forme de craintes de la démocratie électronique restent plus sociétaux que techniques. Certains parlent de fracture entre ceux qui pourront utiliser les outils et ceux qui ne pourront pas [Compiègne, 2011]. D'autres pensent que l'informatisation de la société risque de mener à une société du contrôle où tout le monde serait surveillé par tout le monde [?]. Il ne faut pas oublier que la démocratie électronique ne fonctionnera que si

il y a la volonté des décideurs et des citoyens de l'utiliser [Vedel, 2003]. Enfin, les projets gouvernementaux de démocratie électronique restent pour l'instant cantonnés à quelques expérimentations, bien souvent réduites à de la consultation en ligne. La démocratie électronique n'en est encore qu'à ses balbutiements malgré son âge.

Nous ne considérons dans nos travaux que les verrous propres aux débats 2.0. Même si nous restons concernés par les problèmes de confidentialité des données personnelles. Ces verrous, que l'on retrouve aussi dans [Flichy, 2008], concernent la problématique de la navigation dans une grande quantité d'information produite par les citoyens et pour les citoyens, la problématique de la légitimité des opinions exprimées et des leaders d'opinion et la problématique de l'insularisation et la radicalisation des communautés d'opinions. Enfin, nous soulignons la problématique de l'interopérabilité entre applications dédiées à la démocratie électronique.

3.2 Les débats en ligne

3.2.1 Vers des débats 2.0

Les citoyens sont demandeurs d'espaces publics ou communautaires d'expression d'opinion. Les sites institutionnels sont de plus en plus participatifs. Les citoyens se regroupent et s'expriment par le biais de forums, wikis ou autres outils, centrés sur des opinions ou des causes communes. Certains affichent personnellement leurs opinions via des SocNets ou des Blogs personnels.

Les citoyens sont demandeurs, et en l'absence de véritables outils et espaces consacrés aux débats 2.0, ils pallient ce manque en utilisant l'existant des applications de discussions et de communications. Parfois, ils innovent et on voit émerger de nouvelles fonctionnalités dans ces outils ou de nouvelles applications.

Nous allons dans cette partie présenter les outils classiques adoptés pour l'expression d'opinions en ligne. Nous présenterons quelques usages originaux qui ont attiré notre attention. Enfin, nous présenterons l'existant en matière de représentation des données issues des outils classiques de discussion et de communication ainsi que l'existant en matière de représentation de structures argumentatives.

3.2.2 Les outils du Web Social

Cette section présente un ensemble d'outils du Web social qui ont nourri notre réflexion pour la conception d'une plateforme de débats 2.0. Ces outils couvrent les usages du travail collectif, le partage de ressources, la discussion, le partage de connaissances ...

Nous adoptons la classification suivante tirée de [Breslin *et al.*, 2009] :

- Les outils de discussion en ligne.
- Les outils de partage de connaissance.

- Les outils de partage multimédia.
- Le tagging et le bookmarking.
- Le partage de sources de programmes.
- Les outils de réseaux sociaux en ligne, SocNets.
- Les outils permettant de regrouper les activités éparses d'une communauté, d'un utilisateur. On parlera d'agrégateurs d'activité.
- Les applications sociales en Entreprise, souvent désignés par outils d'Entreprise 2.0.

Nous présenterons parfois un outil précis et parfois nous considérerons l'ensemble des outils désignés sous une même appellation. Par exemple, nous traiterons de l'outil de discussion Quora et nous traiterons des Forums (regroupant sous forums plusieurs outils comme phpBB ou ip.Board).

Discussion en ligne

- Les mails ou courriers électroniques permettent à des utilisateurs de discuter en s'adressant des messages via un système de boîtes aux lettres. Il est possible d'adresser un même message à plusieurs destinataires et la possibilité de créer des listes de destinataire permet de faciliter la gestion de la discussion dans un groupe. Les individus sont ajoutés à la liste, les mails sont adressés à la liste et donc transmis à chaque membre de la liste. Les mails permettent de créer des fils de discussion car il est possible de répondre à un mail. A l'origine, le mail est une application d'Internet. La gestion des boîtes aux lettres est gérée par un serveur. L'utilisateur se connecte via une application cliente au serveur pour récupérer les messages de sa boîte aux lettres et en adresser de nouveaux. Aujourd'hui, il existe des clients accessibles par navigateur Web. L'utilisateur passe alors par un site pour consulter ses mails. Le mail est un outil de discussion asynchrone. Pas besoin que le correspondant soit connecté en simultanée à sa boîte mail pour pouvoir lui adresser un message. Les mails sont aussi utilisés pour partager une information ou pour transmettre des notifications, alerter l'individu qu'une certaine action s'est produite quelque part.
- La messagerie instantanée permet à plusieurs utilisateurs de discuter de façon synchrone. Les premiers serveurs IRC (Internet Relay Chat) apparaissent dès 1985. Avec IRC, les utilisateurs s'inscrivent à des salons et peuvent alors discuter avec toutes les personnes du salon. Aujourd'hui, la messagerie instantanée propose des variations autour de ce principe et est devenue multimédia en permettant d'utiliser le son ou la vidéo ou encore d'envoyer des images. La plupart des applications de messagerie instantanée permettent de gérer des contacts en les classant dans des groupes. La présence en ligne des contacts est affichée par le biais d'icônes traduisant des statuts (disponible, ne pas déranger, inactif depuis plus de 5 minutes, ...). La discussion peut toujours se faire par salons, mais aussi uniquement avec un seul interlocuteur. La messagerie instantanée en introduisant la possibilité d'afficher des statuts indiquant la disponibilité permet une forme de discussion élective et non intrusive [Caseau,

2009]. L'idée est que la personne choisit avec qui elle va discuter et que l'interlocuteur a signalé par son statut son accord (le statut disponible signifiant que la personne est prête à discuter avec ses contacts). Les outils de messagerie instantanée sont pour la plupart dotés d'un historique permettant de consulter les anciennes discussions. Il est aussi devenu possible d'adresser des messages à des personnes non connectées (mode asynchrone), ceux-ci recevant alors le message lors de leur prochaine connexion. La messagerie instantanée est essentiellement utilisée par le biais d'applications clientes mais il existe aujourd'hui des solutions pour intégrer une fonctionnalité de messagerie instantanée dans un site Web.

- Les forums d'aujourd'hui sont des applications Web permettant à des utilisateurs de déposer des messages dans un fil ou un arbre de discussion. Une discussion commence en général par la création d'un sujet (un titre, une phrase, parfois accompagné d'un message de départ). Les forums sont l'évolution des Bulletin Boards et des newsgroups, applications existantes aux premiers âges d'Internet. Ces espaces de discussions sont souvent cités comme les premiers outils communautaires car il y a une procédure d'entrée pour pouvoir participer aux discussions. Il faut partager les centres d'intérêts des autres inscrits à un forum pour s'y inscrire à son tour. La discussion est ici centrée sur les sujets ou Topics et non sur les interlocuteurs. On choisit de discuter d'un sujet, on ne choisit pas ses interlocuteurs. [Caseau, 2009] parle de "stigmergie dirigée", la discussion se fait par dépôts de message dans un espace où les lecteurs sont filtrés par leurs positions (centre d'intérêts, membres de la communauté). Les messages dans un Forum sont appelés des Posts, ils sont généralement sous forme textuelle. Deux messages peuvent être liés par un statut de réponse. Si la discussion est structurée uniquement par l'ordre chronologique d'émission des Posts, la discussion est qualifiée de Fil de discussion. Si la discussion est structurée par les statuts de réponses, la discussion est qualifiée d'arbre de discussion. La modération dans les forums est l'un des points crucial pour assurer la dynamique du forum. En règle général, les forums sont administrés par des utilisateurs ayant des droits particuliers leur permettant de supprimer ou modifier des messages, ou encore de supprimer l'accès à certains utilisateurs. Certains forums comme par exemple le site Slashdot¹ adoptent des systèmes de votes pour permettre une sorte de modération par les utilisateurs. Ainsi, les utilisateurs peuvent filtrer les discussions ayant le plus de votes favorables ou défavorables. Dans de nombreux forums, l'identité des utilisateurs est pondérée par une note "au mérite". Celle-ci est basiquement calculée sur la participation de la personne dans le forum. Dans le site Slashdot, celle-ci est calculée en fonction des votes des autres utilisateurs et donne de nouveaux droits à l'utilisateur.
- L'application Quora² est très proche des systèmes de forum. Quora est un système

1. <http://slashdot.org/>

2. <http://quora.com>

de questions-réponses. Il est possible de créer des fils de discussion mais cette fois, le point de départ de la discussion est nécessairement une question. La discussion est ensuite indexée en fonction de sa date et de son activité mais aussi en fonction de tags associés par les utilisateurs. Chaque élément de la discussion, chaque commentaire peut à son tour être commenté. Le fil de discussion peut donc devenir un arbre de discussion. Il est possible de voter pour ou contre chaque commentaire, et il est possible à tout moment de la discussion de créer une synthèse de la discussion.

- L'application TalkMap³ est un outil de discussion développé en Ruby et Javascript. TalkMap présente les discussions sous la forme de graphes noeuds-liens. Chaque noeud est l'équivalent d'un Post dans un forum. Les noeuds sont reliés par une relation de réponse et TalkMap intègre aussi des jugements permettant de signaler des accords et des désaccords. Il est possible de lier un Post à plusieurs autres et il est possible de regrouper les Posts dans des zones colorées associées à des thématiques. Enfin, la modération dans TalkMap se fait au mérite par le biais d'un système de vote.

Partage de connaissance

- Les blogs sont des outils de communication particuliers étant donné qu'un utilisateur expose par le biais d'articles ses idées. Et par le classement chronologique des articles d'un blog, il devient possible de suivre l'évolution de ses idées, son modèle, sa personnalité. Chaque article dans un blog peut être commenté par les lecteurs du blog. Là encore, les commentaires sont appelés des Posts et la structuration des Posts s'apparente à celle des Forums. La véritable différence entre un blog et un forum, repose sur la différence de statut entre l'auteur du blog et les lecteurs du blog. Les blogs sont souvent dits multimédia car l'inclusion d'une vidéo ou d'images dans un article est facilité dans la plupart des applications de blogs.
- Disqus⁴ se positionne comme une solution offrant la possibilité de commenter toute page d'un site comme dans un blog. Un code javascript est fourni par le système, en ajoutant celui-ci à votre site, toutes les pages du site deviennent commentables avec en plus des fonctionnalités comme la possibilité d'indiquer un "*I like*" ou "*I dislike*", de lier des tweets, d'ajouter des fichiers multimédia ...
- Les wikis sont des outils de co-rédaction. Le premier wiki date de 1995, inventé par Ward Cunningham. Un wiki permet la création, l'édition et la mise en réseau d'un ensemble de pages à la volée, depuis un simple navigateur Web, de manière très simple et rapide. La mise en réseau des pages repose sur l'utilisation de wikimots, une syntaxe particulière qui fait que lorsque un wikimot est écrit dans une page, celui-ci crée en même temps une autre page dont le wikimot est l'identifiant. Les deux pages sont alors liées par un lien hypertexte. Le premier wiki n'est pas qu'une invention technologique, c'est aussi une vision du travail collectif. Ward Cunningham dans son pre-

3. www.talk-map.com

4. <http://disqus.com>

mier wiki a volontairement laissé la possibilité à tout le monde d'éditer le contenu des pages, sans attribuer de statut particulier aux auteurs et sans proposer de dispositif d'inscription ou de filtrage des utilisateurs. Les wikis permettent à des individus de créer collectivement un ensemble de pages. Une page étant le résultat d'une somme d'édition les unes après les autres par des utilisateurs différents, celle-ci peut alors être vue comme étant une représentation textuelle d'un modèle partagé par les différents auteurs. L'édition dans un wiki est en règle générale asynchrone. Mais de nouveaux outils comme l'application Google Docs permet aux utilisateurs de co-rédiger de façon synchrone un document.

- Le MindMapping est l'utilisation de cartes heuristiques pour représenter des associations d'idées. Avant d'être une application Web, le mindmapping est surtout un outil de production de synthèses de réunions. Le modèle est explicité sous la forme d'un graphe dont l'idée principale est au centre et autour gravitent toutes les idées reliées. Cette représentation nécessite peu de conventions et est simple à lire. Un Mind-Map peut être construit au cours d'une réunion puis être partagé pour produire un compte-rendu de la réunion. Elle n'est pas nécessairement éditée à plusieurs, mais est souvent distribuée, elle sert donc de support à la communication. Le MindMapping repose sur la création de liens entre des idées, des phrases. En cela, c'est une pratique à comparer à l'annotation.
- DebateGraph⁵ est un outil en ligne utilisé par un groupe d'experts pour proposer une vision structurée sous forme de graphes noeuds-liens de débats. Les experts créent des noeuds à partir de phrases récupérés sur d'autres ressources (Web, multimedia) comme par exemple une phrase dans un blog ou un extrait d'une interview. Les experts relient ensuite les noeuds par des liens dont ils fixent la sémantique (opposition, soutien, cause, ...). Les débats structurés sont ensuite proposés en consultation aux autres utilisateurs qui peuvent les critiquer par un système de commentaires. Pour l'instant, la base de donnée de DebateGraph est intégralement remplie et structurée par des experts, mais à terme, les développeurs de DebateGraph espèrent pouvoir automatiser une partie de ces tâches.
- PearlTrees⁶ est un outil développé en flash permettant à chacun d'organiser son réseau sémantique de ressources en ligne. Il s'agit pour un utilisateur de créer là encore un graphe noeuds-lien dans lequel les noeuds sont appelés des perles. L'utilisateur est représenté par une perle centrale. Ensuite, il peut créer des perles thématique (des grandes catégories, des centres d'intérêt ...) qui seront reliées à son noeud utilisateur. Puis il peut créer des perles dédiées à des ressources en ligne, des pages Web qu'il a consulté. Ainsi PearlTrees s'apparente à un système de bookmarking mais dont la représentation sous forme de graphes noeuds-liens permet de créer des cartes

5. <http://debategraph.org>

6. <http://www.pearltrees.com/>

heuristiques (comme une mindmap) et de faciliter l'association d'idées. Les cartes produites sont partagées entre les utilisateurs.

Partage Multimédia

De nombreux sites permettent aux utilisateurs de déposer des contenus multimédias et de les partager avec les autres. Dans le contenu vidéo, le site le plus connu est probablement Youtube⁷. Le site Flickr⁸ permet lui d'échanger des photos. Le principe est en général toujours le même. Le site permet à un utilisateur enregistré de déposer du contenu accessible par tout le monde. Il peut catégoriser le contenu par un système de tags. Et la plupart de ces systèmes permettent de voter et de laisser un commentaire sur les ressources qu'ils hébergent, afin notamment d'améliorer l'indexation du contenu. La fonctionnalité de base de ces systèmes est avant tout l'hébergement pour faciliter le partage. Le contenu est stocké sur le serveur de l'application et peut ensuite être inclus dans n'importe quelle page Web.

Tagging et Bookmarking

Le tagging fait référence à une pratique visant à associer des mots clefs à des ressources pour les catégoriser. Selon les approches, l'ensemble des mots-clefs est plus ou moins structuré (ensemble à plat, organisé en hiérarchie, lié à une ontologie ...). Cette pratique peut désormais se faire de façon collaborative et rejoint celle du bookmarking. Le bookmarking consiste à enregistrer les pages Web que l'on consulte et que l'on souhaite surveiller, ou garder comme référence (c'est une façon de gérer ses sites favoris). L'application Delicious⁹ permet d'enregistrer les pages Web que l'on consulte et de leur associer des mots-clefs. Ces informations que l'utilisateur crée sont ensuite partagées avec l'ensemble des utilisateurs de l'application. Delicious fournit ainsi une nouvelle façon de chercher et de naviguer sur le Web en se basant sur une catégorisation des ressources créée par les utilisateurs. Ce type de catégorisation créée collaborativement par un groupe est appelé une folksonomie [Vander Wal, 2007].

Partage de code source

Dans le domaine du développement en informatique, et en particulier dans les communautés du libre ou le travail est distribué entre un grand nombre d'acteurs, le partage de code source est une fonctionnalité centrale. Si à l'origine, les premiers systèmes de partage reposaient uniquement sur le protocole FTP (File Transfer Protocol), aujourd'hui les systèmes de partage de code source proposent des fonctionnalités évoluées et sont parfois

7. <http://youtube.com>

8. <http://www.flickr.com>

9. <http://www.delicious.com>

accessibles via une application Web. Les fonctionnalités de base des systèmes de partage de code source sont la possibilité de déposer du code et de récupérer du code, la possibilité de restreindre les droits d'accès à des parties du code en fonction des acteurs, le suivi des versions et modifications entre chaque dépôt de code. L'arrivée d'applications Web pour partager le code, comme par exemple l'application Github¹⁰, permet d'intégrer de nouvelles fonctionnalités donnant une meilleure visibilité de l'activité et facilitant les interactions entre développeurs. Par exemple, il est possible de consulter des statistiques comme le nombre de téléchargement des sources ou le nombre de modifications entre les versions, le nombre d'acteurs différents ayant ajouté des modifications, il est aussi possible de discuter autour du code facilitant ainsi la transmission de celui-ci ou encore de noter, juger la qualité du code ou son intérêt.

SocNets

Les SocNets ou outils de réseaux sociaux permettent à des utilisateurs de gérer des contacts et d'échanger des informations avec ces contacts. L'application de référence est probablement Facebook regroupant plus de 500 millions d'utilisateurs actifs en 2010¹¹. Un SocNet permet de filtrer les personnes qui ont accès à votre espace personnel, de choisir ses interlocuteurs. Un peu comme un blog, un SocNet encourage l'utilisateur à afficher ses idées, mais dans un SocNet, en réglé générale, les Posts sont courts et consacrés à de la notification. Sur Twitter¹² par exemple, les Posts sont limités à 140 caractères. On n'écrit pas vraiment sur Twitter ou Facebook pour créer une discussion mais plutôt pour afficher son statut, notifier de sa dernière action, d'un goût particulier ... On fait rapidement partager une information à son réseau. Twitter, désigné comme application de micro blogging, est centré sur cette fonctionnalité de notification. On ne peut pas vraiment créer une discussion et lier les posts entre eux. On les accumule dans son espace personnel, ils sont listés selon leurs dates d'arrivée et sont transmis à tout votre réseau par le biais d'un flux d'activités partagé.

Agrégateurs d'activité

Avec l'affluence d'applications se créent des problèmes d'identification des personnes. Les utilisateurs se retrouvent à jongler avec des couples de nom d'utilisateurs et de mots de passe pour chaque application à laquelle ils souscrivent. Il y a là une double problématique :

- Comment faciliter la tâche d'identification de l'utilisateur, ou comment lui fournir un système d'identification simplifié pour s'inscrire plus facilement à de nouvelles applications sans devoir saisir à chaque fois l'ensemble de ses données. Ceci concer-

10. <https://github.com>

11. <http://www.facebook.com/press/info.php?timeline>

12. <http://twitter.com/>

nant notamment l'ensemble de ses contacts qu'il devra ressaisir d'une application à une autre.

- Comment permettre à un utilisateur de facilement visualiser l'ensemble de son activité dispersée sur de nombreuses applications.

Le système Openid¹³ est proposé pour répondre à la première des deux problématiques. Openid fournit aux utilisateurs un identifiant unique. Les applications compatibles avec Openid permettent donc à l'utilisateur de s'inscrire très simplement en utilisant cet identifiant. Mais aujourd'hui, certains services, comme Facebook ou Google, permettent aux autres services d'adopter leurs données d'identification. Ainsi, l'identifiant Facebook d'un utilisateur va aujourd'hui lui permettre de s'inscrire facilement à de nombreuses applications.

Bien que la plupart des applications tiennent à conserver les données concernant leur réseau d'utilisateurs, elles fournissent aujourd'hui des solutions pour accéder à certaines de ces données. Elles fournissent souvent des flux de données facile à parcourir et traiter. Des API ou des services Web basés sur HTTP permettent de manipuler les données des applications. Ainsi, il devient possible de bâtir des applications composites (appelés Mashup) qui vont aller récupérer les données d'un utilisateur dans différents services et lui proposer une visualisation globale de ses activités. Nous appelons ces applications des agrégateurs d'identité. L'application FriendFeed¹⁴ permet par exemple à un utilisateur de regrouper des flux de données le concernant générés dans d'autres applications comme son wall Facebook, ses tweets, mais aussi des flux de syndication RSS ou Atom.

Enfin, l'utilisateur adopte en général une information unique lors de son inscription à la plupart des services, son adresse mail. L'adresse mail d'un utilisateur étant unique, il est possible d'identifier un utilisateur au travers de celle-ci dans plusieurs applications. C'est peut-être l'un des biais utilisé dans l'application 123people¹⁵ pour agréger des données concernant un utilisateur.

Entreprise 2.0

Le terme d'Entreprise 2.0 englobe toutes les pratiques visant à améliorer le travail en entreprise par le biais d'outils du Web social, autrement dit, d'outils du Web 2.0. La plupart de ces outils ont déjà été détaillés précédemment. Cependant, nous allons parler de deux classes particulières d'outils souvent adoptés en entreprise : les CMS et les Groupware.

- Les CMS pour *Content Management System* sont des outils visant à permettre la mise en place simplifiée d'un site Web dont le contenu est éditabile. Un administrateur a la possibilité, par le biais d'une interface d'administration appelée *backoffice* de modifier tout le contenu et la mise en page du site. Les CMS sont adaptés pour créer des sites d'information ou l'administrateur édite les articles, du contenu et ou les

13. <http://openid.net>

14. <http://friendfeed.com>

15. <http://www.123people.fr>

lecteurs peuvent parfois laisser des commentaires comme sur un blog. Les CMS sont livrés avec de nombreuses fonctionnalités et peuvent être enrichis par le biais de plugins. Par exemple il est possible d'ajouter des modules de vente en ligne dans un site en Joomla¹⁶. Certaines applications de blog comme Wordpress¹⁷ développent tellement leurs possibilités fonctionnelles et leurs plugins qu'elles tendent à proposer des offres proches de celles des CMS.

- Les Groupware sont des outils visant à remplir les besoins propres à l'entreprise. On retrouve donc des outils de gestion de projets avec des agendas partagés, des carnets d'adresses ... mais aussi tous les outils de capitalisation de connaissance comme des wikis, des blogs, des outils de tagging et de bookmarking. Le terme groupware en lui-même désigne l'ensemble des outils adoptés pour le travail de groupe. Certains logiciels comme egroupware¹⁸ se spécialisent dans ce service. Là encore, les plateformes de groupware adoptent les fonctionnalités des outils du Web social. Finalement, l'avantage de ce type de solutions réside dans l'agrégation des fonctionnalités dans une même plateforme.

L'ensemble des outils que nous venons de présenter permet de souligner l'existence de solutions complémentaires. Et ce sont aujourd'hui les outils les plus adoptés dans un contexte de travail collaboratif ou de concertation citoyenne en ligne (en ajoutant l'usage des sondages). Ces outils sont en constante évolution. Des solutions proposant une agrégation d'outils comme des CMS ou groupware sont de plus en plus fréquentes. Enfin, l'arrivée d'autres médias comme la vidéo et le son en ligne souligne qu'aujourd'hui, les formes de communication en ligne ne sont plus limitées qu'à une expression textuelle.

Nous avons vu dans cette partie des outils très populaires dans leurs catégories. En particulier, nous avons présenté ici les outils classiques de discussion. Avoir une connaissance de ces outils est indispensable pour dresser un état de l'art des débats 2.0 mais n'est pas suffisant. Les outils de discussion présentés ici manquent généralement de structure pour permettre des discussions à grande échelle. Ils n'offrent pas de mécanismes simples pour présenter une vision d'ensemble de la discussion. Nous soulignons dans la section suivante l'importance de l'expression d'opinions comme élément structurant des discussions.

3.2.3 L'existant en matière d'expression d'opinion

La première approche pour une consultation en ligne est en général d'utiliser un outil de sondage. Le sondage présente l'intérêt d'être facilement traitable à moindre coût quelque-soit le nombre de personnes participant. Cette solution passe à l'échelle. Cependant, le sondage ne permet d'obtenir des opinions qu'en fonction des questions posées et,

16. <http://www.joomla.fr>

17. <http://www.wordpress-fr.net>

18. <http://www.egroupware.org>

de plus, la formulation des questions peut influencer pour beaucoup les réponses au sondage. Pour qu'un sondage soit représentatif, il faut aussi avoir sélectionné la population interrogée. Enfin, l'opinion mesurée n'est valable qu'à l'instant du sondage. Un sondage ne présente pas l'évolution d'une opinion mais une opinion "figée", à un instant donné. Le sondage est donc une bonne solution lorsque l'on souhaite consulter l'opinion des citoyens sur une question précise à un moment précis. Le sondage a mauvaise image auprès des utilisateurs. Il faut donc, en plus de sélectionner les utilisateurs pour qu'ils soient représentatifs, sélectionner des utilisateurs qui soient volontaires pour remplir le sondage.

Une autre approche consiste à se baser sur des pratiques sociales ou les utilisateurs partagent explicitement des objets communs d'interaction comme des centres d'intérêts, l'appartenance à un parti ou une association, la déclaration de soutien d'une cause ... Cette solution peut se faire via les SocNet, Facebook par exemple propose d'être "fan" d'une page particulière. Il est possible d'être fan de la page d'une association, d'une personnalité politique ou d'un parti politique. En observant les mots-clefs partagés par des utilisateurs via un outil de tagging ou de bookmarking, on peut aussi proposer une extraction des tendances de l'opinion des utilisateurs. De la même façon, on peut récupérer des mots clefs utilisés dans des Blogs. [Gaines et Mondak, 2009] proposent par exemple de regrouper les utilisateurs partageant une même opinion politique dans Facebook. Ils se basent sur la possibilité dans une application de déclarer sa position entre : très libéral, libéral, modéré, conservateur, très conservateur, apathique, libertaire, ou autre.

Dans cette approche, on considère que les citoyens expriment naturellement leurs opinions via les pratiques sociales du Web. Les opinions que l'on va recueillir par cette solution sont dirigées mais non bornées par l'objectif de la consultation (qui se traduit plus ici comme une sorte de moissonnage de l'information, vu que l'on se contente de récupérer une information existante). Si l'on cherche à avoir une représentation de l'appartenance des français à des partis politiques par exemple, on pourra récupérer sur Facebook le nombre de fans des différents partis. Il est possible dans ce cas que l'on découvre sur Facebook de nouvelles tendances comme l'existence de nouveaux partis ou d'associations proches de partis politiques. Un avantage est que les données sont ici fortement contextualisées. On pourra par exemple recouper l'appartenance à des partis politiques via les pages fans Facebook avec d'autres informations personnelles comme l'emploi actuel ou la région de résidence par exemple. Enfin, avec cette approche, il est souvent possible d'observer une évolution temporelle de l'opinion. La déclaration "fan" dans Facebook est datée et l'utilisateur peut devenir fan d'autres pages. Par contre, cette approche peut s'avérer couteuse en temps de traitement des informations. Il faut souvent se familiariser avec des formats de données ou des apis propres à l'application ou les données doivent être récupérées. Et la quantité d'information peut entraîner un coût de récolte et de traitement. Certaines applications permettant de générer ces données ont aussi un coût de modération pour garantir que les données soient pertinentes.

Les outils de communication et de discussions en ligne sont souvent le siège de débats d'opinions. Cependant, les outils ne sont pas prévus pour donner un aperçu de cette opi-

nion, une synthèse de la discussion. En règle générale, ces outils permettent de s'exprimer sous la forme de commentaires libres textuels. Cette forme de saisie laisse toute liberté à l'utilisateur d'exprimer son opinion comme il lui semble. Mais cette liberté entraîne un fort coût de traitement pour extraire des données quantitatives sur les opinions émises et sur la polarité de ces opinions. [Plantié *et al.*, 2008] présente une approche d'extraction d'opinions applicables à des articles de blogs, des Posts sur des forums ... des commentaires de nature plein texte. D'autres méthodes d'extraction d'opinions par fouille de textes existent. Elles reposent sur les technologies du domaine du Traitement Automatisé de la Langue Naturelle. Ces méthodes sont en règle générale bien adaptées pour des contextes donnés ou dans des corpus de discussions clos. Les limites de ces méthodes sont atteintes quand l'expression d'une opinion est faite via une expression non connue à l'avance ou lorsque les utilisateurs usent de l'ironie dans leurs messages. Comme pour les autres approches se basant sur des interactions sociales en ligne, cette approche dépend aussi de la pro-activité des utilisateurs.

Les approches innovantes sur la question adoptent quand à elle une autre position. Elles permettent de structurer l'argumentation, de préciser des liens sémantiques entre des commentaires, d'expliquer l'opinion exprimée. Cette explicitation de la structure argumentative et du sens peut se faire au moment de la saisie d'un commentaire ou après coup, par retraitement des données et enrichissement de celles-ci. Par exemple, plusieurs sites comme convinceme¹⁹ proposent des jeux de débat où les utilisateurs peuvent poster un commentaire en précisant si il est en faveur d'une position ou opposé à cette position. Dans ce cas, l'opinion est clairement explicitée lors de la saisie. L'outil debategraph²⁰ permet à des professionnels d'expliquer la structure d'un argumentaire en saisissant des commentaires et en les liant sémantiquement. L'un des usages qui en est fait est de définir des liens sémantiques entre des commentaires récupérés sur le Web. Les liens sémantiques permettent de définir une structure de débat en ajoutant des liens de conséquence, d'opposition, de soutien, mais aussi de composition. Cet outil demande une intervention humaine pour restructurer et lier des débats en cours ou ayant eu lieu dans divers espaces de discussion-communication sur le Web. L'outil TalkMap²¹ permet de créer un graphe de discussion en liant les posts les uns aux autres au moment de leur rédaction ou après coup. Il permet de typer les relations entre les posts (*agree*, *disagree* ou *answer*), de créer des régions dans lesquelles sont regroupées des parties de la discussion en fonction d'une thématique (par exemple dans la discussion "mac vs pc", il y a une région où les arguments échangés portent sur les Jeux vidéos, une région qui parle de linux ...).

Talkmap est un outil de discussion récent et plusieurs échanges ont eu lieu entre les développeurs de cet outil et notre équipe de recherche. Des outils comme Talkmap ou debategraph constituent à nos yeux l'état de l'art des outils de discussion. La structure de la

19. <http://www.convinceme.net>

20. <http://debategraph.org/home>

21. <http://www.talk-map.com>

discussion n'est pas réduite à un fil, dans les deux cas il est possible de structurer la discussion sous la forme d'un graphe dont les relations entre les arguments sont typées. Ce qu'il manque encore à ces outils, et c'est là que nous nous positionnons, c'est l'adoption de formats de données permettant l'interopérabilité. Aujourd'hui, les discussions en ligne ne se font pas dans un environnement unique et contraint. Il faut donc à des outils comme Talk-Map ou DebateGraph la possibilité d'enrichir leur contenu en agrégeant des discussions émises sur des blogs, des forums, sur youtube ...

3.2.4 L'existant en représentation des débats

Nous avons défendu que les outils de discussion-communication sont centraux dans la mise en place de débats 2.0. Représenter les données issues de ces outils est donc l'un des besoins de la représentation des débats. Un autre besoin est de pouvoir clairement représenter une opinion et de lier les différents éléments de débats entre eux. C'est notre approche pour représenter la structure argumentative des débats. Enfin, l'importance des relations entre individus, la possibilité de représenter le réseau social des citoyens prenant part aux débats permettra d'avancer des solutions pour faire face aux verrous de l'insularisation des communautés et de la légitimité des leaders d'opinion.

Nous allons présenter dans cette partie le vocabulaire SIOC qui permet de représenter les données issues d'outils de discussion-communication. Nous présenterons aussi le vocabulaire FOAF qui permet de représenter un réseau social. Concernant l'expression d'opinion, nous présenterons des approches différentes que nous comparerons : l'utilisation de vocabulaires dédiés à la représentation des données issues de pratiques de tagging et l'utilisation de vocabulaires dédiés à la représentation d'émotions. Enfin, nous présenterons des vocabulaires dédiés à la représentation de structures argumentatives.

Représentation des données de la discussion

Le vocabulaire SIOC (Semantically Interlinked Online Communities) permet de représenter les productions d'un utilisateur dans un outil de discussion-communication. Plus largement, il permet de représenter les productions d'un utilisateur dans un espace d'interactions sociales en ligne, un espace communautaire. Les classes centrales pour notre travail sont la classe *UserAccount* qui permet de représenter le compte de l'utilisateur et la classe *Post* qui permet de représenter un Post dans un blog, un forum ou un autre outil de discussion. Grâce aux relations *reply_of* et *has_reply*, il est déjà possible de structurer la discussion comme dans un Forum classique. L'image 3.1 reprend l'ensemble des classes du vocabulaire SIOC et leurs relations.

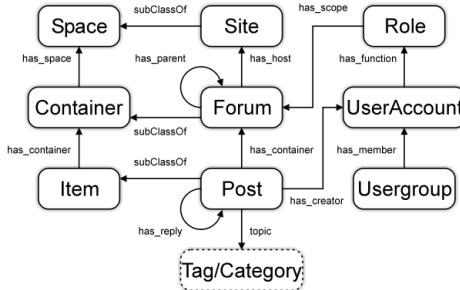


FIGURE 3.1 : Classes et relations du vocabulaire SIOC

Représentation du réseau social

Le vocabulaire FOAF (Friend of A Friend) permet de représenter un réseau social. L'utilisateur y est représenté en tant que personne (via la classe *Person*) et une relation entre deux personnes est définie via la relation *knows* du vocabulaire. Cette solution pour représenter un réseau social est très simple et ne permet pas en soi de représenter des réseaux d'interaction, des réseaux hétérogènes et multi-relationnels. Cependant, le vocabulaire FOAF et le vocabulaire SIOC sont conçus pour être complémentaire. Une personne (*Foaf:Person*) peut avoir plusieurs comptes utilisateurs (*Sioc:UserAccount*). De plus, les vocabulaires pouvant être étendus, il est tout à fait possible de définir des extensions de SIOC et FOAF permettant de représenter des réseaux d'interaction et des réseaux hétérogènes multi-relationnels. De nombreuses extensions proposent par exemple des relations étendant la relation *Foaf:knows*.

Le travail de Guillaume Ereteo a permis de définir l'ontologie SemSNA (pour Semantic Social Network Analysis) [Erétéo et al., 2009b]. Il propose un usage de cette ontologie pour déduire un réseau social hétérogène et multi-relationnel à partir d'un réseau d'interactions [Erétéo et al., 2009a]. SemSNA permet de représenter les résultats de nouvelles formes d'analyse de réseaux sociaux basées non seulement sur la topologie du réseau social mais aussi sur la sémantique des relations dans le graphe [Erétéo et al., 2008]. SemSNA est conçu pour être utilisé avec FOAF et SIOC.

Représentation de structures argumentatives

Nous avons vu que de nombreuses applications de discussion en ligne existent. Celles-ci offrent des formes d'interactions différentes adaptées à des besoins parfois complémentaires parfois indépendants, parfois concurrents. Cependant, qu'en est-il de la représentation des données dans ces contextes ? La plupart de ces applications ne sont pas interopérables, les discussions faites dans un contexte, dans une application ne peuvent pas migrer vers une autre application. Nous avons vu dans la partie précédente que pourtant, il existe des solutions pour représenter les données de la discussion en se basant sur des vocabu-

laires RDF comme SIOC et FOAF. Ces solutions permettent de représenter des structures simple de discussion comme une suite de questions-réponses, un fil de discussion, les commentaires associés à une ressource... Mais elles n'offrent pas les concepts nécessaires à la représentation d'une structure argumentative complexe. Nous allons ici présenter différentes approches qui s'intéressent à la représentation de méta-données "plus riches", permettant d'imaginer des structures de discussion plus complexes.

Dans le domaine de l'intelligence artificielle, la modélisation de l'argumentation (processus de création et de gestion des arguments) est devenu un champ d'étude pour la mise en place de systèmes de prise de décisions. L'argumentation est vue comme un cadre conceptuel permettant une nouvelle lecture de formes de raisonnements logiques [Villata, 2010]. Il s'agit alors d'adopter un modèle de l'argumentation pour construire un système logique permettant à un système multi-agent de prendre des décisions à partir de bases de connaissances ou tout n'est pas connu (monde ouvert) et où il peut y avoir des contradictions. L'argumentation est présentée comme une méthode permettant d'évaluer en cas de conflits entre arguments lequel est meilleur, plus acceptable que les autres. Dans ce domaine, deux théories sont souvent présentées :

- Le modèle de Toulmin [[Rahwan et Sakeer, 2006](#)], [[Chesñevar et al., 2006](#)] qui décompose un argument.
- Le modèle de Dung [[Villata, 2010](#)], [[Devred et Doutre, 2008](#)], [[Coste-Marquis et al., 2007](#)], [[Nielsen et Parsons, 2006](#)] qui structure une argumentation par la présence de relations d'attaque ou de soutien entre les arguments.

Le modèle de Toulmin [[Toulmin, 1958](#)] décompose un argument à partir des éléments suivants :

- **La revendication** (claim en anglais). L'idée défendue, ce que vous voulez que l'autre accepte comme vrai.
- **Les fondements** (grounds en anglais), parfois aussi appelé les données. Ce sont en général des faits indiscutables, on parle aussi de preuves, d'expertise.
- **La justification** (warrant en anglais). Ce qui permet de lier les fondements à la revendication. C'est ce qui permet de répondre à la question "en quoi votre revendication est juste en connaissance des fondements ?".
- **Les soutiens** (backing ou support en anglais). C'est un ensemble d'éléments qui viennent encore renforcer la justification.
- **Les réfutations** (rebuttal en anglais). C'est un ensemble de parades pour désamorcer des contre-arguments déjà connus (par exemple, "nous n'aborderons pas la discussion dans son aspect technique" permet d'éviter tous contre-arguments visant à souligner la difficulté technique de mise en œuvre de la revendication).

Le modèle de Dung [[Dung, 1995](#)] définit un cadre argumentatif comme étant une paire constituée d'un ensemble d'arguments AR et d'une relation binaire sur AR définissant les relations d'attaques entre les arguments. Le modèle est très simple, il ne s'intéresse pas à la structure interne des arguments et permet de construire des systèmes logiques non monotones. L'argumentation est ici abordée en vue d'être implémentée pour créer des systèmes

de décision. L'argumentation est alors ici à rapprocher à des systèmes logiques comme la logique défaisable [Nute, 1994]. Un ensemble de règles permet de dire si un argument est acceptable en prenant en compte un certain ensemble d'arguments. Par exemple, si un argument A n'est attaqué par aucun autre argument ou s'il est attaqué par un argument B lui-même attaqué par C alors l'argument A est acceptable. Dung introduit l'argumentation comme étant un système de prise de décision ou celui qui l'emporte est celui qui a le dernier mot, autrement dit, celui qui arrive à défendre ses arguments contre toutes attaques. L'expressivité du modèle de Dung est donc très limité mais permet de construire des systèmes de décision efficaces (en terme de calculabilité).

Enfin, des travaux venant du domaine du Web sémantique se sont intéressés à la représentation du contenu généré par les utilisateurs en ligne. Nous présentons ici des approches complémentaires à celles déjà présentées de FOAF ou SIOC, en ce sens qu'elles s'intéressent à représenter précisément ce que ne représentent pas FOAF et SIOC et qui ont trait à l'expressivité du langage. Le projet EmotionML [Baggia *et al.*, 2011] propose un cadre permettant de représenter des émotions. Le vocabulaire RDF NiceTag [Monnin *et al.*] permet de représenter la sémantique des actions de tagging. Enfin, ArgumentML [Delan-noy, 1999] et AIF [Rahwan et Sakeer, 2006] s'intéressent à la représentation de structures argumentatives.

Le langage EmotionML est bâti sur le parti pris qu'il est impossible de lister dans un vocabulaire toutes les émotions humaines existantes. Par contre, la structure de base, les concepts utilisés pour parler des émotions semble être communément admise par les scientifiques du domaine (psychologie, sciences cognitives ...). Du coup, le langage propose de fournir ces outils, ces concepts de base comme élément du langage pour pouvoir décrire des émotions. Il est question d'utiliser des dimensions, des catégories, de parler de déclencheurs, d'actions, de comportements, de changements physiques ... La motivation qui est à l'origine de ce langage est de permettre à un composant technologique de représenter et traiter des données liées aux émotions et de permettre l'interopérabilité entre plusieurs composants. EmotionML est une DTD pour XML. Les 3 cadres d'application visés par les auteurs sont :

- les annotations manuelles de données par des utilisateurs
- la reconnaissance automatique d'états émotionnels à partir de comportements utilisateurs
- la génération de comportements systèmes en réponse à des émotions utilisateurs

Bien que nous n'utilisions pas EmotionML dans nos travaux, nous nous sommes intéressés à ce langage dans le cadre applicatif des annotations manuelles de données par les utilisateurs. Nous avions empiriquement défini dans nos premières applications un ensemble d'émotions pour l'annotation manuelle. Nous défendons qu'il est plus utile de définir un ensemble d'éléments structurels permettant de représenter et définir via les applications, selon les contextes d'utilisation, de nouveaux tags émotionnels ou de jugements et de définir leurs relations (notamment opposition, contradiction et implication).

L'approche adoptée dans la définition du vocabulaire NiceTag part du constat qu'il

existe déjà de nombreuses solutions permettant de représenter un tag et la définition d'un tag. Cependant, ce n'est pas suffisant pour permettre de représenter la fonction du tag associé à une ressource par un utilisateur. Par exemple, si un utilisateur tagge une ressource avec le mot-clef "Tim Berners-Lee", est-ce que cela signifie que la ressource est écrite par Tim Berners-Lee, qu'elle s'adresse à Tim Berners-Lee ou qu'elle représente Tim Berners-Lee ? NiceTag propose donc des primitives suffisamment génériques pour pouvoir être étendue dans chaque cadre applicatif. La propriété "*is related to*" permet de dire qu'un tag est associé à une ressource. Cette propriété générique peut-être étendue pour exprimer la fonction de tagging voulue. Par exemple, dans le vocabulaire est proposé une extension de cette propriété "*makes me feel*" qui permet d'indiquer que le tag utilisé dénote une émotion provoquée par la ressource taggée. Ce mécanisme d'extension permet plus de réutilisabilité du vocabulaire. Celui-ci est plus flexible, adaptable selon les contextes d'utilisation. Par contre, cela constitue un frein à l'interopérabilité puisque chaque application redéfinit ses propres concepts.

L'objectif du projet Argument Markup Language (AML) est de représenter les données produites par des outils dits d'argumentation et de permettre l'interopérabilité dans ce contexte. AML est une DTD pour XML. La DTD définit un ensemble de concepts, de tags XML qui permettent de catégoriser des éléments textuels. La classification est proche de celle de Toulmin mais définit de nombreux concepts supplémentaires de façon empirique. Par exemple les éléments du discours peuvent être identifiés comme étant des justifications, des évidences, des concessions, des suppositions ... [Delannoy, 1999] résume l'ensemble des classes possibles. Le vocabulaire définit aussi des concepts permettant de modéliser le domaine abordé dans la discussion et permet de modéliser les résultats d'une analyse de l'argumentation qui permettrait de mettre en évidence des contradictions dans le discours ou de dire si une argumentation semble valide ou non. Le cadre applicatif d'AML proposé à l'origine par l'auteur est celui de l'analyse d'un texte pour en extraire une représentation claire de la structure argumentative. L'outil va segmenter et annoter le texte automatiquement en s'appuyant sur des méthodes de traitement automatique de la langue. Puis, l'utilisateur peut enrichir cette représentation, la corriger, la compléter, si nécessaire. AML couvre donc de nombreux aspects de l'argumentation que ce soit le typage des actes de langage, l'analyse logique de l'argumentation ou la prise de positions déclarées par les utilisateurs mais il ne définit aucunes possibilités d'extension du modèle ou de couplage de celui-ci avec d'autres.

A l'inverse, le projet AIF pour Argumentation Interchange Format, s'appuie sur RDF pour fournir un langage extensible de représentation des arguments. Le cœur du vocabulaire AIF est très simple, il est constitué de deux types de concepts : les noeuds informatifs et les noeuds de modèle. Les noeuds informatifs permettent de représenter tous les éléments du discours. On pourrait par exemple redéfinir toutes les catégories de AML ou les éléments du modèle de Toulmin (ce qui est d'ailleurs fait dans [Rahwan et Sakeer, 2006]). Les noeuds de modèles sont là pour représenter des patterns de raisonnement comme des règles d'inférences de logiques déductives. Le projet AIF s'appuie sur la technologie RDF,

il ne cherche pas à proposer un modèle exhaustif, complet, mais plutôt un modèle suffisamment générique pour pouvoir être adapté, étendu selon les applications. Cependant, AIF semble prévu pour créer des cartes d'argumentation conçues par des "experts" et non pour représenter les données de discussions en ligne.

Hormis les applications développées autour de Nicetag, nous n'avons pas trouvé de travaux sur la représentation des données de l'argumentation qui permettent de représenter la structure argumentationnelle choisie par les acteurs de l'argumentation. Bien que AML soit utilisé dans une application ou les utilisateurs peuvent annoter eux-mêmes, les données ici annotées ne sont pas leurs productions dans un contexte de débat. En règle générale, la représentation d'une argumentation est un travail de retraitement des données que celui-ci soit manuel ou automatique. Il n'y a pas pour l'instant, à notre connaissance, d'outil d'expression d'opinion qui adopte un format de représentation des données argumentatives.

On observe deux tendances dans la définition de formats de données, la définition d'un vocabulaire exhaustif et complet ou la définition d'un vocabulaire simple et générique permettant d'être étendue. Le deuxième solution présente l'avantage d'être plus flexible et donc d'être adaptable à plusieurs contextes applicatifs mais repose la question de l'interopérabilité. Si chaque application se base sur sa propre extension d'un vocabulaire générique commun, les seules données pouvant être échangées entre ces applications sont celles que le modèle générique permet d'exprimer sans extension. Il faut donc trouver un compromis entre générativité et exhaustivité.

Les approches de représentation de l'argumentation proposent toutes de représenter la structure argumentative et essentiellement en vue d'aider à décider sur l'argumentation la mieux construite, la mieux défendue. On retrouve toujours cette notion d'argument et d'attaque entre arguments. Cependant, à l'exception d'AML, aucun vocabulaire ne s'intéresse à l'acceptabilité des arguments auprès des acteurs concernés. Au mieux, on peut savoir qui a créé un argument. Mais seul AML souligne l'importance de la connaissance des coalitions ou des acteurs soutenant un argument. Dans un système démocratique, l'importance du nombre d'acteurs soutenant ou s'opposant à un argument est une mesure importante de l'acceptabilité d'un argument.

3.2.5 Structuration des débats via les n-opposition

Dans notre travail, nous proposons de structurer les débats en fonction des jugements émis par les utilisateurs. Pour ce faire, il est nécessaire de définir des relations entre les jugements, des relations d'implication et d'opposition.

La définition de ces règles nous amène à définir une structure entre les jugements. Ces structures peuvent alors être représentées sous une forme géométrique pour en faciliter la lecture.

Cette forme de structuration et de représentation des structures se base sur la théorie des n-opposition [Moretti, 2005] que nous présentons succinctement ici.

La théorie des n-opposition est une généralisation géométrique de la théorie de l'opposition d'Aristote. En règle générale, la structure présentée pour introduire la théorie des n-opposition est le carré logique des oppositions d'Aristote.

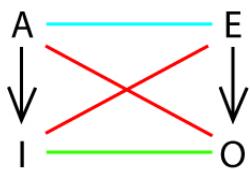


FIGURE 3.2 : Le carré des oppositions d'Aristote

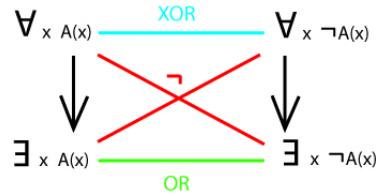


FIGURE 3.3 : Le carré des oppositions d'Aristote pour la logique des prédictats

3.2 présente 4 types de relations (représentées par des flèches ou des lignes) entre 4 jugements de valeurs traditionnellement représentés par des lettres A, E, I, O. A et E représentent des jugements universels (A est l'affirmatif universel, E est le négatif universel). I et O représentent les jugements du niveau pratique (I est le jugement pratique affirmatif, O est le jugement pratique négatif). Les flèches entre les deux niveaux représentent des relations de sub-alternation. La ligne supérieure, entre A et E représente une relation de contrariété. La ligne inférieure entre I et O représente une relation de sous-contrariété. Les diagonales représentent des contradictions. La figure 3.3 est un exemple de carré des oppositions en logique des prédictats.

Les définitions suivantes sont tirées de [Luzeaux *et al.*, 2007].

- La **sub-alternation** est définie comme l'impossibilité d'avoir le premier terme vrai sans avoir le second vrai également. Cela correspond à l'implication logique.
- La **contrariété** entre deux termes est l'impossibilité que les deux termes soient vrais simultanément et la possibilité que ces deux termes soient faux simultanément.
- La **sub-contrariété** entre deux termes est l'impossibilité que ces deux termes soient faux simultanément mais la possibilité qu'ils soient vrais simultanément.
- La **contradiction** entre deux termes est l'impossibilité pour ces deux termes d'être soit vrais simultanément, soit faux simultanément.

Ce type de représentation permet de présenter les relations définies entre les jugements que nous proposons aux utilisateurs. Ceux-ci n'ont pas nécessairement besoin de connaître la structure sous-jacente et beaucoup de ces relations sont traduites par des impossibilités dans les formulaires d'annotation de choisir simultanément deux jugements par exemple.

Ces relations permettent dans un contexte multi-agents de déterminer des conflits entre les agents si par exemple un agent est d'accord sur un message et un autre agent n'est pas d'accord sur ce message.

Enfin, en guise de perspective, soulignons que D.Luzeaux [Luzeaux *et al.*, 2007] a démontré la qu'il était possible de traduire les structures d'oppositions en logiques épistémiques.

Nous présentons dans la partie expérimentation les différentes structures adoptées au cours de nos expérimentations. Aujourd'hui, nous travaillons à l'implémentation dans notre système de la possibilité de définir ces structures par le biais de l'interface utilisateur.

3.2.6 Synthèse des fonctionnalités pour les débats 2.0

Nous allons dans cette section revenir sur les fonctionnalités présentées dans cette partie. Il s'agit de souligner les fonctionnalités existantes et éparses qu'il nous semble judicieux d'intégrer dans un outil de débats 2.0. Nous avons présenté dans cette partie les fonctionnalités suivantes :

1. La discussion. Cette fonctionnalité est primordiale, il s'agit de permettre aux utilisateurs de formuler des éléments de discours, des arguments.
2. L'expression d'opinions ou la structuration de la discussion par le biais de jugements de valeurs. Nous défendons que au delà de la simple structuration des discussions sous la forme de "réponses à", il est nécessaire dans un débat de pouvoir dire que un élément de discussion s'oppose à un autre ou en supporte un autre ou plus simplement de pouvoir exprimer son accord ou son désaccord sur un élément.
3. Le partage de connaissances. Les outils de partage de connaissances apportent la possibilité de considérer certains éléments comme étant des références. Dans notre approche, il nous semble primordial de pouvoir ramener dans la discussion des éléments produits en dehors des débats. Par exemple, un rapport scientifique, une directive ministérielle ...
4. Le partage Multimédia ne fait qu'étendre la fonctionnalité suscité à des contenus multimédia.
5. Le tagging est une fonctionnalité ayant fait ses preuves permettant à chaque utilisateur d'organiser les éléments de discussion comme il l'entend.
6. Le suivi des modifications. Avoir une vision d'ensemble de la discussion à un moment donné est essentiel pour pouvoir prendre une décision. Avoir une vision de l'évolution de la discussion à travers le temps est également une source d'indices judicieux. Ceci permet notamment de mesurer quel individu ou quel élément de discours a pu faire changer l'opinion. Il est donc nécessaire d'avoir une sorte de suivi des versions du débat permettant de voir les modifications, l'évolution de la discussion.
7. La structuration du réseau social. Les relations entre individus dans le débat sont également un élément de structuration. Certaines parties prenantes sont identifiées dès le début du débat, d'autres vont se constituer ou se défaire au fur et à mesure des discussions.

8. Fonder une identité numérique du débattant. Il ne s'agit pas ici de caractériser un profil ou de faire le lien entre un pseudonyme d'utilisateur et une personne. Il s'agit de garder la trace des actions d'un utilisateur, lui permettre de conserver un historique et construire sa représentation dans l'application. Pour les autres, cette identité sera une référence pour savoir à qui s'adresser et suivre le cours des échanges.
9. L'existence de formats d'échanges des données. Enfin, une application de débats 2.0 devant être utilisée dans le cadre de problèmes concrets mettant en œuvre l'usage d'autres applications, il est nécessaire de fournir des formats de données permettant de représenter l'ensemble des données du débat pour une exploitation de celles-ci dans d'autres applications.

3.3 L'annotation

Cette partie est dédiée à l'annotation. L'annotation est une nouvelle forme de communication écrite sur le Web. Pourtant, elle se pratique dans de nombreux contextes depuis des siècles. Nous allons présenter cette pratique tout d'abord dans son contexte historique puis nous présenterons les différentes formes d'annotation aujourd'hui sur le Web. En particulier, nous détaillerons les deux formes d'usage de l'annotation que sont l'annotation dédiée à la discussion, l'annotation discursive et l'annotation dédiée à l'explicitation du sens des données, l'annotation sémantique.

3.3.1 Histoire

Cette section présente les pratiques classiques de l'annotation et l'apport de cette forme d'écriture à l'herméneutique, activité d'interprétation des textes [Lalande, 1900].

L'annotation consiste, dans son approche la plus classique, à associer un commentaire ou autre chose (une image, une référence ...) à un document ou une partie d'un document. Les notes de lecture, commentaires laissés dans la marge d'un livre sont des annotations.

Dans de nombreux contextes professionnels, on retrouve une pratique d'échange de documents et d'enrichissement de ces documents par l'ajout de nouveaux documents ou de notes critiques. Un dossier médical est une suite de documents lus et enrichis par les membres de l'équipe [Thorne, 2000]. On retrouve cette pratique dans tous les domaines où des documents sont coproduits par plusieurs acteurs.

L'annotation, appelée glose, était pratiquée au Moyen-âge. Les manuscrits religieux étaient commentés et distribués avec leurs commentaires à travers l'Europe. Chaque commentaire précisant comment un passage devait être interprété. L'annotation permet la construction collective du sens d'un document et la construction collective d'opinions. Dans le cadre d'une co-construction, l'annotation permet de conserver les échanges qui ont conduit à l'état actuel de la co-construction. Les annotations constituent dans ce cas un ensemble de données permettant de réinterpréter le texte dans un contexte et c'est

ainsi que l'annotation s'avère centrale pour l'herméneutique.

Le W3C crée en 2001 le protocole Annotea²², protocole d'échange et de stockage d'annotations Web au format RDF. La page Web d>Annotea recense plusieurs projets et permet de récupérer des implémentations libres de clients et de serveurs d'annotation. De nombreux outils d'annotation de segments de textes existent déjà sur le Web. Le projet Annotea a donné naissance à plusieurs prototypes comme le plugin Annozilla pour le navigateur Firefox. De son côté, le projet Marginalia-geof²³ fournit un ensemble de bibliothèques permettant d'ajouter la fonctionnalité d'annotation à une application Web. La société Sopinspace propose l'outil Co-ment²⁴ dont un démonstrateur en ligne permet d'annoter un document d'exemple. LogiLogi²⁵ invite à discuter philosophie par le biais d'annotations. Enfin, l'application Diigo²⁶ permet de partager des bookmarks et d'annoter des fragments de pages Web.

3.3.2 Les différentes formes d'annotation sur le Web

En première approche, l'annotation peut être comparée à la pratique de la lecture d'un article avec un stabilo et un stylo. L'annotation revient à surligner un ou plusieurs passages dans un article au stabilo et associer un commentaire à ces passages en écrivant au stylo dans la marge.

L'annotation a été classiquement utilisée de deux façons sur Internet. La première est qualifiée de sémantique et fait référence à l'utilisation des technologies du Web Séman-tique. Il s'agit de rendre explicite des informations concernant un document ou un passage de document. Ces informations, appelées métadonnées, sont ajoutées par le biais d'annotations sémantiques dans le but de rendre explicite pour un programme des informations concernant le contexte de production ou le contenu du document. Par exemple, les annotations sémantiques peuvent préciser qui est l'auteur du document, à quelle date il a été produit et les thèmes qu'il traite. L'exemple 3.5 est tiré du site de DublinCore²⁷. L'annotation sémantique est souvent utilisée afin d'améliorer l'indexation de documents.

L'autre forme d'annotation est l'annotation discursive. Cette pratique s'apparente énormément à celle des Forums ou Blogs. Les annotations sont constituées de deux éléments : la référence à une ou des partie(s) surlignée(s) et un commentaire. L'annotation discursive est utilisée comme outil de discussion dans un contexte de travail collectif.

22. <http://www.w3.org/2001/Annotea/>

23. <http://www.geof.net/code/annotation>

24. <http://www.sopinspace.com/?p=70>

25. <http://en.logilogi.org>

26. <http://www.diigo.com/>

27. <http://dublincore.org/documents/usageguide>

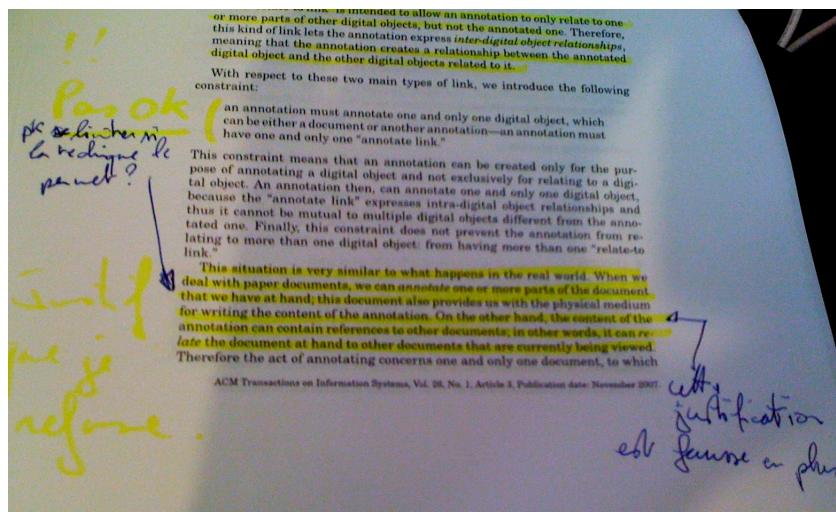


FIGURE 3.4 : Article surligné au Stabilo et commentaire dans la marge

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <rdf:Description rdf:about="http://media.example.com/audio/guide.ra">
    <dc:creator>Rose Bush</dc:creator>
    <dc:title>A Guide to Growing Roses</dc:title>
    <dc:description>Describes process for planting and nurturing different kinds of rose bushes.</dc:description>
    <dc:date>2001-01-20</dc:date>
  </rdf:Description>
</rdf:RDF>

```

FIGURE 3.5 : Exemple d'annotation sémantique utilisant le vocabulaire Dublin Core

L'annotation sémantique

A partir de l'instant où l'on peut associer toute nouvelle information à tout ou partie d'une ressource déjà existante, alors cette nouvelle information est une annotation. Au titre de cette définition, nous pouvons aussi ajouter la pratique du tagging [Wu *et al.*, 2006] comme étant une forme d'annotation. Mais le tagging n'est pas a priori une forme de discussion. L'objectif du tagging est bien de fournir une nouvelle information pour l'indentation des documents. Le tagging permet une catégorisation par les utilisateurs, créant ce que l'on appelle une folksonomie ou taxonomie créée par les utilisateurs (folks-taxonomie) [Vander Wal, 2007].

Avec le tagging, on entre dans un usage de l'annotation visant à expliciter le sens de la donnée taggée, et c'est exactement l'objectif du Web Sémantique. L'annotation sémantique consiste en l'association de métadonnées (données sur les données) à des docu-

Nayla Farouki: Quelques réflexions sur l'homme augmenté...

4 mars 2010
Par nanosciences

On pourrait déjà dans un premier temps se demander ce que signifie un être humain augmenté. En effet, l'outil - quel qu'il soit - augmente forcément la personne qui l'utilise en ce sens qu'il lui donne des moyens qu'elle n'aurait pas sans lui. Quels moyens? Prenons quelques exemples :

- a - suis-je augmenté(e) grâce à mes lunettes sans lesquelles je ne vois rien?
 - b-suis-je augmenté(e) grâce à mon vélo, ma voiture, mon bateau, mon avion qui me permettent de me déplacer d'une manière par ailleurs impossible?
 - c - Suis-je augmenté(e) en raison de la prothèse qui vient remplacer mon col du fémur cassé?
 - d - ma mémoire est-elle augmentée grâce à mon carnet de notes?
- d-plus hasardeux : mon intelligence est-elle augmentée du fait que je peux comprendre des mathématiques que Thalès ou Pythagore n'avaient jamais imaginées? ou encore - ma liberté est-elle augmentée du fait que j'ai de loin plus de choix dans les Grands Magasins qu'au temps de Zola?

Auteur: Nayla Farouki



Je m'interroge
Il semble que deux choses soient présupposées ici :
1- la liberté de l'homme est augmentée par plus de choix dans les Grands Magasins...
2- un homme dont la liberté est augmentée est un homme augmenté

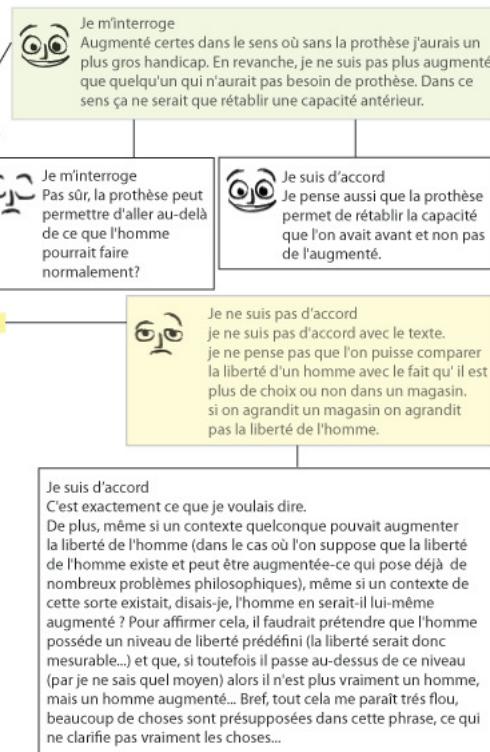


FIGURE 3.6 : Discussion à base d'annotations produites au cours de l'une de nos expérimentations

ments. Les métadonnées sont généralement ancrées dans un document par le biais de balises. Ce type d'annotation repose sur l'utilisation de langages de description interprétables par des agents logiciels et suffisamment verbeux pour pouvoir être lus par des humains. L'objectif est de faciliter la compréhension de la nature et du sens d'une ressource pour des agents logiciels ou humains.

L'un des usages des annotations sémantiques est de s'en servir pour améliorer l'indexation des documents. Les producteurs d'annotations sémantiques sont alors souvent des professionnels de la représentation de connaissances qui utilisent des vocabulaires, aussi appelés ontologies [Gómez-Pérez et Corcho, 2002], portés par les technologies du W3C et souvent définis par des experts. Elles servent notamment à l'amélioration de la recherche documentaire dans des bibliothèques numériques [Agosti et Ferro, 2007]. Mais la tendance actuelle, comme nous le verrons plus loin, est de faire produire ces annotations de plus en plus par les utilisateurs d'applications Web 2.0.

L'annotation discursive

L'annotation discursive est un modèle de discussion au même titre que ceux des forums ou blogs. Il repose sur l'échange de messages entre utilisateurs autour d'une ressource Web. L'utilisateur sélectionne un passage dans une ressource et lui associe un message qui devient une nouvelle ressource. Ainsi, il est possible de créer des discussions par des chaînes de messages.

La plupart des pratiques discursives du Web 2.0 entrent dans le cadre de la définition de [Veronis, 2000] : est annotation tout apport d'information aux données brutes originales.

Ainsi, les commentaires sur un blog sont des annotations de l'intégralité de l'article auquel ils sont rattachés. Chaque nouveau post sur un forum est une annotation de la discussion en cours. Chaque contribution dans un wiki est localisée dans un document, et donc est une annotation du document. Pour aller plus loin, chaque élément de discussion dans une conversation orale est une nouvelle annotation de la discussion.

La thèse de Gaelle Lortal [Lortal, 2006] présente l'usage d'un outil d'annotation, l'outil Annozilla, dans le contexte du travail collectif. Lortal définit l'annotation comme une note critique ou explicative qui accompagne un texte, c'est une note scripturale qui est ancrée dans un document. Pour Lortal, l'annotation discursive permet de considérer un texte inclus dans son contexte, c'est à dire en le liant à ses conditions de production et de réception. Les annotations dans le cadre expérimental de Lortal, le travail collectif, sont l'ensemble des remarques échangées entre les participants d'un même projet autour des documents produits au cours du projet. Les annotations peuvent par exemple être de nouveaux éléments à introduire dans une prochaine version du texte sur lequel elles portent.

Vers un modèle unifié

Nous proposons dans notre travail d'unifier annotations sémantiques et annotations discursives en annotations que nous qualifions de socio-sémantiques. Agosti et Ferro [Agosti et Ferro, 2007] proposent un modèle formel unifiant ces deux approches. Nous allons ici développer le modèle d'Agosti et Ferro pour pouvoir souligner dans notre modèle ce que nous apportons de différent. Chez Agosti et Ferro, l'annotation sémantique est appelée "méta-donnée" et l'annotation discursive fait référence à un ajout de contenu (*annotation as meta-data* et *annotation as content*). [Marshall, 1998] parle d'annotations formelles quand il s'agit de meta-données et d'annotations informelles pour "*ces notes écrites dans la marge du style de ce que l'on s'écrit à soi-même quand on lit un article*". Ils soulignent dans leur modèle que pour modéliser l'annotation il faut préciser la nature des ressources annotées (l'annotation est multimédia, des sites comme nico nico douga²⁸ permettent d'annoter de la vidéo par exemple) et la nature du contenu de l'annotation (il est possible d'annoter avec du texte, des icônes, des images ...) et si l'annotation est fortement structurée ou non. L'approche classique des outils de discussion est de peu structu-

28. <http://www.nicovideo.jp/>

rer les commentaires (qui se font par le biais de champs de texte à saisie libre) alors que au contraire, les métadonnées de l'annotation sémantique sont fortement structurées.

La définition du modèle d'Agosti et Ferro commence par la définition des objets numériques de l'ensemble considéré (les documents, les annotations et d'autres objets numériques). Pour Agosti et Ferro, les annotations et les documents sont des objets numériques mais des annotations ne sont pas des documents. Les deux ensembles sont bien distincts. L'ensemble des objets numériques est défini comme suit :

- D est un ensemble de Documents et $d \in D$ est un document générique. U_D est l'ensemble universel des documents et donc $D \subseteq U_D$
- A est un ensemble d'annotations et $a \in A$ est une annotation générique. U_A est l'ensemble universel des annotations et donc $A \subseteq U_A$
- $DO = D \cup A$ est un ensemble d'objets numériques et $do \in DO$ est soit une annotation soit un document. $U_{DO} = U_D \cup U_A$ est l'ensemble universel des objets numériques et donc $DO \subseteq U_{DO}$

Agosti et Ferro soulignent ensuite l'importance de la dimension temporelle dans l'annotation : il y a, pour eux, nécessairement une relation d'ordre temporelle entre les annotations et les documents. Une annotation ne peut pas être antérieure à un objet numérique qu'elle annote. Les ensembles Universels précédemment cités, U_D, U_A, U_{DO} sont les mêmes en tout instant puisqu'ils sont universels. Par contre, les ensembles D, A, DO changent avec le temps. Les deux opérations de modification des ensembles sont la création et la suppression d'un élément. La création d'un nouvel objet numérique se décrit de la façon suivante :

1. Prenons un ensemble d'objets numériques à un instant k : $DO(k)$
2. Créer un nouvel objet numérique revient à sélectionner un objet dans l'ensemble universel qui n'est pas aussi dans l'ensemble $DO(k)$: $do \in \overline{DO(k)} \subseteq U_{DO}$
3. On obtient un nouvel ensemble à l'instant $k+1$ contenant l'ensemble à l'instant k auquel est ajouté le nouvel objet numérique : $DO(k+1) = (DO(k) \cup (do)) \in 2^{U_{DO}}$

La suppression d'un élément se décrit de la façon suivante :

1. Prenons un ensemble d'objets numériques à un instant k : $DO(k)$
2. Choisir un objet numérique existant dans l'ensemble $DO(k)$: $do \in DO(k)$
3. On obtient un nouvel ensemble à l'instant $k+1$ contenant l'ensemble à l'instant k auquel on a enlevé l'objet numérique choisi : $DO(k+1) = DO(k) \setminus (do)$

Chaque objet numérique doit être identifié de façon unique. Cette problématique n'est pas nouvelle et les auteurs font un rapide état de l'art des différents mécanismes proposés pour fournir des identifiants uniques. De notre côté, nous adopterons le système d'URI [Berners-Lee, 1994]. Quoi qu'il en soit, l'ensemble des identifiants nécessaires se définit de la façon suivante :

- I est un ensemble d'identifiants tel que $|I| = |DO|$ et $i \in I$ est un identifiant. U_I est l'ensemble universel des identifiants tel que $|U_I| = |U_{DO}|$, $I \subseteq U_I$. Une fonction bijective h est définie de U_I dans U_{DO} , $h : U_I \rightarrow U_{DO}$. Cette fonction h permet d'assos-

cier un identifiant unique à un objet numérique : $\forall do \in U_{DO}, \exists ! i \in U_I \mid h(i) = do \Rightarrow h^{-1}(do) = i$.

Les auteurs définissent ensuite des ensembles permettant de représenter des notions de droits d'accès et de droits d'auteurs. Nous ne détaillerons pas ici cette partie du modèle d'Agosti et Ferro qui traite des utilisateurs, de groupes d'utilisateurs et de niveaux de droits (privé, public, partagé, interdit, lecture uniquement et lecture écriture). Cette partie du modèle n'est pas nécessaire pour notre travail. Nous soulignerons juste que dans leur modèle, Agosti et Ferro ont prévu qu'une annotation puisse avoir plusieurs auteurs. Ceci définit donc un ensemble d'auteurs qui est un sous-ensemble des utilisateurs du système. Si nous partageons le fait qu'une annotation puisse être produite par plusieurs auteurs, nous défendons aussi le fait qu'une annotation puisse être produite par des agents logiciels qui ne sont pas des utilisateurs du système au sens où l'entendent Agosti et Ferro.

Il faut maintenant développer un système permettant de lier des annotations à des objets numériques et de les ancrer dans ces objets numériques (localiser le passage précis d'un document sur lequel porte l'annotation par exemple). Le mécanisme proposé par Agosti et Ferro repose sur l'utilisation d'identifiants uniques. Ils envisagent deux sortes de liens entre annotations et objets numériques. Les liens d'annotation pour dire que l'annotation porte sur cet objet numérique et le lien de référence, si l'annotation fait référence à un autre objet numérique que celui annoté. Là encore, nous considérons que le lien d'annotation est primordial mais par contre, en ce qui concerne la définition d'un autre type de lien, nous pensons que toute autre relation liant une annotation à un objet numérique est une extension de la relation d'annotation. Ce n'est que l'association d'une nouvelle sémantique à un mécanisme qui est toujours le même. Tout ce que traduit la deuxième relation proposée par Agosti et Ferro est qu'il est possible de faire des liens hypertextes dans le contenu d'une annotation pour renvoyer vers un autre objet numérique.

De plus, le parti pris de définir ces deux types de relation contraint les auteurs à une restriction sur l'usage des annotations qui ne nous semble pas légitime technologiquement et qui limite les possibilités d'usage de l'annotation. Les auteurs contraignent une annotation à ne porter que sur un seul objet numérique. Autrement dit, il ne doit y avoir qu'un seul lien d'annotation par annotation. Si l'objectif de l'annotation est de créer des liens sémantiques entre des ressources, alors cette contrainte nous semble fortement préjudiciable. Les auteurs justifient cette contrainte en soulignant que, après tout, cela revient exactement au même que ce que nous faisons lorsque nous annotons un document papier. Nous ne pouvons pas ajouter une annotation dans la marge d'un document qui porterait sur deux articles différents. Là encore, cette justification n'a pas lieu d'être. Une personne peut tout à fait souligner plusieurs passages dans un même document et les annoter d'un même commentaire dans la marge. De plus, les nouveaux usages en ligne ne doivent pas être conçus comme des reproductions de pratiques déjà existantes mais comme des enrichissements de ces pratiques, voir comme de nouvelles pratiques.

- L'ensemble des types de liens est appelé LT. Un élément $lt \in LT$ correspond à l'un des deux types de liens possibles.

Agosti et Ferro définissent la notion de Stream comme étant une séquence ordonnée de symboles représentant le contenu d'un objet numérique ou d'une partie de l'objet numérique. Ce concept est central pour permettre de définir l'ancre des annotations dans les objets annotés.

- Un stream sm est une séquence finie $sm : S = \{1, 2, \dots, n\} \rightarrow \Sigma, n \in \mathbb{N}$, où Σ est un alphabet de symboles. Il existe un stream vide : $esm = \emptyset$. SM est un ensemble de streams et $sm \in SM$ est un stream. U_{SM} est l'ensemble universel des streams tel que $SM \subseteq U_{SM}$.

Grâce à cette définition, un objet numérique est vu comme étant constitué d'une séquence d'éléments. Il devient possible de découper cette séquence en segment. Un segment est identifié par son symbole de début et son symbole de fin.

- Soit un stream $sm : S = \{1, 2, \dots, n\} \rightarrow \Sigma, n \in \mathbb{N}, sm \in SM$, un segment est une paire :
- $$st_{sm} = (a, b) \mid 1 \leq a \leq b \leq n, a, b \in \mathbb{N}$$

Un segment est une restriction du stream à un intervalle $[a, b]$. Cette définition s'adapte parfaitement dans le cadre de l'annotation textuelle. Chaque caractère a un indice qui lui est associé représentant sa position dans la suite de caractères du texte. Une sélection ou segment dans le texte peut donc être ancrée en utilisant l'indice du premier caractère sélectionné et l'indice du dernier caractère sélectionné. Cette définition s'adapte bien à des documents indexés sur le temps. Par exemple, une vidéo ou un son ont une durée. L'annotation d'un document indexé sur le temps se fait en utilisant le code temporel du début et de la fin du passage sélectionné. Dans le cadre de documents XML ou HTML, des mécanismes comme XPath et XPointer permettent de représenter ces sélections par le biais d'indices de début et de fin de la sélection. Nous adoptons la même solution que Agosti et Ferro concernant des annotations portant sur des documents HTML. Cependant, qu'en est-il d'annotations portant sur des sélections dans des images ou des zones géographiques ? Le mécanisme de streams et de segments n'est pas adapté dans ces cas là. Enfin, le mécanisme de streams et de segments est valable dans le cadre de documents figés, qui n'évoluent pas dans le temps. Hors, les documents Web aujourd'hui sont amenés à changer.

Pour résumer, chez Agosti et Ferro, une annotation est liée à un document numérique (le document annoté) en utilisant l'identifiant unique de ce document et en utilisant le mécanisme de segment pour ancrer l'annotation dans le document.

Dans la suite de leur modèle, les auteurs parlent de signe de l'annotation pour dénoter le contenu de l'annotation. Ils listent plusieurs types de signes possibles : les signes textuels, les signes graphiques, les signes vidéos et les signes sonores. Ceux-ci pouvant être combinés pour créer des annotations multimédias. Quand on parle de signe, on considère la forme de l'annotation, c'est à dire la nature de son contenu (alpha-numérique, image, son ...), sa présentation éventuellement mais aucunement la sémantique de celle-ci. Pour Agosti et Ferro, la sémantique d'une annotation est en fait constituée des sémantiques associées aux différents signes composant le contenu de l'annotation. Une annotation est donc composée de signes qui ont eux-mêmes chacun au moins une sémantique associée.

Dans notre approche, nous ne parlons pas de signe mais de ressource. Une annotation chez nous est composée de plusieurs ressources (du texte, un tag ...) ayant chacune leur sémantique. La sémantique de l'annotation est alors une composition des sémantiques des différentes ressources constituant l'annotation.

En conclusion de cette partie, nous allons présenter la définition générale de l'annotation d'Agosti et Ferro. Une annotation est l'expression d'un ou plusieurs signes. Chaque signe a au moins une sémantique qui lui est associée. Une annotation annote un et un seul document numérique identifié par un identifiant unique. L'annotation est ancrée dans le document numérique par le biais du mécanisme de segments et de streams.

- Une annotation $a \in A(k)$ est un tuple : $a = (i_a I(k), au_a \in USR(k-1), G_a \in 2^{GR(k-1)} x P, sp_a \in SP, A_a \subseteq SN(k) x LTxST(k) x SM(k-1) x I(k-1))$
- i_a est l'identifiant unique de l'annotation
- au_a est l'auteur de l'annotation
- G_a est l'ensemble des groupes d'utilisateurs avec leurs droits respectifs d'accès pour l'annotation a
- sp_a est la portée de l'annotation (privé, public, partagé)
- Chaque n-uplet de A_a signifie que l'annotation a par le biais d'un signe dans $SN(k)$ et un type de lien dans LT annote ou fait référence à un segment dans $ST(k)$ d'un stream dans $SM(k-1)$ d'un objet numérique identifié par son identifiant dans $I(k-1)$.

La présentation du modèle d'Agosti et Ferro est l'occasion de présenter quelques concepts de base de l'annotation : l'identification des ressources, l'ancrage par le biais de segments dans des streams, la distinction entre signe et sémantique. Nous présenterons plus loin notre modèle en discutant les limites de celui d'Agosti et Ferro. Nous considérons que ce modèle présente les limites suivantes :

- Les annotations ne sont pas considérées comme des documents mais comme des objets numériques. Hors, chez Agosti et Ferro, seuls les documents sont annotables... Donc les annotations sont des objets numériques non annotables.
- Une annotation porte nécessairement sur un document pré-existant. Il n'est donc pas possible de lier une annotation déjà existante à un document plus récent que l'annotation elle-même.
- Une annotation a nécessairement un auteur humain. Les annotations produites par des programmes ne sont pas considérées dans ce modèle.
- Une annotation ne peut porter que sur un seul document.
- Le concept de stream fondamental pour permettre l'annotation de documents textuels ou de documents s'apparentant à des flux (vidéo, son) est ici l'unique solution proposée. Malheureusement, celui-ci ne permet pas d'imaginer des annotations portant sur des documents autres que textuels, vidéo ou sonores (par exemple les images ou les fonds de cartes géographiques) et si la sélection a une forme "complexe" (si par exemple on veut annoter une zone circulaire dans une image).

- Enfin, le modèle ne considère pas l'évolution temporelle des documents. Que deviennent les annotations portant sur un document qui vient d'être modifié ?

Outils d'annotation

Nous allons dans cette partie présenter le protocole Annotea et des applications d'annotation en ligne. Le but est de présenter l'état de l'art des outils d'annotation.

Annotea

Le projet Annotea est porté par le W3C. C'est un "*LEAD project*". Le statut de "*LEAD project*" signifie que le W3C a voulu via Annotea prouver la faisabilité de l'annotation de documents Web. Le projet date de 1999-2000 et les dernières mises à jour sur le site officiel du projet datent de 2006. Pour autant, la proposition principale du projet n'est pas dépassée technologiquement. L'objectif du projet est d'améliorer le travail collaboratif par le biais du partage de métadonnées et de bookmarks (le terme de bookmark renvoie à la notion de marque pages dans un livre, sur le Web un bookmark est l'enregistrement d'une ressource comme étant utile, intéressante, un favori ...). Les annotations sont présentées comme étant des commentaires, des notes, des explications, ou d'autres remarques qui peuvent être attachées à n'importe quel document Web ou partie de document Web.

Annotea définit un protocole de communication client-serveur basé sur HTTP permettant de récupérer des annotations existantes, de les éditer, en créer de nouvelles et les lier entre elles. La page présentant ce protocole est en statut de brouillon (Draft) et n'a pas été éditée depuis 2002²⁹. La création d'une annotation se fait par le biais de requêtes POST HTTP. La récupération des annotations se fait par le biais de requêtes GET. La suppression par une requête DELETE et la mise à jour par une requête PUT. Le principe est simple et efficace, il suffit de s'appuyer sur le protocole HTTP.

L'autre apport du projet Annotea est la définition d'un vocabulaire RDF minimal permettant de représenter des annotations³⁰. Ce vocabulaire ne définit qu'une seule classe, la classe *Annotation*. Et lui associe un ensemble de propriétés :

- **annotates** : lie l'annotation à la ressource qu'elle annote.
- **author** : lie l'annotation à son auteur.
- **body** : lie l'annotation à son contenu.
- **context** : lie l'annotation à la partie précise de la ressource annotée.
- **created** : lie l'annotation à sa date de création.
- **modified** : lie l'annotation à une date de modification.
- **related** : lie l'annotation à une autre ressource que celle annotée.

Par ailleurs, le vocabulaire parle d'une propriété "*hasAnnotation*" comme étant l'inverse de la propriété "*annotates*". Mais celle-ci n'est pas clairement définie (pas de la façon conven-

29. <http://www.w3.org/2001/Annotea/User/Protocol.html>

30. <http://www.w3.org/2000/10/annotation-ns>

tionnelle prévue dans RDFS). Il faut aussi noter que la définition de la propriété "annotes" semble indiquer qu'une annotation est liée à une et une seule ressource. Pourtant, aucune contrainte n'est posée dans le modèle pour restreindre les annotations à ne porter que sur une seule ressource à la fois. La propriété "body" est une spécialisation de la propriété "related" ce qui, au passage, sous-entend que le corps d'une annotation est une ressource. Enfin, les propriétés "author", "created" et "modified" sont des spécialisations de propriétés définies dans le vocabulaire Dublin Core. La figure 3.7 résume les apports essentiels de ce vocabulaire. Nous proposons au passage de renommer certains noms du vocabulaire pour être conforme aux bonnes pratiques actuelles de la définition de vocabulaires RDF (utilisation de verbes pour les propriétés donc *hasBody* et *hasContext*).

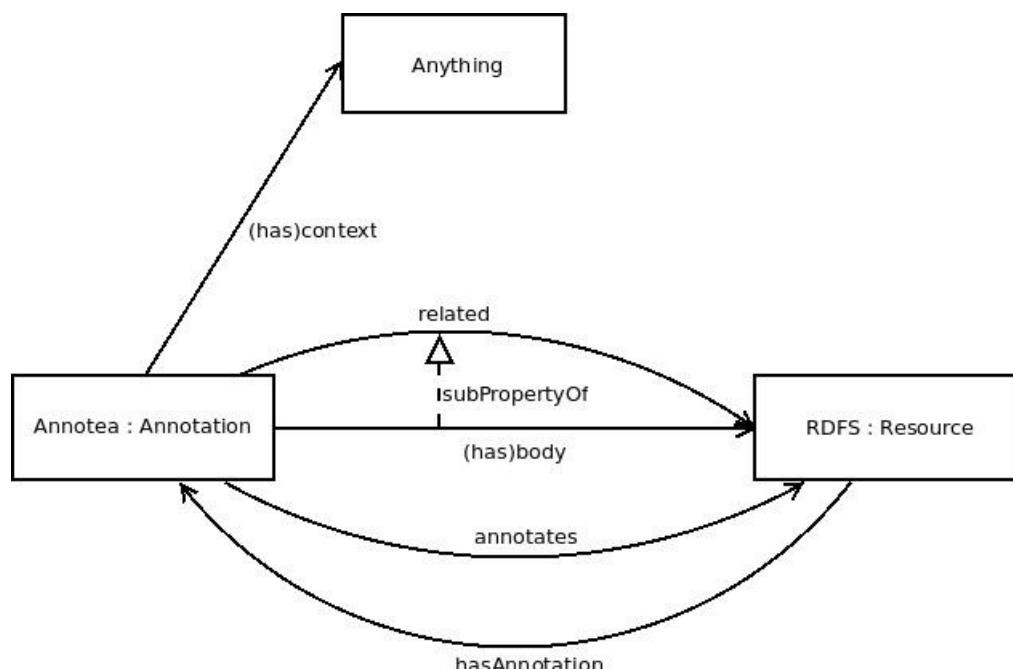


FIGURE 3.7 : Classes et propriétés du vocabulaire Annotea pour l'annotation

L'une des forces d'Annotea est de rester minimaliste. La représentation du contenu des annotations n'est pas définie dans le modèle. Étant donné que RDF est utilisé pour définir ce modèle, et que le contenu d'une annotation (ce qui est pointé par la propriété *body*) est une *Ressource* RDFS (l'une des classes les plus générique du vocabulaire RDF), toute classe définie dans n'importe quel vocabulaire RDF peut être utilisée pour constituer le corps d'une annotation. Nous utiliserons cette possibilité afin de représenter les contenus saisis par les utilisateurs comme étant des *Post* tels qu'ils sont définis dans le vocabulaire SIOC.

Enfin, la propriété *context* pointe vers *Anything*. Cela veut dire que l'on peut mettre ce que l'on veut en relation avec une annotation par le biais de cette propriété. La propriété

context lie l'annotation à une chaîne de caractères ou à une *Ressource* RDFS. Les annotations présentées dans le projet Annotea utilisent la technologie XPointer³¹ pour représenter des sélections de passages dans des documents HTML et les lier aux annotations.

Un Xpointer est une chaîne de caractères composée de la concaténation d'une URI et de caractères permettant de représenter un ou plusieurs points précis dans le document (que ces points soient des balises XML ou des caractères dans un champ texte). XPointer peut être utilisé pour représenter une sélection en listant un point de départ et un point de fin dans le document sélectionné. Cette solution, l'utilisation de XPointer pour la propriété *context*, est adaptée à l'ancre d'annotations dans des documents HTML conformément au principe de streams et de segments de Agosti et Ferro, mais nous proposerons une autre pratique pour permettre d'étendre la portée des annotations à d'autres types de documents.

La chaîne de caractères suivante est un Xpointer :

`http://example.com/page.html#xpointer(id("Main")/p[2])`

Cette chaîne permet de localiser dans une page HTML, identifiée par l'URI `http://example.com/page.html`, un élément particulier de l'arbre HTML, ici le deuxième noeud `<p>` après l'élément HTML d'identifiant "Main". Dans la syntaxe Xpointer, il est possible de désigner un caractère précis par son indice dans une chaîne de caractère constituant le contenu d'un noeud. Par exemple, si on a le noeud HTML avec le contenu suivant

`<p>un exemple de texte</p>`

Le "u" est le premier caractère donc d'indice 0, le "n" a l'indice 1, l'espace qui suit a l'indice 2 ... Le XPointer dans ce cas pour sélectionner le caractère "n" serait :

`http://example.com/page.html#xpointer(point(id("Main")/p[2].1))`

Enfin, pour sélectionner toute un intervalle de texte, on peut utiliser les mots-clefs *start-point* et *end-point*. Pour sélectionner le mot "exemple" dans notre texte, le XPointer devient :

`http://example.com/page.html#xpointer(start-point(id("Main")/p[2].3),end-point(id("Main")/p[2].9))`

Nous ne détaillons pas toute la syntaxe de XPointer. Pour plus d'information il est possible de consulter la recommandation du W3C³².

Annozilla

Le projet Annotea proposant un protocole de communication et un format de représentation des annotations, il ouvre la voie de l'interopérabilité dans le domaine et fournit les recommandations nécessaires pour proposer de nouveaux outils compatibles avec

31. <http://www.w3.org/XML/Linking>

32. <http://www.w3.org/TR/xptr-xpointer>

ceux existant respectant le protocole et le format de données. Annozilla³³ est un client pour l'annotation compatible uniquement avec le navigateur Firefox maintenu à ce jour jusqu'à la version 3.5 de firefox. Le projet Annotea propose le téléchargement d'autres outils clients ou serveurs.

Le client Annozilla doit être paramétré pour lui indiquer avec quel serveur il va pouvoir communiquer, récupérer des annotations et en stocker de nouvelles. Annozilla est développé en javascript et permet d'annoter n'importe quelle page HTML. Lorsque le navigateur Firefox accède à une ressource, le module Annozilla envoie une requête GET au serveur et liste l'ensemble des annotations faites sur la page dans une colonne à gauche du navigateur. Il colore les passages annotés en les encapsulant dans une balise ** ou *<div>* dont le fond est coloré en jaune.

Le client permet de créer de nouvelles annotations par le biais d'un formulaire. Le formulaire propose une liste de types d'annotation. La liste de types est la suivante :

- Commentaire (comment)
- Exemple (example)
- A voir aussi (see also)
- Question (question)
- Explication (explanation)
- Changement (change)
- Recommandation (advice)
- Propos malveillant (flame)

Les figures 3.8 et 3.9 présentent les deux usages du client Annozilla.

Quand on observe le RDF généré par une annotation de l'un de ces types, on se rend compte que chacun de ces types est censé être défini comme une spécialisation de la classe *Annotation* du vocabulaire proposé par Annotea. Ce mécanisme d'extension de la classe *Annotation* est celui que nous adoptons et recommandons pour définir de nouvelles formes d'annotations, cependant, nous allons proposer des formulaires d'annotation, en fonction des types, plus riches que ceux proposés dans Annozilla.

Marginalia

Marginalia³⁴ est le terme anglais pour parler de notes dans la page. Le projet Marginalia est né suite aux travaux de Andrew Feenber et Cindy Xin. Le code du projet est essentiellement écrit par Geof Glass. L'objectif de Marginalia est de fournir un outil d'annotation pour améliorer les forums classiques dans un contexte pédagogique. Le projet a été initialement soutenu par the British Columbia Campus notamment en attribuant un prix au projet récompensant les innovations dans le domaine de l'éducation.

La démonstration de Marginalia est intégrée à la plateforme Moodle qui est un CMS spécialisé dans la création de contenus pédagogiques. Moodle est aussi désigné comme

33. <http://annozilla.mozdev.org/>

34. <http://www.geof.net/code/annotation>

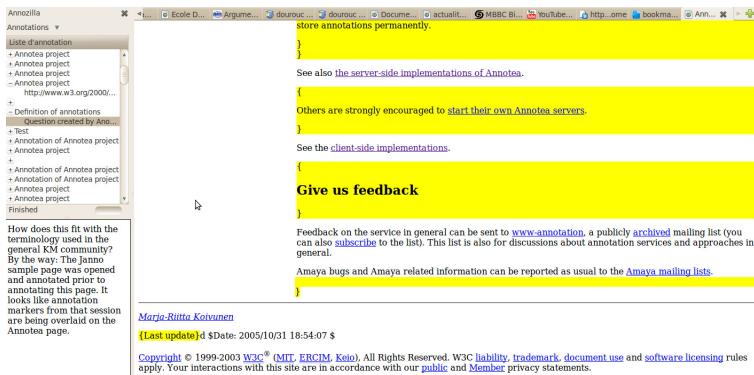


FIGURE 3.8 : Utilisation d'annozilla pour lire des annotations

The screenshot shows the Annotea Project website. On the left, there's a sidebar with links like 'About Annotea', 'How to ...', 'Annotea', 'Mailing list', 'Contributors', and 'Related Links'. The main content area has a heading 'Annotea Project Overview'. Overlaid on the page is a 'Annotation' dialog box. The dialog has fields for 'Annotation title' (set to 'Annotation'), 'Author' (set to 'Anonymous'), 'Server' (set to 'http://annotate.w3.org/annotations'), 'Type' (set to 'Comment'), and 'Language' (set to 'Anglais'). There are 'Save' and 'Cancel' buttons at the bottom. To the right of the dialog, there's explanatory text about the Annotea project.

FIGURE 3.9 : Utilisation d'annozilla pour créer des annotations

étant un environnement virtuel d'éducation. Marginalia est développé en javascript, distribué en plugin pour Moodle ou en code source à intégrer dans n'importe quel environnement. Marginalia se résume à un ensemble de fichiers javascript facilement intégrables dans n'importe quel site Web.

Marginalia permet aux utilisateurs de faire des sélections de passages dans des pages HTML du site ou le code javascript est intégré. A chaque passage sélectionné, l'utilisateur peut associer un commentaire, créant ainsi une annotation. Un espace sur le côté du site, une marge, est réservé pour afficher le contenu des annotations. Les passages sont mis en couleur.

En terme de fonctionnalité, Marginalia ressemble fortement à ce que permet de faire Annozilla. A la différence que Marginalia est intégré dans un site Web et non au navigateur.

teur, et ne permet donc pas d'aller annoter des ressources externes au site Web intégrant le code. De plus, Marginalia , n'intègre pas le protocole Annotea et les annotations de Marginalia ne peuvent pas être échangées avec une autre application.

Là où Marginalia a vraiment quelque chose de nouveau à apporter, c'est au niveau de son intégration dans des environnements type CMS dont le contenu peut changer. Par exemple, les annotations faites dans Moodle prennent en compte le fait que le contenu du site Moodle peut changer. Si une ressource annotée change et qu'il devient impossible de localiser l'annotation dans la ressource, alors celle-ci change de statut et une notification signale que l'annotation est peut-être obsolète.

Co-ment

La société Sopinspace développe l'outil Co-ment³⁵ permettant d'annoter un document HTML. L'outil existe sous forme de service, l'utilisateur peut alors déposer un texte dans le site, celui-ci devient alors annotable. L'outil est aussi développé sous la forme d'un module Drupal (Drupal est un CMS en php). Là encore, le langage utilisé est javascript, ce qui facilite l'intégration du code dans une application Web. Les annotations apparaissent dans le texte en étant colorées en jaune ou en bleu si l'annotation est celle en cours de lecture. Il est alors possible de répondre à une annotation par le biais d'un formulaire de saisie. La figure 3.10 présente une annotation en cours de lecture dans la version d'essai gratuite de co-ment.

The screenshot shows a web-based annotation tool. At the top, it says "Excerpt from Solon Robinson" and "modifié il y a 2 semaines, 4 jours / 1 version". Below this is a toolbar with "Texte" and "Versions (1)". A navigation bar indicates "6 discussions, parcourir par : emplacement" and shows page 5/6. The main content area displays a historical text by Solon Robinson. An annotation is visible in the text: "liberty which God in his wisdom and mercy deprived him or...". A blue box highlights this text, and a yellow box highlights the word "God". A sidebar on the left shows a user profile for "taefr" (created on 08-07-2011 21:16) and a "Nouvelle réponse" form with fields for "Nom d'utilisateur:", "Email:", "Titre:", "Re: taefr", and "Contenu:". The right side of the interface includes social sharing buttons and a vertical sidebar with icons.

FIGURE 3.10 : Utilisation de co-ment pour lire une annotation et y répondre

35. <http://www.co-ment.com/>

LogiLogi

L'application LogiLogi³⁶ développée en Ruby et en javascript est présentée comme étant une nouvelle forme de forum. Un Logi est l'équivalent d'un Post dans un forum. Une fois saisi, un Logi devient un élément de discussion. L'auteur peut le catégoriser par un système de tags, mais surtout, les autres utilisateurs peuvent y répondre. Il y a plusieurs façons de répondre à un Logi. La première s'apparente aux forums classiques. Un utilisateur peut attacher un nouveau Logi comme étant une réponse à un Logi déjà existant. Enfin, il est possible d'annoter des passages d'un Logi. Les passages annotés sont alors écrits en bleu et au survol, le texte ajouté lors de l'annotation apparaît dans une info-bulle. LogiLogi intègre aussi un système de vote permettant de signaler les Logi que l'on apprécie ou non, permettant ainsi de mettre en place un système de modération au mérite.

La véritable différence de LogiLogi est que chaque contenu ajouté lors d'une annotation est à son tour un Logi. Du coup, il devient possible d'annoter le contenu d'une annotation. La figure 3.11 présente un Logi annoté dans LogiLogi.

The screenshot shows a detailed view of a Logi annotation in the LogiLogi application. The main content area is titled "Content (definition)" and is attributed to "Matthijs van der Meer". Below the title, there is a "Rating" section with a star icon and the text "Rating: XXX★★★★★". To the right of the rating, there is a "Positive votes" section showing "0 votes".

The main content area contains several paragraphs of text, each with a small blue "info" icon. One paragraph discusses the "PFR" scheme, another about R0 content, and others about R1, R2, and R3 content. There is also a section for "Logi links" with a single entry: "1. PFR=Matthijs_Van_Der_Meer_165".

On the left side of the main content area, there is a sidebar with the heading "Content (definition)". It includes a "Content either denotes intrinsically carrying meaning." section and a "Expressions along the scheme of PFR yield the following:" section. At the bottom of the sidebar, there is a button labeled "Attach something to a phrase".

At the very bottom of the screen, there is a footer bar with the text "Average rating: XXX★★★★★" and "Rating weight: ~".

FIGURE 3.11 : Affichage d'un Logi-annotation dans LogiLogi

Diigo

Le terme de Social Bookmarking fait référence à l'activité de classer ses lectures par un système de tags et de partager la classification obtenue avec d'autres utilisateurs. Citons par exemple l'application de social bookmarking Delicious³⁷. Cependant, il existe d'autres

36. <http://logilogi.org>

37. <http://delicious.com/>

solutions et Diigo³⁸ est l'une d'entre elles. Diigo permet de faire du social bookmarking tout comme Delicious, mais l'application permet aussi de faire des annotations.

L'architecture de Diigo est assez classique, un serveur regroupe toutes les données et des clients permettent d'accéder à ces données. Les clients sont en fait des plugins pour les différents navigateurs Web existant ou bien des applications pour plateformes mobiles (ipad, iphone et android). Le serveur regroupe les données de l'activité de chaque utilisateur dans un espace appelé "My Library". L'utilisateur gère aussi une liste de contacts, "My Network", et peut ainsi accéder et partager les données d'autres utilisateurs. Les pages annotées peuvent être sauvegardées avec les annotations dans la plateforme Diigo. Diigo réalise une copie du HTML de la page consultée au moment de la création d'annotations. Les pages sont aussi copiées en tant qu'images (ou screenshots), permettant d'ajouter des annotations ancrées par le biais de formes plus complexes que juste un passage surligné (par exemple la possibilité de dessiner un cercle). La figure 3.12 illustre ce propos.

Diigo est probablement l'outil d'annotation le plus en avance en terme de fonctionnalités. Il permet d'annoter toutes ressources du Web à partir de plusieurs navigateurs et de plusieurs plateformes. Il permet de partager ses annotations avec un réseau social. Cependant, l'outil ne se positionne pas vraiment comme un outil de discussion mais plus comme un outil de lecture active ou de social bookmarking amélioré. S'il est possible d'ajouter des commentaires suite à des annotations ou à un bookmark, il n'est pas possible d'organiser ces commentaires autrement que dans un fil de discussion et avec un autre statut que celui de simple commentaire. Le contenu d'une annotation ne devient pas non plus une nouvelle ressource à son tour annotable. Enfin, s'il est possible de choisir la couleur d'une annotation, celle-ci n'est pas vraiment associée à une action particulière dans l'application. Toutes les annotations, quelques soient leurs couleurs, ont au niveau de l'application la même sémantique.

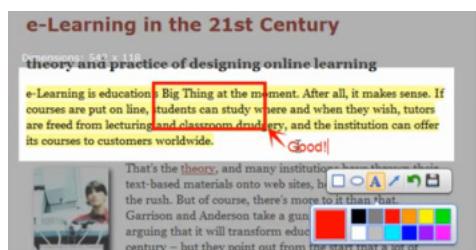


FIGURE 3.12 : Annotation se référant à une forme et non à un surlignage dans Diigo

38. <http://www.diigo.com/>

Google Docs Discussions

L'application Google Docs³⁹ permet à plusieurs utilisateurs de partager et éditer de façon synchrone un document. Google Docs se positionne comme un outil de co-rédaction, dans la même catégorie que les wikis. Google Docs est doté d'une fonctionnalité, "google discussions" permettant de surligner un passage et d'y associer un commentaire. Chaque commentaire devient un fil de discussion, il est possible de répondre aux commentaires.

On est loin de l'annotation de toutes les pages du Web, mais google discussion est une fonctionnalité ajoutée à une application de co-rédaction. Celle-ci a au moins l'intérêt de souligner la problématique du devenir des annotations dans le contexte de ressources annotées en évolution dans le temps. Au niveau de la sémantique des annotations, celles-ci ne sont là que dans le but de faciliter la co-rédaction. Et donc, annoter un passage c'est en quelque sorte une façon de souligner une amélioration à apporter au document.

3.4 L'analyse des SocNets

Internet et le Web sont bâtis sur la notion de réseau social. Il s'agit de permettre la communication, l'échange entre des individus. Cependant, certaines applications sur le Web mettent particulièrement cet aspect en avant, faisant de la création de liens entre individus et de la gestion de ces liens les fonctionnalités primordiales. Ces applications sont souvent appelées Réseaux Sociaux Numériques ou encore SocNets. Nous allons dans cette partie présenter la notion d'identité numérique, l'identité construite au sein des SocNets puis proposer une terminologie permettant de classifier différentes formes de SocNets.

3.4.1 Identité numérique

La forte importance de l'utilisateur en réseau dans les applications actuelles d'Internet et du Web crée un ensemble de questionnements sur l'identité numérique, sur les modes de représentation de l'individu et sur sa façon de se présenter en ligne. Anne Sophie Béliard s'intéresse aux modes de présentations dans les forums [Béliard, 2009]. Elle souligne la contradiction entre l'usage du pseudonyme comme moyen de rester anonyme sur Internet mais aussi comme premier marqueur de l'identité numérique. L'utilisateur est attaché sentimentalement à son pseudo. Il joue un rôle en ligne dont le pseudo est le premier révélateur et il va chercher souvent à pérenniser son usage d'espace en espace. Julie Denouël-Granjon [Denouël-Granjon, 2008] en décortiquant les différents types d'identifiants de la messagerie en ligne montre à quel point le pseudo est le premier révélateur de notre identité.

Durant nos travaux, nous avons eu l'occasion de publier à plusieurs reprises en co-auteur avec Fanny Georges sur les problématiques de l'identité numérique. Notre pre-

39. <https://docs.google.com/>

mier article ensemble [Georges *et al.*, 2009] présente une comparaison de la représentation de l'identité dans Facebook et MySpace. Nous présentons également notre approche transdisciplinaire dans [Georges *et al.*, 2010b]. Nous défendons que le modèle de l'identité numérique tel qu'il est défini par Fanny Georges est un outil permettant de rapprocher informatique et sciences sociales dans le cadre d'analyses ou de conceptions d'applications Web 2.0. Dans [Georges *et al.*, 2010a], nous dressons une première liste de pratiques d'identification en lignes dans lesquelles la préservation de son anonymat joue un rôle important. L'anonymat dans les débats en ligne est un sujet non négligeable, il permet de libérer la parole mais aurait aussi un impact sur la responsabilisation des propos et l'un des freins les plus souvent invoqué à la mise en place de la démocratie électronique est encore la peur du contrôle des citoyens par l'analyse des données qu'il produit.

Fanny Georges propose une décomposition de l'identité numérique en trois composantes :

1. L'identité déclarative constituée de tous les renseignements que fournit l'utilisateur pour être identifié dans le système. En particulier toutes les informations remplies lors de son inscription au système et tous les liens qu'il déclare entretenir avec d'autres utilisateurs.
2. L'identité agissante constituée de toutes les actions de l'utilisateur dans le système pour lesquelles le système renvoie un signal. Par exemple un Post dans un forum est affiché donc signalé, un changement de statut dans Facebook produira une notification, un signal.
3. L'identité calculée, constituée d'un ensemble d'indicateurs produits par des calculs par le système. Par exemple des statistiques d'usage ou le nombre d'amis dans Facebook.

L'image 3.13 tirée de la thèse de Fanny Georges, représente les 3 composantes de l'identité numérique.

3.4.2 Terminologie

Dans sa thèse [Erétéo, 2011], Guillaume Erétéo propose la catégorisation suivante des types de réseaux sociaux souvent considérés en analyse de réseaux sociaux issus d'usage du Web 2.0 :

- Les réseaux explicites ou les réseaux considérant des relations déclarées entre les individus. Les premières analyses de réseaux sociaux se basent sur des données relationnelles obtenues par le biais d'interviews ou de questionnaires dans lesquels les acteurs du réseau sont invités à déclarer eux-mêmes les relations qui les lient aux autres. Ce genre de données correspond à l'usage actuel de nombreux outils de réseaux sociaux comme Facebook par exemple.

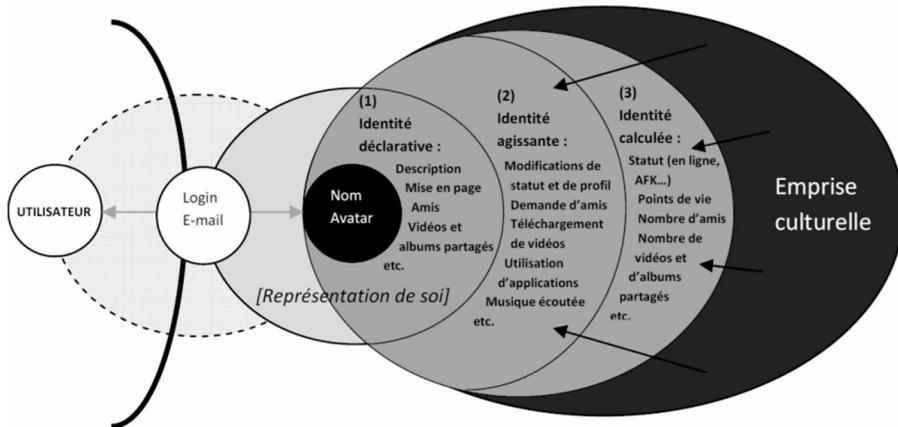


FIGURE 3.13 : Le modèle de l'identité numérique

- Les réseaux d'interactions ou les réseaux dans lesquels toutes interactions entre individus sont considérées comme des relations. Ce type de données est en général issu d'observation d'actions ou de traces d'actions. La plupart des applications dites sociales du Web permettent des interactions entre utilisateurs et donc de nourrir ce type de réseaux.
- Les réseaux d'affiliation dans lesquels sont inférés des liens à partir de l'observation de similarités entre acteurs.

Nous adoptons la terminologie suivante :

- Réseau explicite
- Réseau implicite
- Réseau déclaratif
- Réseau d'interactions
- Réseau calculé

Réseaux explicites

Nous n'adoptons pas tout à fait la même définition que Guillaume Erétéo concernant les réseaux explicites. Nous considérons comme réseau explicite tout réseau qui est clairement présenté à l'utilisateur. Si dans une application, l'utilisateur peut avoir un retour, une représentation du réseau social ou d'une partie du réseau social définie par l'application, alors nous parlons de réseau explicite. Beaucoup de réseaux déclaratifs, comme par exemple Facebook sont des réseaux explicites. Mais il pourrait très bien y avoir des réseaux explicites qui ne sont pas déclaratifs. Par exemple, l'application ArnetMiner⁴⁰ présente un réseau à l'utilisateur qui est le résultat, en partie, d'un moissonnage de données et d'inferences sur ces données. ArnetMiner présente les relations de co-auteur mais aussi les

40. <http://arxiv.org/>

relations organisationnelles du domaine de la recherche (comme par exemple la relation encadrant-encadré durant une thèse). ArnetMiner déduit aussi à partir des thèmes des papiers un réseau d'affiliation entre les utilisateurs.

Réseaux implicites

A l'opposé des réseaux explicites, les réseaux implicites sont des réseaux qui ne sont jamais présentés aux utilisateurs. Bien souvent, on retrouve un système de recommandations, comme par exemple dans l'application de vente en ligne Amazon, qui se base sur une analyse des activités du réseau. Le système va, par exemple, recommander un achat à partir de l'historique d'achats de plusieurs utilisateurs. Dans ce cas, le réseau existe, mais il n'est pas explicité aux utilisateurs. On parlera alors de réseau implicite. Précisons enfin qu'un réseau implicite peut exister sans pour autant être utilisé. C'est le cas actuellement dans la plupart des outils de discussion types wikis, blogs ou forums.

Réseaux déclaratifs

Depuis peu, sur le Web, s'est développé un nouveau type d'outils, appelés SocNet ou plus couramment "outil de réseau social". Facebook en est l'exemple le plus connu. Dans ce nouveau type d'outil, les utilisateurs déclarent leurs relations. Nous qualifions ce type de réseaux de "déclaratif" conformément à la terminologie de l'identité numérique de Fanny Georges [Georges, 2007]. Un réseau déclaratif est composé de relations entre utilisateurs déclarées par les utilisateurs.

Réseaux d'interactions

Depuis les débuts d'Internet, l'idée de communauté est présente. Les outils proposés par Internet et par le Web offrent toujours plus de possibilités d'interactions entre utilisateurs. Les réseaux d'interaction, réseaux déduits à partir d'action des utilisateurs impliquant d'autres utilisateurs, ont été le cadre de nombreux travaux. [Mutton, 2004] étudie les interactions entre utilisateurs via IRC. Il définit des relations entre les utilisateurs en fonction de la fréquence des messages ou du contenu des messages (si un utilisateur explicite clairement à qui il s'adresse dans le corps de son message, ou s'il répond peu de temps suite à l'intervention d'un autre). [Tyler *et al.*, 2005] étudie les échanges de mail pour déduire un réseau. Les réseaux usenet, les newsgroups ou encore les forums ont fait l'objet de nombreuses études [Rheingold, 1993], [Revillard, 2000]. Dans des réseaux d'interaction, on déduit des relations entre individus à partir de leurs interactions. La plupart des réseaux d'interaction ne sont pas explicités aux utilisateurs. Les réseaux d'interaction sont déduits des actions utilisateurs et font donc référence à l'identité agissante du modèle de Fanny Georges.

Réseaux calculés

Enfin, pour finir l'analogie au modèle de l'identité de Fanny Georges, nous proposons le terme de réseaux calculés pour désigner l'ensemble des réseaux déduits à partir d'un calcul. On retrouve donc les réseaux d'affiliation dans le groupe des réseaux calculés. Des réseaux calculés peuvent être le résultat d'analyses d'autres réseaux et souvent de réseaux d'interactions. Par exemple, on peut observer de nombreuses interactions mail entre deux individus et en déduire une relation entre ces individus et la qualification de cette relation peut être le résultat d'un calcul.

3.5 Limites de l'état de l'art

De nombreux outils dits Web 2.0 permettant aux utilisateurs d'exprimer des opinions en ligne existent. Ces outils héritent des problématiques du Web 2.0, répondent au besoin utilisateur de pouvoir créer du contenu mais ne s'adressent pas au besoin de prendre des décisions en concertant le grand nombre et les parties prenantes.

Nous allons par la suite proposer une solution de débat 2.0 basée sur l'annotation socio-sémantique permettant l'expression d'opinions, à partir de laquelle une première structuration du réseau social des intervenants peut être extraite et mettant fortement l'accent sur l'utilisation de formats de données standardisés pour l'interopérabilité.

Notre solution a été conçue en prenant en compte les problématiques suivantes :

- La problématique de la masse de données produites et de son manque de structuration dans les outils de discussion. La plupart des outils de discussion se contentent d'une structuration à base de relation de "réponse à". Un grand nombre d'utilisateur peut créer un grand nombre de données et donc nécessiter un lourd travail de synthèse.
- La problématique de l'interopérabilité des données permettant de penser à une réutilisation des données des débats dans d'autres contextes, comme l'analyse du réseau social, ou l'aide à la prise de décision. La démocratie électronique devrait produire un ensemble d'applications permettant d'améliorer la pratique de la démocratie dans son ensemble. Il faut donc considérer l'usage d'un outil de débats 2.0 dans un contexte applicatif varié. L'application de débats 2.0 doit fournir des données interopérables pouvant être utilisées par exemple dans un outil dédié à la prise de décision, à l'analyse des interactions entre parties prenantes ...
- La problématique de l'insularisation des parties prenantes, groupes de pressions, communautés d'opinions. En proposant un début de structuration du réseau social des participants des débats, nous espérons créer un ensemble d'indicateurs permettant d'identifier les tendances de ces groupes, les leaders d'opinions, les oppositions entre parties prenantes ... La structuration proposée devrait permettre d'identifier les éléments d'insularisation et à terme, de les prévenir.

Enfin, notre démarche a permis d'évaluer l'acceptabilité de l'annotation socio-sémantique en tant qu'outil de discussion, problématique intrinsèque à l'annotation discursive.

CHAPITRE

4

Proposition

Préambule

Ce chapitre présente l'ensemble de la proposition de thèse, à savoir l'usage de l'annotation comme nouvelle forme d'interaction, la standardisation des outils de communication dans le cadre des débats en ligne et enfin, une première approche pour la détection de groupes d'opinion dans les débats en ligne.

Sommaire

4.1	Spécification d'un outil de débat 2.0	120
4.2	L'annotation	125
4.3	Les réseaux sociaux	133
4.4	Interopérabilité	138
4.5	Exemple complet	145

Notre démarche expérimentale a été fortement guidée par les objectifs du projet Intermed. Le premier besoin était de générer de la donnée qui nous appartienne d'expression d'opinions politiques liées à la gestion de bien communs. Nous avons donc construit notre plateforme expérimentale autour de ce besoin. Notre outil de débat permet de créer des discussions et d'exprimer des opinions. Le choix de l'annotation ayant été pris dès le début mais n'étant pas une fonctionnalité aisée à implémenter, notre prototype actuel est le dernier d'une longue série d'itérations alternant les phases de conception, d'implémentation et d'expérimentation.

Notre prototype repose sur l'utilisation d'annotations pour construire des discussions. La forme de saisie de nos annotations permet de définir des relations entre les données de discussion produites par les utilisateurs. Les relations que nous proposons sont essentiel-

lement des relations d'accord ou de désaccord. A partir de ces relations entre les éléments de discussion entre utilisateurs, nous proposons de déduire des relations entre les utilisateurs de type soutien ou opposition. Le prototype propose ensuite des formats d'échange des données de discussions et du réseau social afin de faciliter le traitement de ces données et l'interopérabilité entre cet outil et d'autres.

4.1 Spécification d'un outil de débat 2.0

Une application de débats 2.0 a pour objectif principal de permettre à un ensemble de citoyens de débattre sur des sujets les concernant. Cependant, il ne peut pas y avoir de débats sans information. Il faut que les citoyens accèdent à certaines informations pour pouvoir prendre part au débat. De plus, le débat doit produire une opinion collective. Celle-ci peut être présentée sous forme de synthèse. L'essentiel est que cette opinion collective rende compte des différentes positions émises et de la répartition ou acceptation de ces positions parmi les citoyens débattant.

Un outil pour les débats 2.0 doit permettre :

- **La discussion.** Un outil de débats 2.0 doit permettre de discuter sur des sujets. Il permet donc à minima de faire ce que permet de faire un outil de discussion type blog ou forum.
- **La synthèse.** Un outil de débats 2.0 doit permettre de synthétiser le contenu d'une discussion en présentant le modèle commun construit par les citoyens ayant participé à la discussion. Cette synthèse pourra par la suite servir de support de communication.
- **L'expression d'opinions.** Un outil de débats 2.0 doit permettre d'exprimer clairement une opinion. En particulier, l'expression d'opinions de support et d'opposition est le minimum.
- **La représentation de l'identité du citoyen.** Il ne s'agit pas nécessairement d'identifier clairement les citoyens qui participent à la discussion, mais il est important de donner au citoyen la possibilité de se créer une identité de débattant. L'usage de pseudonymes est donc nécessaire, l'usage d'un anonymat total ne doit pas être la règle mais rester possible.

Ces 4 fonctionnalités nous semblent être le minimum pour un outil de débats 2.0. Elles définissent déjà une nouvelle forme de structuration des discussions basées sur le modèle de discussion classique mais enrichit par l'usage des opinions pour définir de nouvelles relations entre les éléments de discussion. Elle fixe aussi des règles nécessaires pour le bon déroulement des discussions : imposer l'identification des interlocuteurs permet de construire des discussions sur la durée et réduit la dé-responsabilisation dans l'expression, ce qui entraînerait une hausse des positions sans fondements, juste pour le jeu. Préciser que les débats 2.0 doivent produire des synthèses, quelque-soit au final la forme que celles-ci adoptent (tableau de bord, cartographie, document texte ...), souligne que les dé-

bats ont un objectif. Il ne s'agit pas juste de discuter mais bien de construire à plusieurs une opinion commune.

Ces 4 fonctionnalités permettent à un ensemble de citoyens identifiés de débattre et de créer une opinion commune. Ce service est le service de base défini par les débats 2.0. Il s'agit maintenant d'adopter une implémentation qui permette le passage à l'échelle. Cependant, nous considérons que le passage à l'échelle repose essentiellement sur une bonne structuration des discussions, sur la production de synthèses à n'importe quelle étape de la discussion et au couplage de l'outil de discussion avec un outil de gestion du réseau des citoyens. Il faut que les citoyens puissent, en cours de discussion, s'organiser en groupes d'opinions, en particulier en groupes de soutien et en groupes d'opposition.

Si on se réfère au paradigme SLATES, les 4 fonctionnalités précédentes doivent être accompagnées des fonctionnalités suivantes, pré-requises pour tout espace 2.0 :

- **Search** : Une solution pour chercher du contenu. Les données produites dans un espace de débat doivent pouvoir être classées selon plusieurs critères pour faciliter la navigation. Dans notre cas, nous définissons par exemple un ensemble de métadonnées sur lesquelles les contenus peuvent être indexés : les auteurs, la date, la thématique, l'opinion exprimée, la longueur du contenu, le nombre de personnes soutenant ou s'opposant ...
- **Link** : Il faut pouvoir lier les données entre elles. Les nouveaux contenus produits dans un outil de débat 2.0 sont liés par un statut de réponse ou une relation spécialisant ce statut de réponse (avis d'opposition, avis de soutien, expression d'une incompréhension ...).
- **Authoring** : Le contenu est créé par les utilisateurs du système même s'il est possible que des agents logiciels créent aussi du contenu en complément.
- **Tag** : Tout contenu est classifiable par un système de tags. Le tagging est une des formes d'annotation essentielle pour mieux indexer les données des débats.
- **Extensions** : Pour l'instant, nous ne greffons pas de système de recommandation dans un espace de débats. Mais la structuration que nous proposons des données produites doit permettre de faciliter la création de systèmes de recommandations (par exemple, il est possible de construire ce genre de recommandations à partir des données de notre système : "vous devriez discuter avec untel qui n'a pas les mêmes opinions que vous mais s'intéresse aux mêmes sujets").
- **Signal** : Tout nouveau contenu est signalé minimalement par le biais de flux RSS 1.0 couplé à des vocabulaires RDF.

Comme nous l'avons déjà défendu, un outil de débat 2.0 est une étape dans un processus de participation ou de démocratie électronique. Cet outil doit donc être interopérable avec d'autres. Le service de débat 2.0 doit donc définir des entrées sorties, des données et des formats de données que peuvent lui transmettre d'autres applications et des données et des formats de données que l'outil doit pouvoir fournir à d'autres applications. De plus, notre définition de réseau social est basée sur la prise en compte des interactions entre utilisateurs. Nous pensons que un outil de débats 2.0 fournit un ensemble de données très

enrichissantes pour structurer le réseau social des citoyens.

Nous proposons qu'un outil de débat 2.0 puisse recevoir en entrée un ensemble de ressources servant de support d'information de base aux débats. Cet ensemble de ressources est présenté essentiellement sous la forme d'un ensemble d'URLs et de tags sur ces URLs permettant de catégoriser les ressources en question.

En sortie, nous proposons que l'application de débat 2.0 transmette les données de l'opinion collective. Cette donnée fortement sociale est composée des productions des utilisateurs reliées par des relations traduisant les opinions exprimées.

Bien que notre outil fournit des informations permettant de structurer un réseau social, il peut s'avérer intéressant au niveau de l'application de débat 2.0 d'avoir un réseau social de départ déjà structuré. En sortie, l'application fournira un réseau social dont la structure est enrichie par des relations de soutien ou d'opposition.

Afin de répondre au besoin d'interopérabilité, nous proposons des formats pour ces données reposant sur l'utilisation de vocabulaires RDF déjà présentés et d'extensions.

Avant de présenter les données que permet de produire un outil de débats 2.0, nous présentons quelques scénarios d'usage illustrant notre proposition.

4.1.1 Scénarios d'usage

Nous présentons dans cette partie des scénarios d'usage d'outils de débat 2.0. Ces scénarios sont le fruit d'un travail de co-conception avec la société Intactile Design¹ suite à plusieurs expérimentations. Ce travail a produit 3 scénarios d'usage. L'intégralité des dessins est produite par intactile Design.

Le premier scénario correspond à une expérimentation menée durant la thèse suite à cette co-conception. Il s'agissait d'organiser un débat au sein du laboratoire afin de préparer l'Assemblée Générale du changement de direction du laboratoire. Ce scénario illustre bien un cas d'usage où le débat en ligne n'est qu'un complément de débats réels. Nous soulignons dans ce scénario le besoin de modération qui ici était envisagée comme étant faite par un modérateur à partir de votes sur les opinions exprimées. Nous soulignons le besoin de l'anonymat qui donne un statut différent aux expressions non signées. L'anonymat permet de libérer la parole dans certains cadres. Nous avions d'ailleurs observé au cours de l'expérimentation le besoin de l'anonymat en particulier lors des débats concernant les relations entre doctorants et encadrants. Enfin, nous soulignons le besoin de produire une synthèse pour permettre de restituer le contenu des discussions. Le storyboard en annexe ?? est la version produite par intactile de ce cas d'usage. L'image 4.1 est l'une des pages du storyboard débat en ligne complément de débat en présence, elle montre un moment de débat en présence, l'AG du laboratoire, et la retransmission de ce qui s'est dit en ligne.

Le deuxième scénario présente plus particulièrement les fonctionnalités de l'application de débat. Ici, un personnage fictif, Hiro pêcheur japonais de thon rouge, souhaite dé-

1. <http://www.intactile.com/>

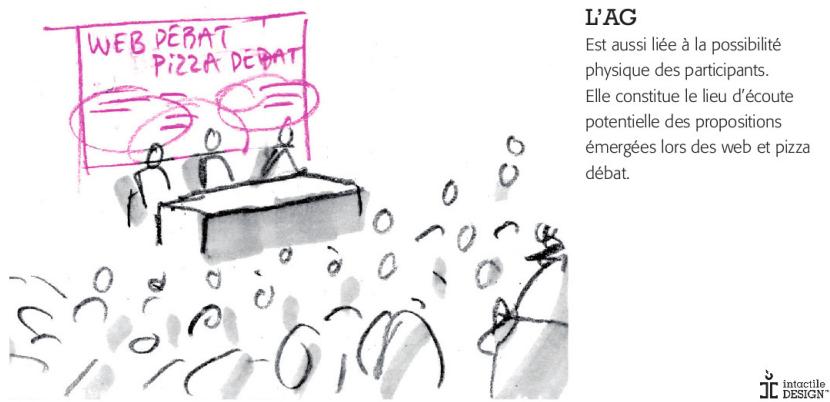


FIGURE 4.1 : AG du labo avec synthèse des discussions en ligne

fendre sa position de partisan de la pêche au thon. L'outil de débat est pour lui l'occasion de trouver du soutien et de développer son argumentation. Le scénario décrit la phase de création d'une identité dans l'application, puis la création d'un sujet de débat (contre l'interdiction de la pêche au thon rouge), et enfin le développement de l'argumentation. Nous souhaitions souligner la présence d'agents assistants guidant Hiro par le biais de recommandations (d'autres débats à lire, des personnes avec qui discuter). Mais aussi souligner le besoin de gérer une partie privée et une partie publique des débats. Certains arguments doivent attendre d'être mûrs avant d'être rendus publiques. Il faut trouver la bonne formulation et prévoir les contre-arguments pour savoir les désamorcer. Construire un bon argumentaire se fait rarement seul. Il faut l'aide d'autres personnes qui vont permettre d'éprouver les arguments en privé avant de les exposer à tous.

Nous avons aussi introduit dans ce scénario l'idée de rôles des débattants définis par un système de compétences. Certains utilisateurs sont considérés comme étant des spécialistes, d'autres comme étant des structurants ... Nous pensons qu'en se basant sur l'activité des utilisateurs, il est possible de proposer une catégorisation des usages et donc des profils.

Ce scénario est aussi le cas d'usage mettant l'accent sur les possibilités de l'annotation en tant qu'outil de discussion.

Le storyboard en annexe ?? est la version produite par intactile de ce cas d'usage.

L'image 4.2 est l'une des pages du storyboard décrivant comment développer son argumentaire en ligne.

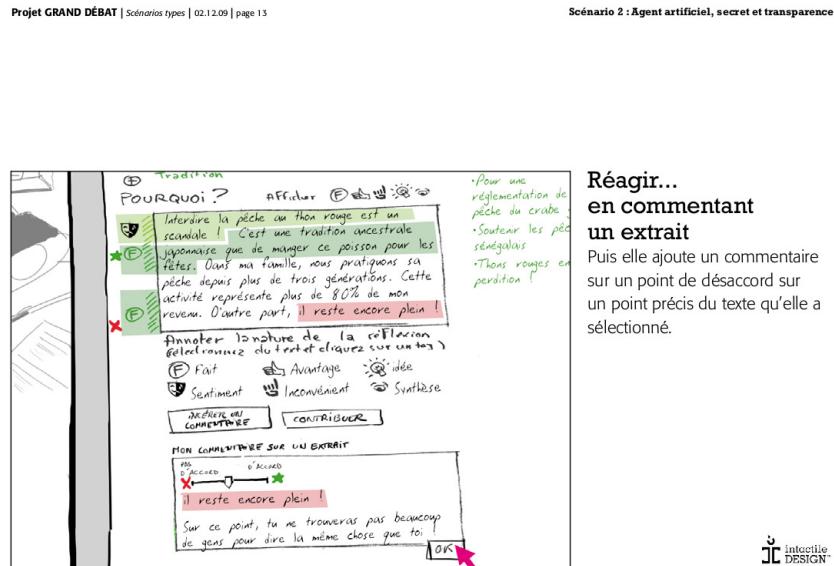


FIGURE 4.2 : interface simplifiée pour l'annotation dans le scénario de construction de cause

Enfin, le troisième scénario met l'accent sur les outils d'assistance à la création de synthèses. Un ensemble d'idées d'interfaces de visualisation des données des débats est proposé. Nous pensons que la cartographie des données est une solution pour aider à naviguer dans la grande quantité de données produites et pour donner une vision synthétique des discussions. La visualisation de données est une forme d'extensions au sens du paradigme SLATES qui nous semble primordial dans ce contexte. Le storyboard en annexe ?? est la version produite par intactile de ce cas d'usage. L'image 4.3 est une page du storyboard d'assistance à la création de synthèses, elle montre la possibilité de colorer les commentaires dans un débat en fonction des jugements qui leur sont associés.

4.1.2 n-opposition et expression d'opinions

Nous avons présenté dans la partie Background et état de l'art la théorie des n-oppositions. Cette théorie propose de représenter des relations d'oppositions entre des jugements de valeur sous une forme géométrique.

Notre proposition dans ce domaine est l'adoption de telles structures dans le contexte de débats en ligne. Il s'agit de permettre la définition d'un ensemble de jugements de valeurs adaptés à l'expression d'opinions dans le contexte de débats et de permettre la dé-

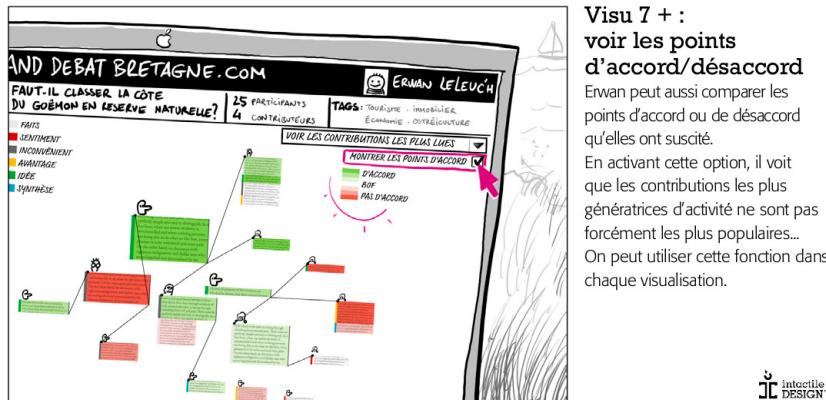


FIGURE 4.3 : Visualisation des accords et désaccords dans un graphe de commentaires, un graphe de débat

finition de relations d'oppositions entre ces jugements. Les structures ainsi définies vont ensuite nous permettre de détecter des conflits dans le discours d'un agent ou des conflits entre des agents. C'est sur ce principe que nous proposons par la suite de définir des formes d'assistance à la création d'argumentation et une analyse du réseau social basée sur des relations d'opposition ou de soutien entre utilisateurs. Ceci est illustré dans l'exemple complet déroulé en fin de ce chapitre.

La représentation des relations d'opposition entre jugements fait partie de nos extensions produites pour les vocabulaires RDF dédiés aux contenus générés par les utilisateurs.

4.2 L'annotation

Nous détaillons dans cette section nos apports propres à l'annotation. En particulier, nous présentons le modèle de l'annotation socio-sémantique. Puis nous détaillons l'expressivité que permet notre système d'annotation.

L'usage de l'annotation comme outil de débat permet de décomposer les éléments de discussions, de sectionner les commentaires des participants. Ceci permet en particulier de mieux définir les éléments d'argumentation. [Desquinabò, (2008)] propose une étude des actes de locution (ou actes locutifs) dans les débats en ligne via forum. Une partie du travail consiste à décomposer chaque commentaire en briques de bases, ou énoncés

d'argumentation. Un énoncé d'argumentation est censé ne développer qu'une seule idée, qu'un seul élément d'argumentation. En général, dans un commentaire via un outil en ligne, ou dans une discussion, chaque participation est composée d'un ensemble d'éléments d'argumentation. Plusieurs éléments s'enchaînent, un argument en soutenant un autre. L'annotation permet de sélectionner des sections dans un commentaire et d'exprimer une opinion vis à vis de cette section. Cette opération permet une décomposition des commentaires en briques de bases de l'argumentation. Nous proposons de standardiser la représentation des argumentations en reposant sur la représentation des annotations et des opinions. Une discussion est dans notre format, un graphe d'annotations liées par des relations de support et d'opposition.

4.2.1 L'annotation socio-sémantique

Nous avons présenté dans l'état de l'art sur l'annotation deux formes d'annotations, l'annotation dédiée à la discussion ou annotation discursive et l'annotation dédiée à la production de méta-donnée, notamment pour améliorer l'indexation, ou annotation sémantique.

[Lortal, 2006] présente les deux formes d'annotations : les annotations sémantiques dont le but est de représenter des méta-données pour améliorer le traitement des données par des programmes et notamment, améliorer l'indexation et les annotations discursives qui sont une forme de communication médiée par ordinateurs au même titre que les Posts dans des forums ou des blogs. L'idée de proposer un modèle commun de l'annotation est développée par [Agosti et Ferro, 2007]. Mais comme nous le présentons dans la partie état de l'art de l'annotation, nous faisons quelques critiques à ce modèle et proposons quelques améliorations. Nous proposons un modèle d'annotation permettant de réunir ces deux formes d'annotations que nous appelons annotations socio-sémantiques.

Les annotations socio-sémantiques sont à la fois sémantiques et discursives. Les utilisateurs de l'annotation pour les débats créent des annotations essentiellement pour la discussion, mais le système génère à partir des saisies utilisateurs un ensemble d'informations précisant la sémantique des annotations. La représentation des annotations dans notre système s'appuie sur l'utilisation de RDF, technologie du Web sémantique.

En particulier, les participants aux débats peuvent via les annotations socio-sémantiques laisser des commentaires libres, utiliser des tags pour catégoriser, utiliser des tags particuliers pour participer à la modération collective, et surtout, ils peuvent utiliser des tags particuliers leur permettant d'exprimer leurs accords ou désaccords. Toutes ces actions peuvent être produites lors d'annotations séparées ou lors d'une seule annotation. Toutes les informations résultant de ces annotations ainsi que celles produites par le système (comme la date de création par exemple) sont représentées sous forme de méta-données RDF afin de faciliter leur indexation et leur réutilisation.

Le terme de socio-sémantique fait référence au livre [Breslin *et al.*, 2009] et concerne l'utilisation des technologies du Web sémantique pour représenter les méta-données

structurelles produites et déduites de l'activité du Web social.

Le principe de l'annotation socio-sémantique est de proposer aux utilisateurs une forme d'annotation permettant de construire des discussions, de tagger des documents, d'être producteur des données dans le système mais permettant aussi de produire les méta-données qui permettront au service rendu par le système d'être toujours plus pertinent, en particulier en améliorant l'indexation des éléments de discussions, des documents du système. La production des méta-données par les utilisateurs n'est pas une nouvelle tâche de l'utilisateur. Les méta-données sont extraits de son activité (fréquentation, indice de participation, place dans le réseau social des utilisateurs ...), et de la saisie du contenu d'annotations par le biais d'un formulaire (titre de l'annotation, opinion exprimée, date de création ...).

4.2.2 Formes d'usage et expressivité de nos annotations

Nous présentons une modélisation de notre système d'Annotations. Ce modèle reprend le formalisme d'Agosti et Ferro en intégrant nos apports. La définition de ce modèle est motivé par le besoin de créer du lien entre les données, principal intérêt de l'annotation, qui nous semble bridé par la restriction classique qui veut qu'une annotation ne peut porter que sur une seule ressource à la fois.

Le principal intérêt du modèle d'Agosti et Ferro est qu'il permet de réunir sous un même concept les notions d'Annotation sémantique et d'Annotation Discursive.

Les principales critiques que nous adressons au modèle d'Agosti et Ferro [Agosti et Ferro, 2007] sont les suivantes :

- Les annotations portent sur des documents et non sur des ressources.
- Les annotations ne sont pas annotables à leur tour.
- Les annotations ne sont produites que par des utilisateurs humains du système.
- Une annotation ne porte que sur une seule sélection dans un document.
- Le système de streams et de segments permet de traiter des annotations dans des textes ou des flux mais ne permet pas de s'adapter à des graphiques (notamment la cartographie).

L'un des concepts fondamentaux du Web est celui d'hyperlien. Le Web est vu comme un graphe de documents se référençant les uns les autres et dont la navigation des uns aux autres est permis par un système de liens entre les ressources. La critique faite aux liens hypertextes de base du protocole HTTP et à laquelle s'attaque le Web sémantique, est qu'il est impossible d'associer une sémantique à un lien entre deux documents. L'objectif du Web sémantique est donc de créer un ensemble de ressources liées les unes aux autres par des relations fortement typées. L'annotation socio-sémantique va dans ce sens. Elle permet de créer des graphes de ressources dont les relations et les entités sont typées.

Contraindre l'annotation à ne porter que sur une seule ressource à la fois, c'est limiter les possibilités de liens entre les ressources. Dans notre modèle de l'annotation, une annotation peut porter sur plusieurs ressources et nous définissons les annotations comme

étant des ressources. Chacune des annotations produites est alors accessible par une URL, chaque annotation peut être consultée comme n'importe quelle page Web. Nos annotations portent sur des sélections de ressources. Une annotation étant une ressource, les annotations sont à leur tour annotables.

Nos annotations sont essentiellement dédiées au débat. Elles permettent d'exprimer des accords et des désaccords mais aussi permettent de créer une modération par les utilisateurs en signalant les comportements malveillants comme le spam, le troll ou le flame. Elles permettent aussi de catégoriser les contenus, les documents ou les annotations par le biais de tags. C'est sur ces principes que repose le typage de nos ressources et de nos liens. La caractérisation de l'intention de l'annotation par l'utilisation d'un tag d'opinion ou de modération permet de définir des relations qualifiées entre les annotations dans le graphe. L'étude réalisée par Julien Cotret [[Cotret et Gouaïch, 2011](#)] dans un forum en PHP pour Facebook montre que les utilisateurs adoptent parfaitement l'usage de certains tags d'opinion. En particulier, les tags "je suis d'accord" et "je ne suis pas d'accord" sont parfaitement adoptés, les commentaires associés à l'usage de ces tags d'opinions confirment en général l'opinion exprimée. Le tag "je ne comprends pas" est également bien adopté. Concernant d'autres tags, les utilisateurs les estiment soit superflus soit confus. Il a aussi été émis que les utilisateurs aimeraient pouvoir définir eux-mêmes des formes de jugements.

Enfin, l'annotation permet de sectionner les commentaires et ainsi de décomposer la structure d'un argument. Mais il est aussi possible de proposer des formulaires de saisie d'annotations ou la structure est prédéfinie. Par exemple, un formulaire utilisé lors d'expérimentations permettait de reformuler les parties sélectionnées via un premier champ de saisie de texte avant de pouvoir proposer un nouveau commentaire dans un autre champ comme montré dans la capture d'écran [4.4](#).

Une grande partie des méta-données produites est propre aux conditions de production de l'annotation et permet d'ancrer l'annotation dans un contexte temporel, humain et applicatif. La date de production est enregistrée. Mais également l'agent ou les agents producteurs de l'annotation. Aujourd'hui, de nombreuses méthodes de fouille de texte permettent de créer des annotations sémantiques. Les systèmes de recommandation créent du contenu qui peuvent alimenter des discussions. C'est dans ce contexte que nous avons pensé nos annotations. Et donc, nous permettons à un agent logiciel de créer des annotations autant qu'à un agent humain. Ceci est possible par le biais de services Web. Ainsi, une recommandation produite par un logiciel est traitée dans notre système comme une annotation. Chaque agent est aussi connu dans notre système. Une ressource permettant de représenter cet agent est créée dans le système. En particulier, nous enregistrons sa date d'inscription dans le système, si l'agent est humain ou logiciel, si d'autres ressources Web déjà existantes lui sont associées (comme la page d'une application ou la page personnelle d'un utilisateur). Enfin, nous représentons ses accointances avec d'autres agents du système, que celles-ci soient déclarées ou calculées.

Le dernier point que nous adressons avec notre modèle d'annotations est celui de l'an-

FIGURE 4.4 : Formulaire d'annotation

notation de toute ressource Web possible. Le système de streams et segments d'Agosti et Ferro est très bien adapté pour des annotations de documents de type textes ou flux. Mais sur le Web, les ressources sont multimédia et de nombreux services affichent des images. L'annotation sur une image peut porter sur des sélections de zones ou formes géométriques. Penser un modèle d'annotation suffisamment générique pour pouvoir s'adapter aussi à ce contexte demande de mettre en place un mécanisme au dessus du système de streams et de segments. Nous nous appuyons pour cela sur le type de ressources défini dans le vocabulaire NiceTag *PartOfWebRepresentation*². Toute ressource Web peut-être segmentée en ressources de types parties de ressources Web. Cela vaut pour les segments dans des flux aussi bien que pour des zones dans des images. Au moment de la création d'une annotation, nous créons donc une ressource qui est l'annotation mais nous créons aussi un ensemble de ressources qui sont les sélections sur lesquelles portent les annotations. Nous inspirant de la démarche adoptée par Diigo³, nous créons aussi une copie de la ressource annotée. Ainsi, même si la ressource annotée évolue, nous gardons une trace de l'annotation dans son ensemble. Nous pouvons présenter à l'utilisateur pour toute ressource la version à la date où elle a été annotée.

-
2. <http://ns.inria.fr/nicetag/2010/09/09/voc.html#PartOfWebRepresentation>
 3. <http://www.diigo.com/>

4.2.3 Définition formelle

Nous allons ici reprendre le formalisme ensembliste d'Agosti et Ferro pour présenter notre modèle de l'annotation socio-sémantique.

Tout d'abord, nous partons de l'ensemble des ressources du Web et plus largement de l'ensemble des ressources possibles. Ensuite, l'ensemble des annotations est un sous-ensemble particulier de l'ensemble des ressources.

- R est un ensemble de Ressources et $r \in R$ est une ressource générique. U_R est l'ensemble universel des ressources et donc $R \subseteq U_R$
- A est un ensemble d'annotations et $a \in A$ est une annotation générique. U_A est l'ensemble universel des annotations et donc $A \subseteq U_A$
- On a aussi $U_A \subset U_R$ et si $A \subseteq R$ alors $a \in R$.

Ressource.

Le concept de Ressource dans le cadre du Web est une notion de haut niveau regroupant à la fois des documents, des lieux, des personnes, des services ... tout ce qui peut avoir une représentation accessible par le protocole HTTP et à ce titre, se voir affecter une URI (un identifiant unique). A titre d'exemple, Tim Berners Lee est une personne. Il est possible de créer un document décrivant qui est Tim Berners Lee. Tim Berners Lee est donc une ressource. Cependant, la description de Tim Berners Lee n'est pas Tim Berners Lee, c'est une autre ressource. Donc Tim Berners Lee peut avoir une URI et le document de description de Tim Berners Lee aura une autre URI que la ressource Tim Berners Lee lui-même. Au niveau du protocole HTTP, un appel HTTP GET sur l'URI de Tim Berners Lee ne permettra pas de récupérer Tim Berners Lee lui-même par le réseau. L'appel HTTP GET renverra comme réponse une redirection vers l'URI de la ressource décrivant Tim Berners Lee. Utiliser le concept de Ressources, concept plus générique que celui de Document, permet d'envisager des usages de l'annotation portant sur d'autres éléments que des Documents.

Document.

Un Document est un type de Ressource. Dans le cadre de notre implémentation, nous ne considérons que des Documents accessibles via HTTP et écrits en HTML.

Annotation.

Une annotation est un type de Ressource. Une annotation est la combinaison d'un ensemble de sélections sur des Ressources (notamment sur des Documents donc) et de données produites par les utilisateurs comme un Tag, un Jugement, un Commentaire ... finalement, d'autres Ressources.

Nous ne détaillerons pas l'aspect temporel comme dans le modèle de Agosti et Ferro car cette partie est triviale. Nous soulignerons juste que dans notre modèle, il est tout à fait possible de modifier une annotation afin que celle-ci annote désormais une ressource qui n'existe pas avant la première version de l'annotation en question. Ceci permet par exemple de prévoir une annotation pour toutes les ressources qui me déplaisent et de la modifier à chaque fois que je rencontre une nouvelle ressource qui me déplaît.

Etant donné que nous parlons de ressources Web, chacune de ces ressources est et doit être identifiée de façon unique par une URI.

- I est un ensemble d'identifiants tel que $|I| = |R|$ et $i \in I$ est un identifiant. U_I est l'ensemble universel des identifiants tel que $|U_I| = |U_R|$, $I \subseteq U_I$. Une fonction bijective h est définie de U_I dans U_R , $h : U_I \rightarrow U_R$. Cette fonction h permet d'associer un identifiant unique à une ressource : $\forall r \in U_R, \exists! i \in U_I \mid h(i) = r \Rightarrow h^{-1}(r) = i$.

Afin de lier une annotation aux ressources qu'elle annote, nous allons devoir préciser l'existence d'un sous-ensemble particulier de ressources que sont les parties de ressources. Nous définissons cet ensemble comme suit :

- PR est un ensemble de parties de ressources et $pr \in PR$ est une partie de ressource. U_{PR} est l'ensemble universel des parties de ressources et donc $PR \subseteq U_{PR}$.
- On définit aussi une fonction s qui permet d'associer à chaque partie de ressource la ressource dont elle est issue. $s : U_{PR} \rightarrow U_R \mid s(pr) = r$.
- La fonction inverse de s renvoie pour une ressource, l'ensemble des parties de cette ressource existant dans le système. Cet ensemble peut-être vide.

Typiquement, les segments de Agosti et Ferro sont des cas particuliers de parties de ressources. Donc l'ensemble des segments est un sous-ensemble de l'ensemble des parties de ressources.

Dans le modèle d'Agosti et Ferro, le corps d'une annotation, ou le contenu saisi par l'auteur (ou les auteurs) d'une annotation, est composé d'un ensemble de signes. La sémantique de l'annotation est alors constituée de la composition des sémantiques des signes. Dans notre modèle, les signes sont, là encore, des ressources. Ces ressources peuvent être des *Posts* ou des *Tags* (et leurs extensions) dans l'état actuel de l'implémentation de notre système. Mais à terme, nous pourrions ajouter tout type de ressources comme des vidéos, des sons ou des images avec leurs sémantiques propres. Il faut aussi souligner que ces ressources peuvent être créées au moment de la création de l'annotation (le cas le plus souvent des Posts) mais peuvent être pré-existantes (le cas le plus souvent des Tags). La sémantique de l'annotation dans notre modèle est alors partiellement définie par la composition des sémantiques des différents signes constituant le corps de l'annotation. Partiellement uniquement car, comme défendu par [Monnin et al., 2010], l'action d'annotation, tout comme l'action de tagging, relie des ressources ayant chacune leur sémantique propre mais l'action elle-même a une signification. La sémantique d'une action d'annotation est donc définie par la composition d'un choix de formulaire d'annotation et par le contenu de l'annotation. Actuellement, la définition d'un nouveau type d'annotation est réservée

à l'administrateur du système. Nous définissons l'ensemble des types d'annotation de la façon suivante :

- TA est un ensemble de types d'annotation et $ta \in \text{TA}$ est un type d'annotation. U_{TA} est l'ensemble universel des types d'annotation et donc $\text{TA} \subseteq U_{\text{TA}}$

Il reste à préciser qu'une annotation peut avoir plusieurs auteurs et que les auteurs sont des agents logiciels ou humains (utilisateurs du système). Nous définirons donc simplement l'ensemble des agents de la façon suivante :

- AG est un ensemble d'agents et $ag \in \text{AG}$ est un agent. U_{AG} est l'ensemble universel des agents et donc $\text{AG} \subseteq U_{\text{AG}}$

Une annotation est dans notre système composée comme suit : c'est un choix de type d'annotation puis l'association d'un ensemble de parties de ressources (ou ressources annotées) à un ensemble de ressources (posts, tags ... saisies par l'utilisateur).

- Une annotation $a \in A(k)$ est un tuple : $a = (i_a \in I(k), AU_a \subseteq AG(k-1), R_a \subseteq R(k), ta_a \in TA(k), PR_a \subseteq PR(k))$
- i_a est l'identifiant unique de l'annotation
- AU_a est l'ensemble d'agents auteurs de l'annotation
- R_a est l'ensemble des ressources constituant le corps de l'annotation
- ta_a est le type d'annotation choisi
- PR_a est l'ensemble des parties de ressources annotées par l'annotation

A partir de là, une discussion est constituée d'un ensemble d'annotations liées entre elles.

Pour synthétiser :

- Notre modèle s'appuie sur le concept de Ressource comme concept le plus générique. Bien que pour l'instant dans notre implémentation, les Ressource annotables au départ ne sont que des Documents HTML, il nous semble que, au vu des évolutions des fonctionnalités des navigateurs Web vis à vis des contenus multimédia, considérer que les annotations portent sur des Ressources et non seulement sur des Documents HTML permet d'envisager l'intégration de futures formes d'annotations comme des annotations sur des images, des sons ou des vidéos.
- Les Annotations portent sur plusieurs Ressources. Nous sortons de la conception classique de l'Annotation qui ne porte que sur une seule Ressource à la fois. Le premier intérêt de l'annotation est de créer des liens entre des Ressources. Restreindre les annotations à ne porter que sur une seule Ressource à la fois c'est limiter le potentiel de liens que l'on peut créer.
- Le corps des Annotations est constitué d'un ensemble de Ressources. Dans la pratique, nous avons observé que les utilisateurs ne saisissent pas toujours un commentaire dans leurs formulaires d'annotations. Certains utilisent l'annotation uniquement pour regrouper des sélections, ou ne saisissent qu'un mot, un smiley ... Nous avons encouragé cette pratique en permettant de ne saisir qu'un jugement par exemple pour exprimer son accord ou son désaccord sur un ensemble de sélections. A nouveau, considérer que le corps d'une annotation est constitué d'un ensemble de Ressources permet de représenter l'annotation telle quelle est implémentée dans

notre système, nos formulaires permettent d'associer des Tags, des Jugements et des Commentaires par exemple, mais également d'anticiper sur de nouvelles fonctionnalités comme la possibilité d'associer une Photo ou un Son.

- Les Annotations sont elles-mêmes des Ressources. Ceci permet de rendre une annotation annotable à son tour.

4.3 Les réseaux sociaux

4.3.1 Réseau social calculé à partir d'interactions

Le concept de communauté d'intérêts est au centre d'Internet depuis sa fondation. Dans le cadre des débats 2.0, nous parlerons de communautés d'opinions pour représenter des groupes d'individus partageant une opinion collective, une opinion qu'ils auront construit à plusieurs par leurs interactions communicationnelles. Un réseau social est la définition d'un ensemble de relations entre individus. La communication est au centre de la constitution des réseaux en ligne. Les deux approches classiques sont : soit de permettre aux utilisateurs de déclarer eux-mêmes leurs relations les uns par rapport aux autres, soit d'analyser leurs communications pour en déduire des relations.

Les outils qualifiés de réseaux sociaux ou encore de SocNet (pour Social Network) sont des outils privilégiant la déclaration d'un réseau par les individus. L'application la plus connue dans ce domaine est Facebook. L'utilisateur, après s'être créé un compte, peut déclarer parmi les autres utilisateurs du système lesquels il connaît. La relation ainsi définie dans l'application Facebook prend le label d'"ami". Déclarer une relation dans Facebook, c'est déclarer un ami. Ce principe est le même dans tous les SocNet, seul le label d'"ami" peut changer. Dans un SocNet comme LinkedIn se voulant professionnel, la relation est celle de "collègue". La différence fondamentale entre Facebook et les outils de SocNet pré-existants est que Facebook s'est directement placé comme un outil de prolongement des contacts réels entre les utilisateurs. Dans Facebook, l'utilisateur est censé déclarer des relations avec des individus qu'il a déjà rencontré en Face à Face. Les SocNet comme Friendster ou MySpace, présents sur le Web avant Facebook avaient plutôt joué le jeu de la rencontre. L'objectif était de permettre à des gens ne se connaissant pas d'entrer en contact en fonction d'intérêts communs comme les goûts musicaux ou les jeux vidéos par exemple. On retrouve ce principe de recréer un réseau social existant dans pratiquement tous les SocNets récents créés après Facebook.

Cependant, dans les sciences sociales, la relation entre deux individus est généralement définie comme étant une somme d'interactions. Et cette relation peut-être, selon les théories, qualifiée (amitié, relation professionnelle, relation familiale ...) et quantifiée (on parle d'intensité). Les interactions entre les individus sont des actions d'un individu vers l'autre et les actes communicationnels entre individus ont donc une place importante dans la liste des interactions permettant de définir une relation. Aujourd'hui, Internet est un ou-

til de communication incontournable. Il est donc naturel d'étudier les échanges communicationnels médiés par Internet pour proposer d'en déduire des relations entre individus. Cette approche a été adoptée dès l'invention des premiers outils de communication via Internet. Les premiers réseaux d'ordinateurs étaient déjà vus comme des réseaux sociaux [Wellman *et al.*, 1996]. Il existe de nombreuses références sur l'analyse de réseaux sociaux déduits à partir d'outils de discussion. Pour ne citer qu'eux : [Rheingold, 1993], [Revillard, 2000] étudient le réseau social de Usenet, [Tyler *et al.*, 2005] analyse des échanges de mail et [Mutton, 2004] des discussions par messagerie instantanée.

Dans notre approche, nous considérons l'utilisation d'un outil de discussion, l'annotation socio-sémantique et conjointement un réseau déclaré par les utilisateurs (à partir de données fournies dans le système ou à partir de données issues d'un ScoNet comme Facebook). A partir de l'analyse des interactions et d'un réseau déjà déclaré, nous proposons de calculer un réseau où les relations sont qualifiées et quantifiées. Notre réseau est essentiellement structuré par les relations d'opposition et de soutien. Et ces relations sont quantifiées en fonction de la fréquence et de la quantité des oppositions et soutiens exprimés par les annotations.

4.3.2 Identité numérique

Fanny Georges développe un modèle de l'identité numérique où l'identité est décomposée en ce qui est déclaré, ce qui est "fait-notifié", ce qui est calculé. Ce modèle permet notamment de dire d'un système s'il met plus l'accent sur l'activité de l'utilisateur ou sur sa façon de se présenter. L'étude comparée de Facebook et MySpace [Georges *et al.*, 2009] montre clairement cette différence. Facebook remonte énormément d'indicateurs de l'activité dans le réseau tandis que MySpace se concentre plus sur un espace personnel très personnalisable. Notre point de vue sur ce point est que globalement, sur le Web, l'activité devient de plus en plus importante. Ce qui est important n'est plus la façon de se présenter mais la façon d'agir. Les systèmes de recommandation, les moteurs de recherche aujourd'hui incluent de plus en plus des indicateurs de l'activité dans leurs calculs comme la nouveauté, la fréquence de visite, le nombre de liens.

Nous proposons d'adopter la même terminologie que Fanny Georges en l'adaptant aux réseaux. Ce prolongement nous semble naturel. Il y a aujourd'hui une part du réseau qui est déclaré, l'ensemble des relations que l'on déclare entretenir, comme par exemple ses amis Facebook. Les outils phares qualifiés de réseaux sociaux en ligne présentent une vision du réseau essentiellement déclarative. Mais il y a aussi une part du réseau qui est déduite de l'agissant. On parlera de réseau agissant ou réseau déduit de l'activité. Par exemple, lorsque un réseau est déduit à partir de discussions dans un forum, un blog, la musique que l'on écoute, la mise en relation de personnes jouant à un même jeu... Enfin, l'analyse de réseaux sociaux repose essentiellement sur la définition et l'analyse d'indicateurs numériques de l'état d'un réseau. L'analyse des réseaux sociaux est la partie de réseau calculé qui prolonge l'identité calculée du modèle de Fanny Georges.

4.3.3 Déduction d'un réseau à partir de l'annotation

Nous proposons d'utiliser les données produites par les débats en ligne pour construire un réseau. Les données produites par les débats constituent en elles-mêmes un réseau d'interaction. Les utilisateurs sont reliés par le biais de leurs annotations se répondant les uns les autres. Ils sont aussi liés par le fait qu'ils consultent les mêmes ressources, qu'ils utilisent les mêmes tags de catégorisation ... A partir de ces interactions, nous définissons des fonctions de calcul prenant en compte l'intensité des échanges. Nous nous inspirons de la théorie de Marc Granovetter [Granovetter, 1973] pour définir des relations dont l'intensité, la force dépend de la fréquence des échanges, de leur longueur, de leur durée dans le temps. La temporalité est un élément important du calcul des réseaux. Un réseau évolue avec le temps. Nous proposons des éléments de mesure ancrés dans le temps permettant d'analyser la dynamique du réseau.

Les discussions dans un outil de débat 2.0 étant fortement structurées par l'expression d'accords ou de désaccords, nous proposons de déduire de celles-ci des relations pour le réseau social des citoyens. Nous avons brièvement présenté cette idée dans [Seilles et al., 2010b]. Nous définissons notamment des relations de supports et d'oppositions entre citoyens, à partir desquelles il est possible de calculer des communautés d'opinions. Nous proposons donc de représenter des relations entre utilisateurs, extensions de la relation "Knows" de FOAF en utilisant des relations "Supports" et "Attacks". Ces relations sont déduites des échanges entre utilisateurs dans les discussions. Nous proposons également une catégorisation des utilisateurs par le biais de classes ou types d'utilisateurs. Ces classes sont définies de façon empirique et reposent sur l'observation de comportements utilisateurs. Nous définissons donc une liste d'indices sur lesquels nous basons la définition de nos classes.

La liste d'indices définis est la suivante :

- **Activité Connexion** : fréquence de connexion au système.
- **Activité Création de contenu** : fréquence de création de contenu (annotations de toutes formes).
- **Activité de Lecture** : quantité de ressources consultées et temps de lecture moyen de chaque ressource.
- **Taux de Flame** : moyenne du nombre d'utilisateurs jugeant les annotations produites par l'utilisateur considérées comme étant du Flame (propos dérangeants).
- **Taux de Spam** : moyenne du nombre d'utilisateurs jugeant les annotations produites par l'utilisateur considérées comme étant du Spam (répétition de propos déjà tenus ailleurs).
- **Taux de Troll** : moyenne du nombre d'utilisateurs jugeant les annotations produites par l'utilisateur considérées comme étant du Troll (tentative de détournement de la conversation).
- **Taux d'Accord généré** : moyenne du nombre d'utilisateurs signalant leur soutien sur des annotations produites par l'utilisateur considéré.

- **Taux de Désaccord généré** : moyenne du nombre d'utilisateurs signalant leur opposition sur des annotations produites par l'utilisateur considéré.
- **Taux d'Accord personnel** : nombre d'accords exprimés par l'utilisateur sur des annotations qui ne sont pas de lui.
- **Taux de Désaccord personnel** : nombre de désaccords exprimés par l'utilisateur sur des annotations qui ne sont pas de lui.
- **Taux de jugements exprimés** : nombre d'annotations exprimant un jugement par rapport au nombre d'annotations produites par l'utilisateur.
- **Taux de modération personnel** : nombre d'actions de modération produites par l'utilisateur (déclaration de flame, troll ou spam).
- **Taux de lecture de l'utilisateur** : nombre de ressources consultées par l'utilisateur sur le nombre de ressources existantes dans le système.
- **Quantité d'écriture** : longueur moyenne de texte saisi par l'utilisateur par annotation produites.
- **Quantité d'annotations produites** : nombre d'annotations produites par l'utilisateur.
- **Taux de tag** : nombre de tags utilisés par l'utilisateur et combien de fois chaque tag est utilisé.

Nous proposons ensuite la liste de Classes suivantes, sachant qu'un utilisateur peut très bien être dans plusieurs classes. Les noms des classes sont choisis de façon empirique :

- **Actif** : les taux d'activités sont élevés.
- **Inactif** : les taux d'activités ne sont pas élevés.
- **Annotateur** : l'utilisateur a créé beaucoup d'annotations.
- **Lecteur** : l'utilisateur a lu beaucoup de ressources et passe beaucoup de temps à lire des ressources.
- **Modérateur** : l'utilisateur a effectué beaucoup d'actions de modération. Il a un taux de modération propre élevé.
- **Flameur** : l'utilisateur a un taux de flame élevé.
- **Trolleur** : l'utilisateur a un taux de Troll élevé.
- **Spameur** : l'utilisateur a un taux de spam élevé.
- **Charismatique** : les ressources produites par l'utilisateur sont beaucoup lues.
- **Leader** : le taux d'accord généré de l'utilisateur est élevé.
- **Polémiste** : le taux de désaccord généré par l'utilisateur est élevé.
- **Contradicteur** : le taux de désaccord propre de l'utilisateur est élevé.
- **Supporteur** : le taux d'accord propre de l'utilisateur est élevé.
- **Électrique** : l'utilisateur utilise beaucoup de tags différents.
- **Spécialiste** : l'utilisateur utilise peu de tags différents.
- **Expressif** : l'utilisateur s'exprime beaucoup par saisie de texte libre.
- **Impulsif** : l'utilisateur s'exprime peu par saisie de texte libre.

4.3.4 Classification

Nous proposons dans cette partie une classification des données produites par les débats en ligne. Cette classification permet de mieux situer nos apports et se base sur la terminologie proposée par Fanny Georges [Georges, 2007].

Les données que nous présentons dans la partie précédente, les données que l'on a besoin de représenter pour des débats en ligne concernent essentiellement la production de contenu par des individus, et nous considérons que l'un des objectifs des débats est d'aider les individus à mieux argumenter notamment en se regroupant ou en se confrontant à d'autres individus.

Nous proposons de classer les données selon la même terminologie que proposée par Fanny Georges. Les données déclarées concernent les profils d'utilisateurs tels qu'ils sont saisis lors de l'inscription. Sur cette catégorie de données, l'état de l'art des vocabulaires RDF dédiés à la représentation de profils utilisateurs et de réseau social déclaré par l'utilisateur est déjà suffisant. Nous représenterons un profil utilisateur en se basant sur le vocabulaire FOAF. Nous proposerons par contre des extensions de FOAF dès qu'il s'agit de représenter des relations entre utilisateurs. FOAF ne définit qu'une relation dans ce cas, la relation *know*. La définition de relations étendant la relation *know* est une pratique courante et recommandée.

Les posts, ou commentaires ou encore arguments émis par les utilisateurs et de façon plus large toutes les données produites par les utilisateurs lors de l'usage du système (tagging, discussion) sont des données produites par l'action, l'agissant. Pour les représenter, nous reposons essentiellement sur les vocabulaires SIOC et Annotea que nous couplons à Nicetag et MOAT. C'est sur ces données que nous proposons le plus de nouveautés. Nous étendons la classe *Tag* de MOAT pour définir des classes *Emotion* et *Jugement*. Nous proposons une utilisation conjointe de Annotea et SIOC qui remet en question le statut de corps d'une annotation. Une annotation est-elle un Post ou un Post est uniquement une partie du corps d'une annotation ? Nous avons décidé dans notre modèle que les Posts comme les Tags sont des éléments constitutants le corps d'une annotation. Nous proposons là aussi des extensions à la classe Post afin de typer les posts selon leurs objectifs. Par exemple, dans la version distribuée d'Argumentea, notre outil d'annotation, nous définissons les classes *Commentaire* et *Reformulation*.

Enfin, la liste d'indices précédemment cités résulte de calculs effectués par le système. Ces données calculées peuvent être représentées en reposant sur l'usage du vocabulaire SemSNA et en étendant la classe *SNAIndice*.

Les classes d'utilisateurs précédemment proposées reposant sur un calcul du système, elles devraient être qualifiées de données calculées. Cependant, une classe, si elle est acceptée par un utilisateur et que celui-ci modifie son comportement en conséquence (pour changer de classe ou au contraire, renforcer sa position dans une classe), peut basculer dans le statut de rôle et à ce moment là être considérée comme une information déclarée par l'utilisateur.

4.4 Interopérabilité

Dans la vision Web 2.0, les applications sont présentées comme des services. Ces services peuvent être construits à partir de données qui leur sont propres mais aussi à partir de données échangées avec d'autres services. La notion de mashup ou applications composites est présente dans la définition du Web 2.0. Publier ses données et adopter un format clair permet de simplifier la réutilisation des données et par là même, contribue à assurer l'usage du service. Les données peuvent être exploitées de façon non prévues par le service initial, voire être croisées avec d'autres données pour être enrichies et permettre la création de nouveaux services.

La définition de formats de données clairs et servant de normes permet de contribuer à rendre des applications interopérables. L'interopérabilité est l'un des facteurs facilitant l'accès à un service. L'ouverture des données de Facebook a contribué à son succès. Les utilisateurs-développeurs ont pu créer de petites applications enrichissant l'usage de Facebook et petit à petit un véritable marché des applications gravitant autour de Facebook s'est créé. Aujourd'hui, certaines sociétés de développement de jeux ont adopté Facebook comme plateforme et ne distribuent leurs produits que par ce biais.

Si l'on reprend l'histoire d'Internet, la constitution d'Internet tel qu'on le connaît aujourd'hui n'a été rendu possible que par l'invention de protocoles définissant des formats de requêtes et de réponses à des requêtes que les fabricants de machines ou les développeurs doivent adopter pour permettre l'interopérabilité. Le premier protocole est le protocole TCP/IP qui permet à des ordinateurs d'être en réseaux quelque-soit le constructeur des machines. Ensuite, le protocole mail a permis de construire un premier outil de communication. Il en va de même pour IRC et la messagerie instantanée et pour le Web basé sur le protocole HTTP. Aujourd'hui, le Web est une plateforme d'applications. De nombreux langages permettent de définir des formats de données permettant aux applications du Web d'échanger leurs données. Parmi eux, RDF et XML portés par le W3C. Si l'on veut que les débats 2.0 s'inscrivent dans de véritables processus démocratiques, il faut adopter des formats de données qui faciliteront l'interopérabilité entre outils de débats mais aussi entre les outils de débats et d'autres applications comme des outils de sondages, des outils d'aménagement du territoire, des sites institutionnels ...

Les données produites par notre outil de débat concernent les discussions et le réseau social. Un format de données ne valant que s'il est adopté par d'autres, nous avons opté pour l'utilisation conjointe de formats déjà existants. Nous avons opté pour l'utilisation de RDF dont l'un des avantages est de permettre la définition de nouveaux formats et des utilisations conjointes de ces formats. Nous avons donc listé les vocabulaires RDF existants permettant de répondre à nos attentes et adopté des ponts entre ces vocabulaires permettant de les utiliser de façon conjointe. Nous proposons une utilisation conjointe de formats existants et d'extensions pour couvrir les besoins concernant les débats.

En particulier, nous adoptons le vocabulaire Annotea pour représenter les annotations de notre système. Le contenu de ces annotations, ou corps des annotations est représenté

par une utilisation conjointe de SIOC, MOAT, NICETAG, RSS et DC. Les données propres aux utilisateurs sont quand à elles représentées essentiellement via l'utilisation du vocabulaire FOAF. Et les données de structuration du réseau déduites des interactions ou déclarées sont représentées par l'utilisation conjointe de FOAF et SemSNA. [Ces différents vocabulaires sont présentés dans la section traitant de la représentation des données de la discussion, du réseau social et des débats.](#)

Nous utilisons le vocabulaire RDF Annotea annotation-schema pour représenter une annotation. Ce vocabulaire définit la classe *Annotation* et les relations *body* et *annotates*. La relation *body* permet de lier l'Annotation à une ressource qui constitue le corps de l'annotation. Dans notre cas, une Annotation est liée à autant de ressources que définies par son formulaire de saisie. Par exemple, une annotation peut-être liée à un tag d'opinion, et à plusieurs champs textes. La relation *annotates* permet de lier l'annotation à une ressource ou portion de ressource annotée. Dans notre cas, toutes les ressources annotables sont des pages HTML et les portions sont définies par des XPointer. Là encore, une annotation peut être liée à plusieurs ressources annotées.

Le contenu d'une Annotation dans notre système est composé de commentaires et de tags utilisés conjointement ou séparément. Ces commentaires et tags ont des statuts particuliers. Nous détaillons ici les solutions de représentation adoptés.

4.4.1 Représenter les données de la discussion

Cette partie décrit les méta-données que l'on a besoin de représenter concernant les données des débats 2.0. Nous présenterons d'abord les méta-données concernant les Annotations, éléments de discussion de base. Puis nous présenterons les méta-données concernant la discussion dans son ensemble et enfin, les méta-données concernant le réseau social.

Le contenu des annotations peut être constitué de Posts mais aussi de Tags, et de façon générale, de Ressources. Les messages, ou Posts, dans notre application de débats 2.0 sont typés. Un Post peut-être un simple commentaire mais il peut aussi être une reformulation. Il est aussi possible d'entrevoir d'autres types de Posts comme des Questions et des Réponses. Nous ajoutons donc cette méta-donnée concernant le type des Posts aux classiques informations que sont la date et l'auteur d'un Post. Concrètement, cela revient à proposer des classes d'extension de la classe *SIOC:Post*.

Si les Posts sont typés, les annotations et les tags aussi. Cette information est fixée par le formulaire utilisé lors de la saisie de l'annotation. Par exemple, pour les tags, nous définissons les types "*jugement*" pour exprimer une position (d'accord, pas d'accord, c'est bien, c'est mal ...), "*domaine*" pour catégoriser une annotation (économie, sport, sciences ...) et "*émotion*" pour exprimer une émotion (je suis content, ça m'énerve ...). Le tagging dans notre application est pensé pour servir plusieurs fonctionnalités. Il a déjà été constaté que la pratique du tagging servait des propos différents selon les utilisateurs [[Golder et Hberman, 2006](#)]. En particulier, l'un des usages constaté est de juger la ressource, de définir

un rapport entre la personne et la ressource exprimant une appréciation personnelle de la ressource (c'est intéressant, c'est nul ...). Chaque action de tagging est un acte de langage [Monnin et al., 2010] qui exprime plus qu'un simple lien entre une ressource et un signe. Nous proposons également d'utiliser le tagging afin de signaler les comportements malveillants. Le tagging est aussi utilisé comme une solution pour la modération de la discussion par les utilisateurs (ceci est une insulte, ceci est un spam ...). Lorsqu'un administrateur crée un formulaire d'annotation, il crée en fait un nouveau type d'annotation et définit la structure du contenu de ce type d'annotation. Par exemple, l'annotation proposée par défaut pour exprimer un accord dans notre application est constitué d'un tag de type *Jugement* exprimant l'accord (je suis d'accord) et d'un champ texte ou Post de type *Justification*. Il est possible de définir des types d'annotation bien plus complexes avec plusieurs types de Tags et plusieurs types de Posts.

La représentation d'une discussion se fait par la représentation des liens entre les ressources. Cette annotation annote cette ressource. Ce Post répond à celui-ci. Cette ressource fait référence à ce tag. Là encore, les relations "Annotates", "RepliesTo" et "RefersTo" sont étendues. Une annotation va par exemple supporter les passages annotés, ou les critiquer. Un point essentiel à représenter dans une discussion faite par le biais d'annotation est l'ensemble des passages sélectionnés sur lesquels porte précisément l'annotation. Classiquement l'annotation est liée à la ressource annotée et le passage sélectionné est précisé par un système de streams et de segments. Nous proposons de reprendre le concept de NiceTag de PartOfWebRepresentation pour faire de chaque sélection une ressource à part entière. Nos annotations portent donc sur des portions de ressources du Web. Un segment est donc un type de *PartOfWebRepresentation* tout comme pourrait l'être une sélection dans une image. Il faut également souligner que l'annotation ne définit pas uniquement une relation d'une ressource annotée vers une ressource constituant le corps de l'annotation. Nos annotations créent un lien entre un ensemble de ressources annotées et un ensemble de ressources constituant le corps de l'annotation.

Le vocabulaire SIOC définit une classe particulière pour représenter les commentaires utilisateurs dans les blogs ou forum, la classe *Post*. Cette classe est étendue dans d'autres vocabulaires pour représenter des vidéos, des pages de wikis ...⁴ En particulier, certains forums sont structurés par les statuts de question et de réponse entre les Posts. Nous permettons dans notre système au moment de la création de formulaire d'annotation, de définir de nouveaux statuts de commentaires. Ces statuts sont des classes étendant la classe *Post*. En particulier, de base, nous avons proposé un statut de Reformulation et un statut de Définition. Ces statuts représentent des saisies libres ayant une intentionnalité particulière. D'autres informations sur un Post sont représentées en utilisant le vocabulaire RSS et le vocabulaire Dublin Core, en particulier la date de production, le contenu du Post, le titre du post ...

Un *Post* peut-être relié à un autre par les relations *repliesTo* et *replyOf*. Ces relations

4. extension types du vocabulaire SIOC : <http://rdfs.org/sioc/types>

ne définissent qu'une sémantique de temporalité. La réponse à un Post est en fait un Post associé à ce Post par la suite. Ce n'est pas la sémantique de réponse comme entre une question et une réponse par exemple. Ces relations peuvent être étendues en fonction des besoins. Par exemple, on pourrait définir une relation d'opposition ou de soutien entre un Post et un autre.

Systèmes de tagging

Le tagging permet avant tout de catégoriser, de lier une ressource à une catégorie ou domaine. Cependant, il existe d'autres formes d'utilisation du tagging. Nous avons adopté un usage du tagging permettant aux utilisateurs d'exprimer des opinions, des jugements de valeurs mais aussi de signaler des contenus à modérer.

Les vocabulaires pour représenter des tags sont nombreux. Nous avons fait le choix d'adopter MOAT pour représenter un tag. MOAT a la particularité de permettre de lier une ressource à un tag afin de signaler que la ressource définit le tag. Nous étendons ensuite la classe *Tag* de MOAT. Les tags utilisés pour exprimer une opinion dans notre système sont des tags *Jugement*. L'expression d'une émotion se fait par des tags de la classe *Emotion* et la catégorisation par des tags de la classe *Domaine*.

Cependant, MOAT comme de nombreux autres vocabulaires, s'intéresse à représenter les tags et non l'action de tagging. Tagger une ressource avec le pseudonyme d'un utilisateur signifie-t-il que cet utilisateur est l'auteur de la ressource, que la ressource le concerne, qu'il devrait consulter cette ressource ... Le vocabulaire NiceTag est dédié à cette tâche de représentation des actions de tagging. Il permet de représenter sous la forme d'un noeud dans le graphe RDF l'association d'un tag à une ressource. La relation *NiceTag :isRelatedTo* qui permet de lier un tag et une ressource peut-être étendue par des relations précisant l'intentionnalité de l'action de tagging.

Utilisation conjointe de formats existants et extensions

Dans notre système, les annotations ont un statut, elles sont reliées aux ressources qu'elles annotent par des relations précisant la position de support ou d'opposition et elles sont reliées à des ressources qui constituent le corps de l'annotation, ce que l'utilisateur crée comme contenu au moment de la saisie. Les annotations sont représentées en utilisant le vocabulaire Annotea annotation schéma et des extensions de la classe *Annotation* pour chaque statut d'annotation possible. La relation *annotates* est également étendue par des relations précisant le support ou l'opposition. Cependant, de nouveaux statuts d'annotation et de nouveaux tags d'opinion peuvent être définis par les utilisateurs du système. A chaque fois, l'une de ces opérations met à jour le vocabulaire RDF du système en créant une nouvelle extension d'Annotea annotation schema.

Le contenu des annotations est pour l'instant limité à des commentaires en saisie libre ou l'usage de tags. Les tags et le tagging sont représentés par une utilisation conjointe de

MOAT et NiceTag. Les commentaires peuvent avoir des statuts particuliers, comme la reformulation, l'expression d'une question ... Chaque nouveau statut est représenté par une classe étendant la classe *Post* de SIOC.

Cependant, il est maintenant possible de discuter certaines formes d'usage. Par exemple, lorsque un nouveau commentaire est associé par le biais d'une annotation à l'intégralité d'un commentaire déjà présent. Cette action revient exactement à créer une réponse à un Post dans un forum. Doit-on alors représenter la relation entre les deux Posts par le biais d'une annotation ou par le biais de la relation *replies* de SIOC? De même, lorsque un tag est associé à une ressource, doit-on représenter cette relation par le biais d'une annotation ou doit-on utiliser la relation *isRelatedTo* de NiceTag? Dans ces deux cas, nous avons opté pour la représentation en passant par une annotation. Notre position est que les actions de tagging ou de réponses dans des forums sont des actions déjà définies par l'action d'annotation.

4.4.2 Représentation des réseaux sociaux

Le vocabulaire FOAF permet de représenter un réseau social. Il définit une classe *Person* pour représenter les individus et une relation *knows* extensible qui permet de représenter un lien de connaissance entre deux individus. Le vocabulaire FOAF a fait l'objet de plusieurs extensions permettant de représenter une relation de parenté, une relation de collègue de travail ... Le vocabulaire SIOC définit une classe *UserAccount* qui est liée à la classe *Person* de FOAF. Un *UserAccount* représente le compte d'un utilisateur dans une application et celui-ci est lié à une *FOAF:Person*. Ainsi, par le biais de l'utilisation conjointe de FOAF et SIOC, on peut représenter un réseau social et l'activité des membres de ce réseau. Autrement dit, l'utilisation conjointe de FOAF et SIOC permet de représenter des réseaux sociaux déclarés et des réseaux d'interaction, ou réseaux centrés objets. Dans un réseau centré objet, les membres du réseau ne sont plus liés directement mais par le biais d'objets intermédiaires comme par exemple des éléments de discussion, ou pour nous, des annotations.

Cependant, il reste à représenter les résultats de l'analyse du réseau social. Que l'on cherche à déduire des relations à partir d'interactions ou que l'on cherche à caractériser la structure du réseau, les membres ayant une position particulière dans ce réseau, sa densité ... Le travail de Guillaume Eretéo sur l'ontologie SemSNA [Erétéo, 2011] vient parfaitement compléter les vocabulaires FOAF et SIOC dans ce sens. SemSNA permet de représenter des degrés entrants ou sortants et des indices de centralité en fonction d'un type de relation dans le réseau. Par exemple, il est possible de représenter le degré entrant calculé des relations familiales et le degré entrant des relations de travail pour un même individu dans un réseau multi-relationnel. En complément, Guillaume Eretéo a proposé des extensions pour FOAF permettant de représenter par exemple les différentes relations de parentés comme des relations étendant une relation familiale. Nous adoptons la même démarche pour représenter les relations de débats. Une relation débat étendue par une relation sup-

port et une relation opposition.

Nous proposons donc de représenter les données du réseau social en utilisant conjointement ces trois vocabulaires RDF. FOAF est dédié à la représentation du réseau déclaré par les utilisateurs et à la représentation du réseau déduit de l'activité. Les données de l'activité ou réseau agissant sont représentées par SIOC conjointement avec Annotea, NiceTag, MOAT ... Et enfin le réseau calculé, ou les données issues de l'analyse de réseau social est quand à lui représenté en utilisant le vocabulaire SemSNA. La figure 4.5 synthétise cette proposition.

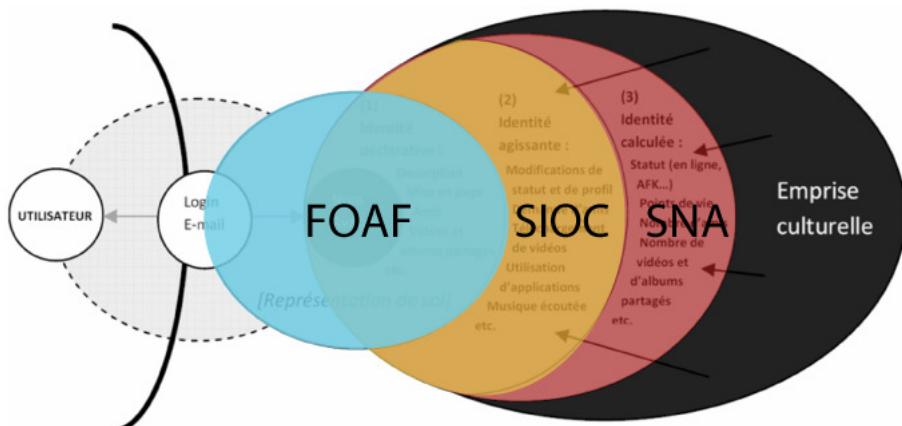


FIGURE 4.5 : Représentation du réseau social déclaré, agissant et calculé

4.4.3 Synthèse du modèle de données

Nous avons présenté précédemment les différents vocabulaires RDF sur lesquels nous nous appuyons pour définir notre format de données. Nous proposons des extensions à ces vocabulaires et un usage conjoint de ces vocabulaires. Nous allons essayer de représenter ici une vue synthétique du modèle global en mettant en avant les articulations entre ces modèles ainsi que les extensions.

Le premier vocabulaire à présenter est indiscutablement le schéma d'annotation Annotea, vocabulaire dédié à la représentation des annotations. Ce vocabulaire ne présente qu'une seule classe, la classe Annotation.

- Une annotation est une ressource, comme toute instance de classe dans RDF. Autrement dit, la classe Annotation étend la super classe Ressource.

Le vocabulaire définit un ensemble de propriétés permettant de lier une Annotation à d'autres objets :

- `annotates` : relie une annotation à la ressource annotée.
- `author` : relie une annotation à son auteur, en l'occurrence une personne représentée via la classe Agent du vocabulaire DublinCore.

- body : relie une annotation à la ressource créée. Dans la plupart des implémentations basées sur Annotea, cette ressource créée est un commentaire saisi dans un champ de texte libre.
- context : permet de préciser la partie de la ressource sur laquelle porte l'annotation. Cette partie de ressource est précisée par une suite de caractères appelé un XPointer dans les implémentations basées sur Annotea.

Nous ne présentons pas les relations "created" et "modified" dans le modèle. Celles-ci servant à lier l'annotation à sa date de création et à sa dernière date de modification. Enfin, la relation "related" permet de lier une annotation à une ressource mais qui n'est pas la ressource annotée. Cette relation n'étant pas réutilisée dans notre approche nous ne la présentons pas également dans le schéma suivant.

Bien que rien dans la spécification d'Annotea n'interdit d'associer plus d'une ressource par la relation `annotates` et également plus d'une ressource par la relation `body`, les implémentations d'Annotea restreignent les utilisateurs à n'associer qu'une ressource annotée et qu'une ressource comme corps de l'annotation. Etant donné que dans nos cas d'usages nous souhaitons pouvoir annoter plus d'une ressource à la fois et ajouter plus d'une ressource, nous précisons sur le schéma que ces relations sont 1.n.

Comme souvent pour les vocabulaires RDF Annotea n'est représenté que sous la forme d'une liste de classes et de propriétés. Nous avons réalisé le schéma suivant^{4.6} reprenant les composantes du vocabulaire Annotea pour en simplifier la présentation.

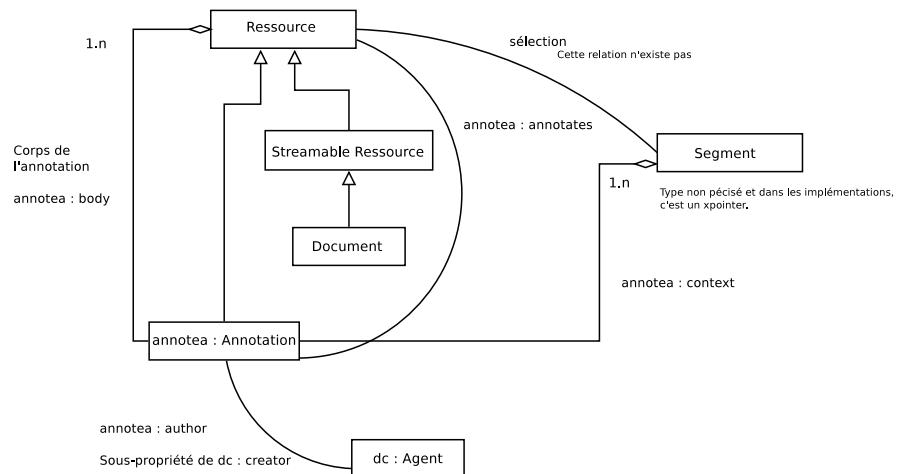


FIGURE 4.6 : Ce que nous retenons du modèle Annotea

La relation `context` de Annotea ne précise pas sur quel type de ressource elle porte et dans les faits, le contexte de l'annotation est précisé par un XPointer. Le fait que ceci ne soit pas conceptualisé comme étant une ressource pose plusieurs problèmes notamment pour retrouver le lien entre le contexte et la ressource dans laquelle est faite la sélection pour l'annotation. Nous avons trouvé dans le modèle NiceTag une classe qui correspond tout à

fait à la représentation de ce concept, la classe PartOfWebRepresentation. Cette classe est de plus liée à la ressource dont elle est une partie par la relation hasPart.

Nous avons déjà présenté le vocabulaire SIOC plus en détail plus tôt. Ce qui est intéressant pour notre travail dans le vocabulaire SIOC c'est essentiellement la classe Post. Cette classe sert à représenter un contenu produit par un utilisateur de type billet de blog, commentaire de blog ou de forum. Un Post est en règle général saisi dans un champ de texte libre. Nous allons utiliser cette classe pour préciser le type de ressources pouvant être utilisé comme corps d'une annotation. Au cours de nos expérimentations, il est remonté comme besoin le fait de pouvoir préciser la sémantique des commentaires saisis. Dans certaines expérimentations nous avons dû définir un champ de "reformulation" ou un champ de "justification". Dans notre implémentation actuelle, il est possible à un utilisateur administrateur de créer de nouveaux types de champs de texte pour les formulaires d'annotation. A chaque nouveau type, une nouvelle classe étend la classe sioc : Post dans notre modèle. Par exemple, pour un champ de justification, nous créons une classe argumentea : Justification qui étend la classe sioc : Post.

Le vocabulaire MOAT définit une classe Tag, une classe Meaning et une relation hasMeaning. La relation hasMeaning permet de lier un Tag à une définition de celui-ci. Nous nous appuyons sur la classe Tag de MOAT pour définir une classe Jugement et une classe Emotion notamment.

La figure 4.7 récapitule le schéma complet d'Argumentea.

4.5 Exemple complet

Nous présentons dans cette partie un exemple complet du processus d'annotation. Nous allons montrer par des captures d'écran comment un utilisateur réalise une annotation et exprime son opinion. Nous présentons ensuite l'export RDF correspondant à l'annotation produite. Cette section permet d'illustrer l'intégralité de notre proposition décrite précédemment.

Nous ne présentons pas la phase d'inscription d'un utilisateur car celle-ci est très classique. L'utilisateur enregistre son nom d'utilisateur et son mot de passe, il peut éventuellement saisir un court texte permettant de le décrire et saisir une url pour communiquer une page personnelle ou un profil FOAF par exemple.

Lorsque l'utilisateur se rend sur le site de débat, il peut choisir de s'enregistrer pour annoter en étant identifié dans le système ou annoter directement sans identification afin de rester anonyme. L'interface de login est là aussi tout ce qu'il y a de plus classique.

Pour créer une annotation, l'utilisateur doit d'abord sélectionner les ressources ou parties de ressources qu'il veut annoter. Cette phase se fait par le biais du navigateur intégré dans le site de débat (l'un des composants de notre solution est un navigateur Web intégré dans une portion de HTML, on se retrouve donc avec un navigateur dans une page Web). Soit l'utilisateur tape directement dans la barre d'url du navigateur intégré l'url de la page

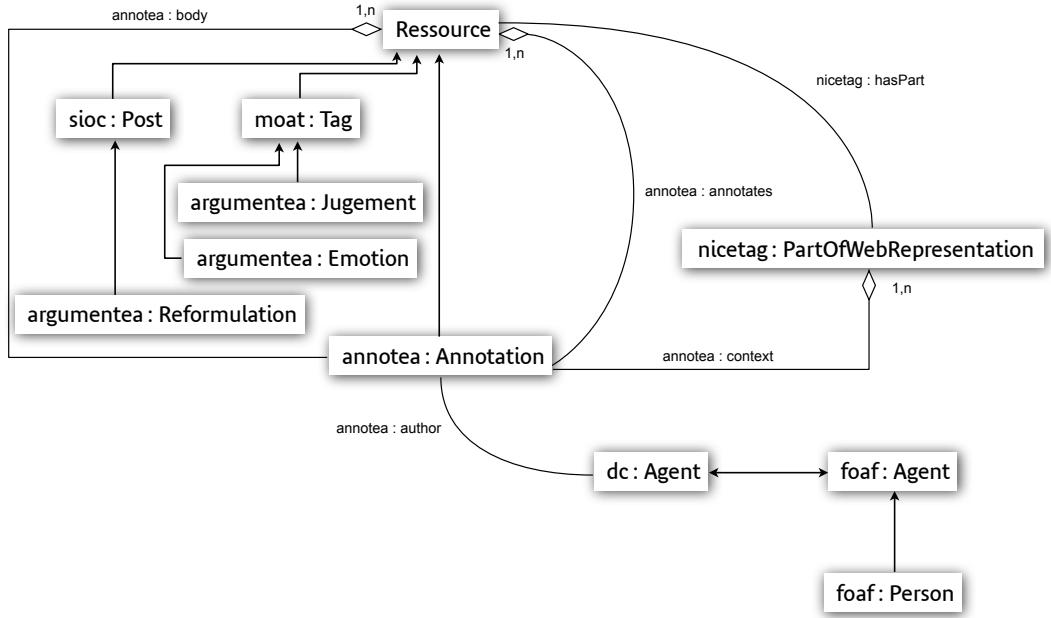


FIGURE 4.7 : Le modèle Argumentea

qu'il veut consulter pour la sélectionner, soit il choisit une page parmi celles proposées dans le portlet de navigation (sorte de gestion des favoris, de collection de bookmarks). La ressource voulue est alors affichée dans le navigateur intégré et l'utilisateur peut soit choisir d'ajouter la ressource entière à son annotation soit sélectionner une partie de texte. L'utilisateur se retrouve alors avec une collection de ressources ou de parties de ressources. Il peut choisir de supprimer ou de garder les éléments de cette collection pour les utiliser dans une ou plusieurs annotations. La capture d'écran 4.8 présente une sélection dans une ressource Web et liste les sélections déjà faites. L'utilisateur coche les sélections qu'il voudra effectivement annoter. La première ressource sélectionnée est une page Web complète.

Une fois qu'il a géré sa collection de ressources, l'utilisateur peut remplir le formulaire d'annotation qu'il souhaite. Pour changer de formulaire, il lui suffit de cliquer sur celui qui l'intéresse dans la liste proposée. Par exemple, dans la capture d'écran 4.9 nous montrons le choix d'un formulaire et dans la capture d'écran 4.10 le contenu du formulaire "à clarifier" permettant d'exprimer que l'on ne comprend pas les passages ou ressources annotés.



FIGURE 4.8 : Sélection dans une ressource et choix des sélections qui seront annotées

Il existe des formulaires bien plus compliqués. Celui-ci ne fait intervenir qu'un tag et un champ texte pour constituer le corps de l'annotation. Mais il est possible d'utiliser des formulaires mêlant plusieurs tags et plusieurs champs textes, chacun ayant sa propre sémantique (commentaire, reformulation ...). Nous pensons que l'ordre de ces champs dans le corps de l'annotation a de l'importance et songeons à mettre cette fonctionnalité en place.

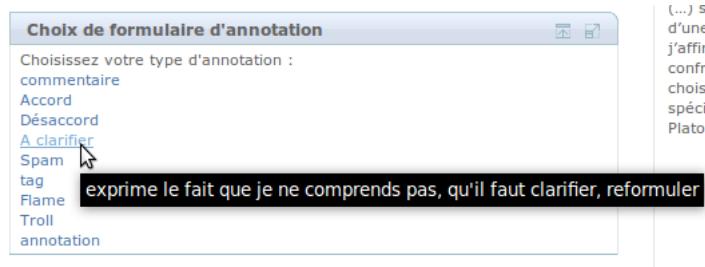


FIGURE 4.9 : Choix du formulaire "à clarifier"

L'annotation, une fois validée, est listée sous les pages Web annotées et peut être colorée dans le texte. La capture d'écran 4.11 présente un ensemble d'annotations listées sous la page Web et une annotation est ancrée dans la ressource par une coloration du passage sélectionné.

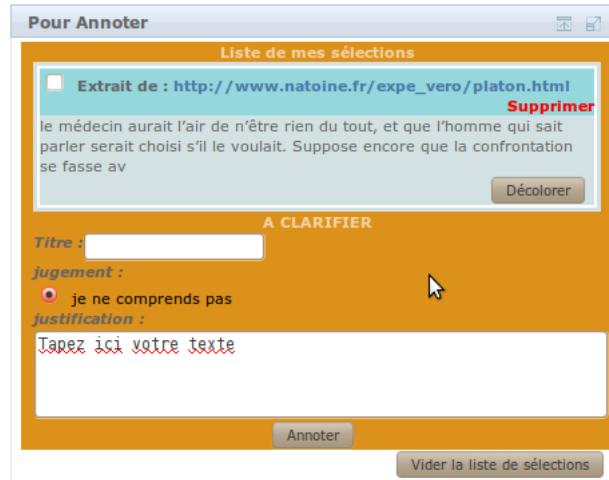


FIGURE 4.10 : Formulaire "à clarifier". Celui-ci est constitué du tag de Jugement "Je ne comprends pas", d'un champ titre pour titrer l'annotation et d'un champ de texte libre ayant le statut de Justification

(...) suppose qu'un orateur et un médecin se rendent dans la Cité que tu voudras, et qu'il faille organiser, à l'Assemblée ou dans le cadre d'une autre réunion, une confrontation entre le médecin et l'orateur pour savoir lequel des deux on doit choisir comme médecin. Eh bien, j'affirme que le médecin aurait l'air de n'être rien du tout, et que l'homme qui sait parler serait choisi s'il le voulait. Suppose encore que la confrontation se fasse avec n'importe quel autre spécialiste, c'est toujours l'orateur qui, mieux que personne, saurait convaincre qu'on le choisit. Car il n'y a rien dont l'orateur ne puisse parler, en public, avec une plus grande force de persuasion que celle de n'importe quel spécialiste. Ah, si grande est la puissance de cet art rhétorique.

Platon, Gorgias, 456b-457c, G.F. 1967, p. 182

natoine : je ne suis pas d'accord, Commentaire : , Ca ne me va pas du tout. Je crois en autre chose !!!

Liste des annotations

Group by : type , Auteur			
De natoine	ok	0 réponses	Répondre à cette annotation
Jugement :	je suis d'accord		Laisser un commentaire
De natoine	Absolument contre !!!!	0 réponses	Répondre à cette annotation
Jugement :	je ne suis pas d'accord		Laisser un commentaire

FIGURE 4.11 : Annotation ancrée dans le texte et annotations listées sous la ressource

CHAPITRE

5

Expérimentations

Préambule

Ce chapitre présente nos choix d'implémentation et notre démarche expérimentale ayant conduit à la version actuelle d'Argumentea.

Sommaire

5.1	Premier test de l'annotation : Betapolitique	150
5.2	Deuxième test de l'annotation : Xwiki	153
5.3	Usages de l'annotation dans Intermed	157
5.4	Moissonnage de données Facebook	165
5.5	Argumentea : Prototype actuel	168

Nous présentons dans cette partie l'implémentation de notre démonstrateur de thèse, l'application Argumentea, et sa mise en œuvre au cours de plusieurs expérimentations s'intercalant dans une démarche de développement itérative. L'objectif de l'application est de permettre la mise en place de débats 2.0 reposant sur l'usage de l'annotation socio-sémantique. Comme objectif secondaire, l'application doit faciliter l'extraction d'un réseau social basé sur l'analyse des interactions entre les utilisateurs. La plupart de nos expérimentations ont consisté en la création de données pour permettre de tester l'utilisabilité de l'application et les formats de données garantissant l'interopérabilité entre des outils dédiés aux débats 2.0. Nous estimons que l'usage de l'annotation permet de faciliter la discussion et le traitement des données produites lors de débats à grande échelle. Cependant, expérimenter véritablement la grande échelle demande la mise en place de débats faisant intervenir un grand nombre d'agents durant une période longue. Nous n'avons pas pu profiter d'un tel contexte qui demande à la fois une réelle volonté politique de mise en œuvre

de tels débats mais aussi de disposer d'une plateforme opérationnelle sur le long terme.

Dans sa première version, l'outil d'annotation était lié à un site internet d'informations politiques : Betapolitique¹. Celui-ci avait l'avantage d'offrir un premier réseau d'utilisateurs. Puis la fonctionnalité d'annotation a été proposée comme extension d'une version du wiki Xwiki². Cette solution offrait l'avantage de pouvoir utiliser l'annotation sur des textes que nous pouvions rédiger via le wiki en fonction des expérimentations (l'annotation a depuis été reprise et intégrée dans les distributions récentes de Xwiki). Ensuite, une solution a été proposée pour conduire les premières expérimentations du projet Intermed. Cette solution reposait sur les choix technologiques du projet à savoir les technologies Java Web et plus particulièrement des Servlets et des Portlets dans un serveur JBoss. Afin de répondre à un temps d'implémentation très court, l'annotation n'était pas directement embarquée dans le portail JBoss mais déléguée à un plugin pour Firefox reprenant le plugin Annozilla. Cette solution dépendante d'un navigateur était vouée à être rapidement remplacée par une nouvelle version. La société Pikko³, partenaire du projet Intermed, a alors proposé un produit reprenant les principes de l'annotation dans un wiki en proposant cette fois-ci une extension pour Mediawiki. Enfin, afin de répondre aux recommandations techniques du projet, le dernier prototype en cours d'utilisation est développé en Portlets et Servlets déployés dans un portail JBoss⁴.

Chacun de ces prototypes a permis aux différents partenaires du projet de mener des expérimentations. Et cette itération successive d'expérimentations et d'implémentations a permis de dresser un cahier des charges complet des fonctionnalités attendues pour un outil d'annotation dédié à l'expression d'opinion, à des débats 2.0. En particulier, la possibilité de proposer aisément de nouveaux tags d'expression d'opinions permet de s'adapter rapidement à des contextes expérimentaux différents. Aujourd'hui, dans nos travaux, l'acceptabilité d'un tel outil auprès de jeunes utilisateurs et les formats de données favorisant l'interopérabilité ont été validés. Nous défendons qu'un réseau social déduit des discussions pourrait permettre de stimuler les débats et en particulier diminuer le phénomène d'insularité des communautés d'opinion en nourrissant un système de recommandations.

5.1 Premier test de l'annotation : Betapolitique

Le travail présenté dans cette section fait référence au travail de stage de Master 2 recherche effectué avant la thèse au LIRMM, sous la direction de Jean Sallantin et Abdelkader Gouaich [Seilles, 2007]. Nous reprenons dans ce document les résultats de ce travail car celui-ci correspond à nos premiers tests d'acceptabilité de l'annotation en tant qu'outil de discussion.

-
1. <http://www.betapolitique.fr/>
 2. <http://www.xwiki.com>
 3. <http://www.pikko-software.com/>
 4. <http://www.jboss.org/jbossportal>

Ce travail fût présenté au cours du carrefour des possibles de la FING en île de France en mars 2007⁵ puis lors du congrès sur le carré des oppositions à Montreux en juin 2007⁶. Il a également donné lieu à une publication suite au congrès du carré des oppositions [Sal-lantin et Seilles, 2009].

5.1.1 Description du dispositif

Lors du stage de Master 2 au LIRMM, nous avons eu l'opportunité de tester l'annotation au sein d'un site d'information politiques en ligne [Seilles, 2007], le site Betapolitique⁷. Cette expérimentation fût réalisée en 2007, le site Betapolitique reposait à l'époque sur la plateforme php SPIP⁸. Nous avions travaillé à l'époque avec la société Pikko software⁹ pour développer l'annotation dans SPIP pour Betapolitique.

Nous avions donc à notre disposition l'ensemble des documents du site Betapolitique, des articles de journalistes à la veille des élections présidentielles. L'annotation constituait en la sélection d'un passage d'article et en l'ajout d'un jeu de jugements puis d'un commentaire. L'annotation apparaissait alors sur le côté de l'article et il était en suite possible de répondre à une annotation à la façon d'un Post sur un forum mais toujours avec un jeu de jugements.

5.1.2 Structure d'oppositions pour les jugements

Nous avions défini 6 jugements organisés en 3 niveaux :

- Le niveau théorique permettant d'exprimer accord et désaccord (je suis d'accord, je ne suis pas d'accord).
- Le niveau pratique permettant d'exprimer si cela fonctionne ou non (ça fonctionne, ça ne fonctionne pas).
- Le niveau émotionnel permettant d'exprimer si l'on aime ou non (j'aime, je n'aime pas).

Nous proposons une organisation de ces jugements reposant sur une structure d'opposition, se basant sur la théorie des n-oppositions. Dans cette structure, le niveau émotionnel est une conséquence du niveau théorique et le niveau pratique est une conséquence du niveau émotionnel. Ce choix correspond à une vision ou l'instinct, l'émotion, peut-être éduqué par la théorie et ou l'action, la pratique est une conséquence de l'instinct.

Nous avions trouvé à l'époque une correspondance entre cette structure de jugement et la structure d'une logique défaisable [Antoniou *et al.*, 2000]. Mais nous n'avons pas investi plus en avant ce point. Les logiques défaisables permettent de prendre en compte

5. http://doc.openfing.org/CDP/PDF/CDP_38_IDF_080307.pdf

6. <http://www.square-of-opposition.org/enter1.html>

7. www.betapolitique.fr

8. www.spip.net

9. www.pikko.fr

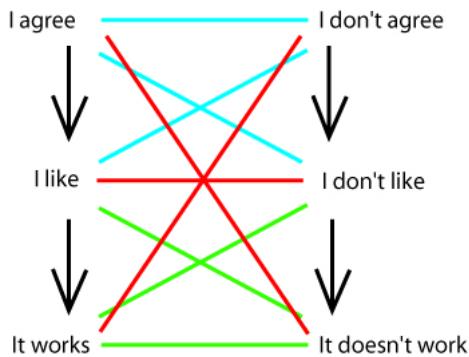


FIGURE 5.1 : Structure d'opposition de l'expérimentation betapolitique

des situations de conflits dans les bases de faits, ce qui est le cas dans les débats. Nous recommandons l'article de référence concernant les logiques défaisables [Nute, 1994].

5.1.3 Analyse

Cette première expérimentation aura durée 3 mois de Février à Mai 2007. Les utilisateurs furent confrontés à l'annotation du jour au lendemain sans aucune notice d'utilisation. Nous n'avions pas défini de sujet de débat à proprement parler. Les utilisateurs étaient identifiés par le biais de pseudonymes ou restaient totalement anonymes. Rien ne contrainait ni n'incitait les utilisateurs à participer, si ce n'est la présence des annotations déjà faites par d'autres sur le côté de la page 5.2. Une application de visualisation de données, un Treemap 5.3, développé par la société Pikko permettait de cartographier les articles les plus annotés. Un Treemap est une forme de visualisation habituellement utilisée pour représenter des données ayant une structure d'arbre. Chaque nœud de l'arbre est représenté par une surface dont la taille est proportionnelle au nombre de fils du nœud. Ici, chaque surface représentait un article et la taille dépendait du nombre d'annotations sur l'article.

La figure 5.2 présente un article de betapolitique avec des annotations sur la droite. La figure 5.4 présente le formulaire de saisie du contenu des annotations dans betapolitique.

Cette expérimentation nous a permis de confirmer l'acceptabilité de l'annotation et de formulaires utilisant des jugements dans un contexte de discussion en ligne. Nous avons eu 870 annotations en 3 mois d'expérimentations et 6000 commentaires de réponses aux annotations 5.5. Tous les jugements furent adoptés. Nous ne prenons pas en compte les annotations manquées ou de test. Nous avons à peu près 20% d'annotations de tests en plus des 870 annotations considérées comme utiles.

Betapolitique blog politique connectif
Chroniques |Décryptages |Participer |Ré-Veilles |Ressources |

Ségolène Royal : les 100 propositions du pacte présidentiel
dimanche 11 février 2007

Thèmes : Ségolène Royal

Répondre à cet article
La candidate socialiste a dévoilé aujourd'hui, à l'issue d'un grand discours programme clôturant la phase de consultation des « débats participatifs », un ensemble de 100 propositions constituant le socle de son « pacte présidentiel ».

Ségolène Royal : les 100 propositions du pacte présidentiel
3 mars 2007 16:32

A_PROPPOS : ""26- Réviser la carte scolaire pour supprimer les ghettos scolaires, assurer la mixité sociale et constituer des réseaux d'éducation prioritaire. MON AVIS :

ENTHEORIE_21 idéalement, j'adhère totalement

ARGUMENTAIRE voici mon argumentation : la mixité devient naturelle pour tous ceux qui l'ont vécue ainsi l'école est LE lieu ideal pour qu'enfin les différentes cultures cohabitant en france deviennent un avantage.

Répondre à ce message

Ségolène Royal : les 100 propositions du pacte présidentiel
1er avril 2007 10:37, par jean

instinctivement, j'aime, voici mon argumentation :
enfin, un jeune content d'avoir appris à vivre dans un monde multiculturel

FIGURE 5.2 : Un texte annoté dans Betapolitique



FIGURE 5.3 : Un treemap représentant les textes annotés

5.2 Deuxième test de l'annotation : Xwiki

Nous présentons dans cette section une expérimentation réalisée suite au stage de Master 2 recherche en attendant le démarrage officiel du projet ANR Intermed. Nous avons profité de l'organisation de la conférence ECAP08 à Montpellier pour utiliser l'annotation comme outil de "reviewing" d'articles.

5.2.1 Description du dispositif

La société XWiki était à l'époque intéressé par l'intégration d'une fonctionnalité d'annotation dans son wiki¹⁰. Notre travail a permis de tester la faisabilité de cette intégration.

10. www.xwiki.com

Instinctivement	<input type="radio"/> j'aime	<input type="radio"/> je ne sais pas	<input type="radio"/> j'aime pas
Par principe	<input type="radio"/> j'adhère	<input type="radio"/> je n'ai pas d'avis	<input type="radio"/> je n'adhère pas
En pratique	<input type="radio"/> cela fonctionne	<input type="radio"/> il n'y a pas d'exemple	<input type="radio"/> cela ne fonctionne pas

FIGURE 5.4 : Formulaire d'annotation dans Betapolitique

		percent
I like	147	16,9
I don't like	127	14,6
I agree	120	13,79
I don't agree	86	9,89
It works	115	13,22
It doesn't work	162	18,62
others	113	12,99
Total	870	100

FIGURE 5.5 : Répartition des jugements dans Betapolitique

Aujourd’hui, les dernières versions de XWiki intègrent l’annotation¹¹.

L’idée de cette expérimentation était de proposer aux participants de la conférence un outil leur permettant de discuter les résumés des articles sélectionnés. Ils pouvaient ainsi signaler les articles qui les intéressaient, proposer des améliorations de syntaxe ou de contenu, et proposer des références de travaux connexes. A terme, nous souhaitions coupler cette approche de l’annotation à un véritable outil de "reviewing" d’article. Les données produites par les débats autour des résumés permettant de mesurer l’attractivité des présentations auprès des participants de la conférence, le planning de celle-ci serait établi en se basant sur ces données. Cependant, le faible niveau de contribution sur l’outil avant la conférence n’a pas permis d’établir le planing selon ce principe.

Une partie du site Web de la conférence ECAP08 était constitué d’un Xwiki que nous avions modifié pour intégrer l’annotation. Les pages du wiki étaient l’ensemble des résumés sélectionnés pour la conférence. Les annotations pouvaient être saisies par le biais de 3 formulaires. L’un permettant de signaler son adhésion ou le contraire, le second permettant de signaler des erreurs de syntaxe et enfin, le troisième permettant de recommander des références supportant ou contre-disant les propos surlignés. Nous avions défini 6 jugements, 3 couples de jugements :

- **Je suis d'accord, Je ne suis pas d'accord.**
- **C'est bien écrit, C'est mal écrit.**

11. [http://www.xwiki.com/xwiki/bin/view/Blog/New Annotations Feature with XWiki Enterprise](http://www.xwiki.com/xwiki/bin/view/Blog/New%20Annotations%20Feature%20with%20XWiki%20Enterprise)

– **Cela confirme, Cela s'oppose.**

A chaque fois, les deux jugements d'un couple étaient reliés par une relation de contradiction. Les formulaires étaient aussi composés d'un champ de saisi de texte portant la mention "reformulation". Nous souhaitions tester si les utilisateurs justifieraient bien leurs jugements ou non.

L'expérimentation a débuté quelques jours avant la conférence et s'est terminée avec la conférence. Nous avions mis un tutoriel en ligne. Les annotations étaient listées sur le côté droit de chaque page [5.6](#). Nous avions couplé l'application d'annotation à un outil de dessin de graphe développé par la société Pikko software pour offrir une vue des annotations sur chaque article [5.8](#).

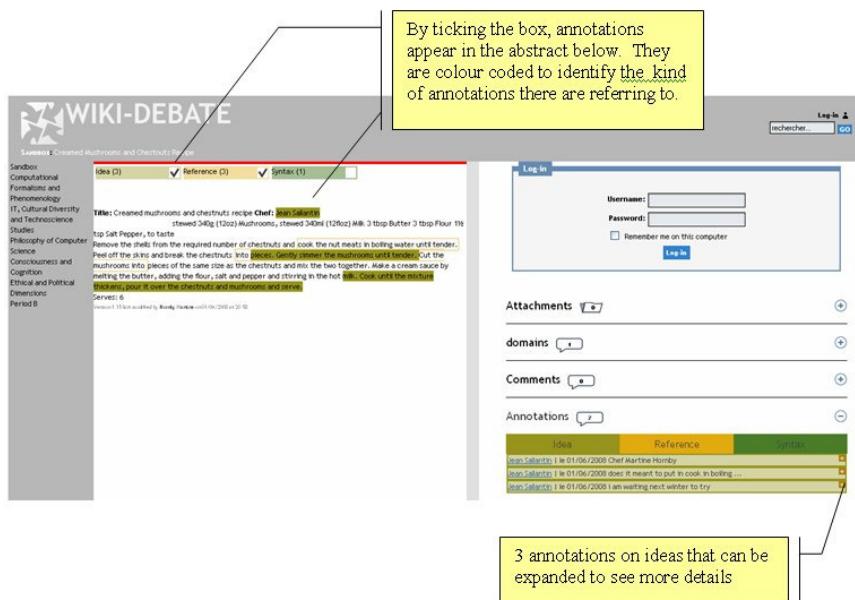


FIGURE 5.6 : Screenshot of annotations for ECAP

5.2.2 Analyse

Cette seconde expérimentation nous a permis de valider l'importance de l'interface d'annotation ainsi que le besoin d'éléments de gratification immédiate comme l'existence d'un tableau de bord de chaque utilisateur ou l'affichage de statistiques par exemple. Peu de participants à la conférence ont utilisé l'outil. Par le biais d'interviews et d'un sondage de satisfaction de la conférence, les participants ont exprimé qu'ils trouvaient l'idée intéressante mais l'outil trop compliqué.

FIGURE 5.7 : Une annotation dans Xwiki pour ECAP

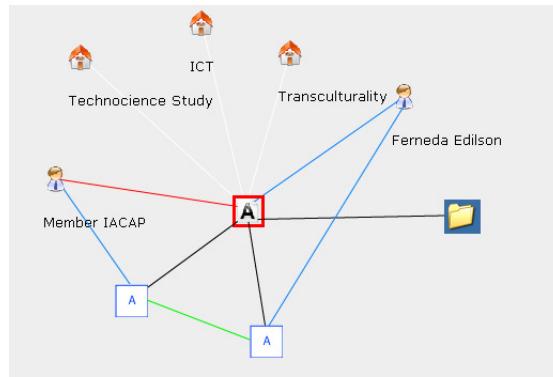


FIGURE 5.8 : Visualisation des accords et désaccords dans un graphe de commentaires, un graphe de débat

L'analyse des visites du site via l'outil Google analytics¹² 5.9 montre que le plus haut taux de connexion est atteint le premier jour de la conférence (le 16 juin 2008) avec 147 visiteurs tandis que nous avions 80 inscrits à la conférence. Nous avons eu un peu plus de 300 visites entre le 15 et le 19 juin (la conférence se terminant le 18). Le temps de visite moyen est de 4 minutes. Nous avons eu un peu moins de 100 annotations (95 exactement), la plupart faites le 19 juin. Sur les 70 résumés présentés, le plus annoté le fut 7 fois. La plupart des annotations n'ont pas donné suite à des réponses. Aucune discussion ne fut amorcée via l'outil. Le comité d'organisation de l'édition suivante de la conférence, ECAP09 adopta à nouveau notre outil d'annotation sous Xwiki pour cette autre édition.

12. <http://www.google.com/intl/fr/analytics/>



FIGURE 5.9 : Statistics of visits of ECAP website

5.3 Usages de l'annotation dans Intermed

L'annotation est l'un des choix de départ du projet Intermed. Nous avons donc participé au développement de prototypes pour l'annotation dédiés à des expérimentations menées par d'autres acteurs du projet.

Nous allons dans cette section présenter les prototypes développés dans ce contexte au préalable du développement de notre outil d'annotation actuel : Argumentea.

5.3.1 Phases de coconception et évocation des besoins

La conception et le choix des outils adoptés dans le projet Intermed ont été le sujet de nombreuses longues réunions permettant aux chercheurs et sociétés du projet de trouver un vocabulaire commun et de définir les besoins communs.

Nous avons souligné dans l'article [Seilles *et al.*, 2011] des difficultés liées à la pluridisciplinarité, fruit de l'observation du travail de coconception dans Intermed. Nous allons synthétiser ici ces phases de coconception.

Béatrice Pary introduit ainsi la méthodologie adoptée pour le projet dans un rapport interne dédié à la préparation des expérimentations Intermed :

La méthodologie retenue est fondée sur une interaction forte avec les terrains partenaires. Elle cherche à concilier l'adéquation et la réponse aux besoins concrets des acteurs du terrain, la faisabilité technique, financière et temporelle par l'équipe du projet, et l'innovation scientifique et technique. Cette

perspective constitue en soi un défi pour le projet, dans la mesure où les services et technologies visées sont à la fois nouvelles pour l'application visée – la démocratie locale -, et potentiellement risquée politiquement pour les gestionnaires et élus. Par ailleurs la plupart des partenaires industriels n'a pas d'expérience dans ce domaine, dans un projet aussi intégrateur, et en se rendant dépendant des besoins terrains. Cela suppose un arbitrage fin entre une gouvernance "descendante", gage de clarté et de continuité dans le plan de développement, et une construction ascendante.

Une première difficulté rencontrée était la participation de gouvernants, d'élus dans les phases d'évocation des besoins. Afin d'être en adéquation avec les demandes utilisateurs, Intermed a jugé bon de faire participer des élus et des citoyens dans la phase de conception. Les élus acceptant de prendre part à cette démarche sont difficilement mobilisables de par leurs emplois du temps et ils émettent parfois des réticences au projet en fonction de l'agenda des prochaines élections.

Les événements clefs de cette méthodologie de conception sont donc :

- les réunions ou forums d'évocation des besoins au cours desquelles est attendu de la part des utilisateurs l'expression des usages et fonctionnalités attendues.
- les expérimentations au cours desquelles des outils censés répondre aux besoins identifiés sont soumis à l'usage.

InterMed a souhaité proposer 3 "protocoles d'expérimentation" chaque année. Le document rédigé par Béatrice Pary propose la définition de ce qu'est un protocole dans Intermed :

Un protocole est la conjonction d'un besoin identifié dans un terrain, d'une solution logicielle envisagée, d'un schéma de test et d'un processus d'évaluation. L'objectif d'un protocole est donc de tester une approche vis-à-vis d'un problème exemplaire.

La figure 5.10 précise l'ordre des enchainements entre les réunions dans le projet Intermed.

Cette méthodologie de conception était prévue originellement pour fonctionner avec 4 partenaires de terrains. Nous en avons retenu 3 au final, la Camargue, le bassin de Thau, la région Bretagne.

Une méthodologie générale a été définie pour les forums des besoins. Son premier objectif était de permettre aux acteurs locaux de décrire leurs activités et projets dans lesquels le public peut être sollicité pour contribuer à la rédaction de documents de gestion. Le deuxième objectif était de définir en quoi et comment un outil Internet pourrait faciliter cette mise en œuvre ou ces processus.

La rencontre avec le groupe était prévue selon le programme suivant, sur la base de 3h de réunion :

- Présentation rapide d'InterMed et discussion sur les attentes du projet
- Étape individuelle de réflexion : identification des besoins perçus par chacun

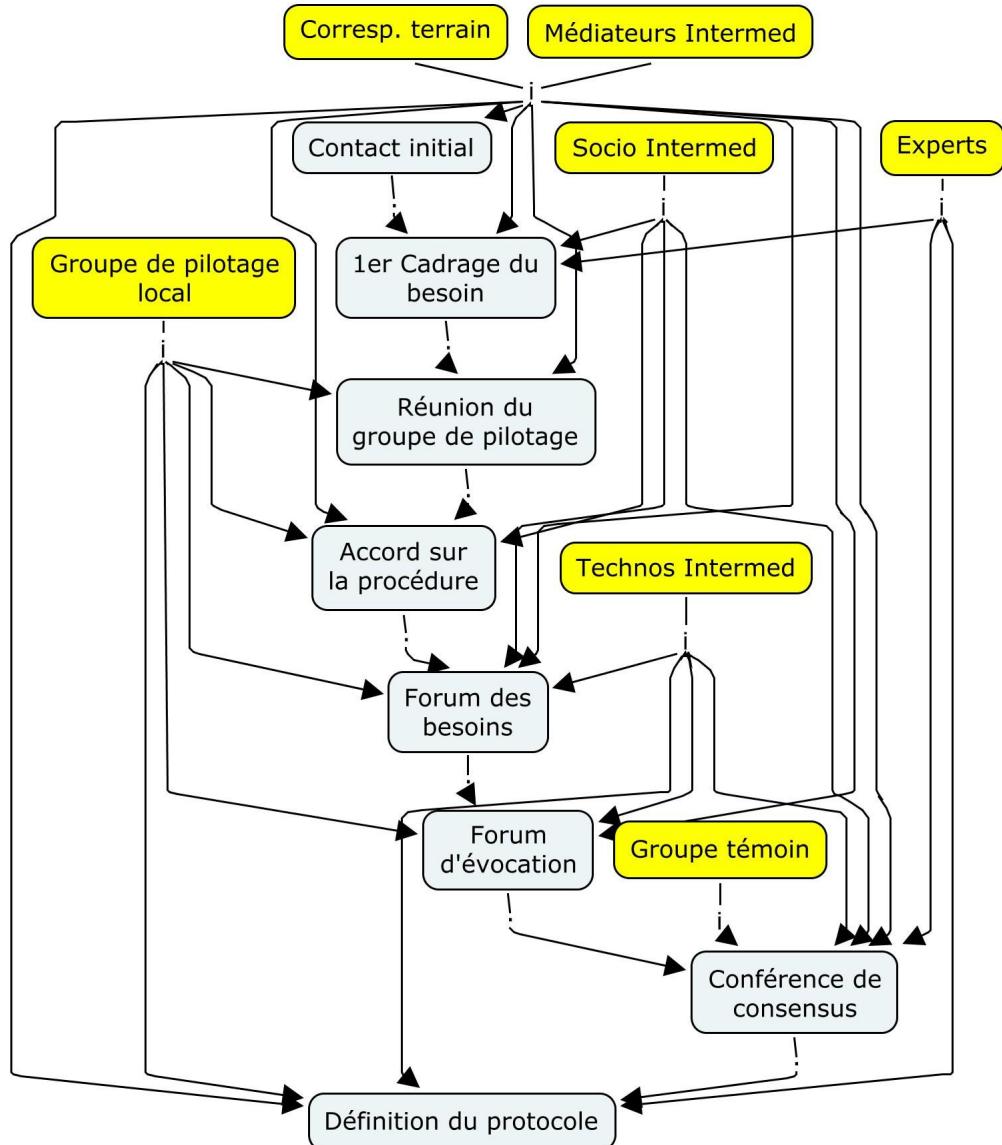


FIGURE 5.10 : Organigramme des dépendances entre les réunions Intermed

- Mise en commun : définition du plan d'action et identification des principales étapes par les acteurs présents
- Identification des étapes pouvant être assistées par l'outil ?
- Conclusion

L'étape individuelle de réflexion sur les besoins perçus demandait aux participants de préparer la réponse aux questions suivantes (sur des fiches individuelles) :

1. Quels sont les projets ou actions dans lesquels vous êtes amenés à rédiger (vous-mêmes) ou faire rédiger (par des experts ou par des groupes de travail) des chartes, des plans de gestion ou d'autres documents ? Préciser le type de document.
2. Comment le public est-il associé ? Selon quelles modalités de participation ?
3. Quels problèmes se posent ? Quelles difficultés sont rencontrées ?
4. Comment pensez-vous qu'Internet pourrait aider à les résoudre ?
5. Quels sont donc vos besoins d'outils Internet en tant que gestionnaires ?
6. Quels sont les besoins Internet des citoyens ?

La synthèse de ces besoins a permis d'établir une liste de scénarios à partir de laquelle furent définis les premiers outils devant être développés par Intermed. Les outils de discussions pour la concertation et les outils de co-rédaction y tiennent une part prépondérante. Ces scénarios étaient ensuite présentés lors de forums d'évocation visant à faire valider l'adéquation de nos propositions avec les besoins identifiés.

Les scénarios proposés lors de ces forums d'évocation sont les suivants :

- Scénario 1 : Assister le débat autour de la production d'une charte ou plan de gestion de territoire, avec consultation étendue de la population et gestion de la procédure
- Scénario 2 : Rédaction d'un projet de territoire dans le cadre d'une action éducative et citoyenne avec des jeunes (projet inter-lycées, SAGE des jeunes, etc.)

Ces scénarios visaient à illustrer un processus d'usage complet, incluant au mieux les fonctionnalités innovantes prévues dans InterMed. Ils étaient choisis pour évoquer des propositions d'outils, sachant que ce qui serait développé finalement pouvait être différent et dépendrait des réactions et demandes des participants.

Les résultats des forums d'évocation permettent de lister un ensemble de problèmes remontés par les structures publiques en charge de la concertation. Ces problèmes sont en adéquation avec les verrous des débats 2.0.

Des problèmes communs relatifs à la participation :

- Mobilisation difficile à maintenir : disponibilité et temps limité des citoyens, retard ou faiblesse des concrétisations politiques...
- Représentativité faible avec "toujours les mêmes" (tête de réseau) et donc un "assèchement" des propositions
- Faible participation en général
- Défiance citoyenne sur l'instrumentalisation de la concertation et défiance parallèle des élus sur risque de contre-pouvoir
- Sentiment fort anticipé de non-intégration des contributions

Autres problèmes identifiés :

- Problème de moyens pour la maintenance du site
- Problème de modération des outils de discussions
- Problème de crédibilité de l'information officielle via la presse et crainte de relancer des polémiques

- Les sites personnels, blogs associatifs et la presse sont perçus comme négatifs, pas constructifs, véhiculant des rumeurs

Plus précisément, au niveau des modules ou fonctionnalités proposées, les observations suivantes ont été faites (nous ne retenons que celles relatives à nos travaux) :

1. La cartographie sociale crée des craintes de cristallisation et d'inhibition. Alors qu'aujourd'hui les outils de réseaux sociaux sont communément acceptés, utilisés, cet aspect poserait problème dans le contexte de débats car elle permettrait d'identifier les individus, les opposants ou les sympathisants. On retrouve la peur du flicage, l'identification pour le contrôle. La cartographie sociale n'est pas perçue comme un plus pour la création et l'organisation de groupes d'opinions, pour la recherche de soutien.
2. La question du contrôle des identités est un point délicat. Si l'on permet trop d'anonymat, on n'assure pas la bonne représentativité et on risque d'avoir des dérives. Les individus déresponsabilisés écriraient n'importe quoi. A l'inverse, si l'identification est trop importante (usage de noms réels par exemple), on risque d'inhiber et d'exclure des usagers qui craignent le contrôle.
3. La catégorisation des documents ou arguments semble trop complexe.
4. Il y a une crainte de transformer en conflits personnels des désaccords exprimés vis à vis de documents.

Suite à ces forums de besoin et d'évocation, Intermed a organisé ses expérimentations autour de trois axes :

1. La citoyenneté chez les jeunes. Ces expérimentations ciblent donc de jeunes usagers et recouvrent la plupart des usages.
2. Rédaction en groupe de Chartes. Ces expérimentations ciblent plus particulièrement la co-rédaction de chartes.
3. Application de Chartes. Il s'agit d'aider à la conception de plans d'actions visant à mettre en application une charte écrite au préalable

Dans nos travaux, nous avons essentiellement participé au premier axe d'expérimentation. Bien que l'annotation fût utilisée dans les 3 axes.

5.3.2 Description du premier dispositif

Le premier prototype développé pour l'ANR Intermed était constitué d'un ensemble de Portlets dédiés à la gestion des documents de référence pour les débats et de gestion des utilisateurs. Ces Portlets étaient couplés à une version légèrement modifiée de l'outil d'annotation Annozilla (plugin pour firefox reposant sur le protocole Annotea).

Dans ce prototype, le formulaire d'annotation pouvait être défini via l'un des Portlets. Le formulaire adopté pour l'expérimentation permettait d'exprimer son opinion via l'utilisation de jugements et de tags d'émotions. Les tags d'émotion étaient dans cette version

venus remplacer les émoticônes que nous n'avions pas pu intégrer dans les formulaires. Les utilisateurs pouvaient reformuler leurs sélections et ajouter un commentaire ayant statut d'explication de leur opinion. L'expression d'opinion se faisait cette fois-ci par un choix de réponse oui ou non à une liste de trois questions : Est-ce prouvé ? Est-ce obscur ? Est-ce incorrect ?

Ce prototype a permis de réaliser une première expérimentation de l'ANR Intermed où les sociologues ont pu comparer des modalités de débats en présence avec l'usage de l'annotation.

5.3.3 Description du deuxième dispositif

Suite à cette première expérimentation, la société Pikko développa à son tour un prototype d'annotation, cette fois-ci lié à l'outil mediawiki¹³, permettant d'annoter les pages d'un wiki.

Une première version du prototype utilisé dans Intermed permettait des annotations simples ne mentionnant que du texte libre. En vu d'une expérimentation auprès de Paris-tech, nous avions demandé à la société Pikko de reprendre les formulaires d'annotation précédemment testés, à savoir des formulaires reprenant le principe des champs de texte libre pour la reformulation et l'explication de l'opinion exprimée, proposant 3 choix de jugements pour l'opinion (je suis d'accord, je ne suis pas d'accord, je m'interroge), et proposant de préciser son humeur via un choix d'émoticône parmi 4 propositions.

La suite de la discussion était structurée par un choix de formulaire de réponse aux annotations. Dans ce cas, à nouveau, les annotations sont le début de fils de discussion. Les formulaires de réponses à une annotation étaient les suivants :

- Un formulaire permettant d'exprimer directement une opinion de soutien. Le tag de jugement "*je suis d'accord*" est automatiquement saisi, accompagné d'un choix d'émoticône et deux champs de texte : commentaire et suggestion.
- Un formulaire permettant d'exprimer directement une interrogation. Le tag de jugement "*je m'interroge*" est automatiquement saisi, accompagné d'un choix d'émoticône et deux champs de texte : commentaire et suggestion.
- Un formulaire permettant d'exprimer directement une opinion d'opposition. Le tag de jugement "*je ne suis pas d'accord*" est automatiquement saisi, accompagné d'un choix d'émoticône et deux champs de texte : commentaire et suggestion.
- Un formulaire de synthèse permettant par le biais d'un champ de texte libre de synthétiser l'ensemble de la discussion (les annotations et commentaires précédents dans la discussion).
- Un formulaire de vote permettant d'associer une note à chaque élément de discussion, permettant une expression différente et plus rapide de soutien ou d'opposition.

13. <http://www.mediawiki.org/wiki/MediaWiki/fr>

Dans ce prototype, les formulaires et les règles d'enchaînement des formulaires étaient à nouveau pré-définis dans l'application. Ce prototype a servi lors d'une expérimentation Intermed menée conjointement avec l'école ParisTech.

Suite à cette expérimentation, nous avons entamé le développement du prototype actuel d'annotation, Argumentea, reprenant les choix de technologie du projet Intermed (Portlets et Servlets). Nous présentons ce prototype dans [la section dédiée à Argumentea](#).

5.3.4 Analyses

Jeunes utilisateurs

Dans l'article [[Desquinabò et al., 2009](#)] Nicolas Desquinabò présente un ensemble de résultats provisoires concernant l'axe expérimental destiné aux jeunes utilisateurs. Des étudiants sont invités suite à une visite sur le terrain à établir un plan d'aménagement d'un territoire. Les étudiants jouent le rôle de différentes parties prenantes et doivent défendre leurs plans vis à vis des autres par le biais d'annotations semi-structurées. Les documents de départ pour les annotations sont les plans d'aménagement de chaque équipe.

Les débats en ligne via l'annotation sont comparés à des débats en présence. Il est attendu de l'annotation qu'elle augmente l'interactivité, permette à tous de s'exprimer et donc améliore la représentativité et la qualité des arguments notamment en augmentant la diversité d'arguments exprimés et en mettant en exergue plus de changements d'opinions. Les premières expérimentations menées sur ce sujet ont plutôt confirmé ces hypothèses.

La qualité des débats est mesurée à partir de la densité d'arguments (qui parle le plus ou le moins ? Quelle quantité d'arguments par personne ? ...), le taux d'échanges entre les utilisateurs, le nombre de conflits et la diversité d'arguments exprimés. Les variables d'impact identifiées sont les changements d'opinion et la dynamique du répertoire des arguments utilisés (à quel moment y a-t-il le plus d'arguments ? Sur quels sujets ? ...).

Nicolas Desquinabò souligne l'intérêt de l'annotation dans le contexte de plans d'aménagements du territoire. Dans ce cas, les documents critiqués font en général plus d'une centaine de pages et les annotations semi-structurées semblent être un bon compromis entre sondage en ligne et concertation en texte libre. Les critiques peuvent ainsi être localisées dans le texte, les annotations semi-structurées permettent de garder une forme d'expression peu contraignante à base de commentaires tout en permettant un minimum de traitements comme l'indexation à partir des jugements, le calcul rapide du nombre d'accords et de désaccords. L'expérimentation fut reproduite deux fois, se basant sur le premier prototype d'annotation que nous proposions.

Dans les deux cas, les débats en ligne via l'annotation ont montré qu'ils étaient moins inégaux que les débats en présence. Il y a moins d'écart à la moyenne de production d'annotation pour chaque individu qu'il y a d'écart de temps de parole.

Mesurer le taux d'interactivité est une tâche aisée dans le cadre d'un débat en ligne. Il suffit d'observer la structure des discussions en se basant sur les relations d'annotation.

Par contre, cet indice est très délicat à mesurer en débats en présence ou la règle est le tour de parole. Ce n'est donc pas un bon critère de comparaison.

Concernant le niveau de conflit, le résultat dépend de ce que l'on mesure entre taux de *flame* (propos injurieux) ou taux de désaccords exprimés. Les débats en présence font état de moins de *flame* et proportionnellement de plus de désaccords exprimés. Cependant, les débattants passent plus de temps à défendre et ré-expliquer leurs projets dans un débat en présence que dans un débat via l'annotation. Enfin, concernant le *flame*, celui-ci est plus élevé dans les débats via l'annotation mais reste bien moins important que le taux de *flame* habituellement constaté dans les débats en ligne non modérés via forum.

Concernant la diversité des arguments émis, le total reste à peu près le même dans les deux conditions. Le nombre d'arguments différents pour ou contre émis est à peu près le même. Par contre il y a plus de messages exprimant chaque argument.

Le taux de changement d'opinions est à peu près le même dans les deux conditions. Mais cette mesure est ici probablement biaisée, les étudiants préfèrent à chaque fois un projet mettant l'accent sur son impact environnemental positif.

La Camargue

Cette expérimentation repose sur l'usage du deuxième prototype d'annotation dans la discussion d'un plan d'aménagement de territoire en Camargue. Il est important de souligner que l'enjeu est ici réel, les données produites ont réellement servi à la concertation concernant l'aménagement de la Camargue. Cette expérimentation est plus précisément décrite dans [Desquinabò et Ferrand, 2009] et [Desquinabò *et al.*, 2010].

Il était attendu dans cette expérimentation qu'il y ait des débats plus équitables en ligne, avec plus d'opinions différentes exprimées, plus de connaissances acquises par les utilisateurs ... en bref, que les débats soient plus intéressants pour les parties prenantes. Il était aussi attendu que ce procédé de concertation soit profitable à un niveau organisationnel en gagnant en qualité d'information pour un coût réduit d'organisation de la concertation.

Cette fois-ci, le dispositif expérimental comprend un système d'inscription contrôlé et un système de modération "légère". Les participants sont donc choisis car connus pour être intéressés par ce débat et ils sont également clairement identifiés dans l'application (pas de pseudonymat ou d'anonymat). La modération "légère" consiste à annoncer que les messages dont le contenu est gênant seront supprimés mais s'arrête là. Finalement, il n'y a pas eu de modérateur. L'annotation dans cette expérimentation se fait de façon totalement libre. Il n'y a pas de formulaire structurant, pas de jugements, juste un champ libre.

Les premiers résultats de cette expérimentation ont confirmé l'intérêt organisationnel d'un tel dispositif. Ils ont aussi permis d'observer une hausse de la diversité d'opinions exprimées sans augmentation du *flame*. Cependant, il est ici observé une plus grande concentration de l'expression qu'en face à face. Certaines personnes ont énormément annoté tandis que d'autres ne se sont pas du tout exprimés. De nombreux utilisateurs n'ayant

pas annoté ont exprimé comme raison que ce qu'ils souhaitaient dire avait déjà été dit par d'autres.

5.4 Moissonnage de données Facebook

Vers la fin de la première année de thèse, nous avons commencé à travailler avec Fanny Georges sur la question de l'identité numérique et de la représentation de soi en ligne, en réseau.

Nous présentons ici les résultats d'une expérimentation menée conjointement avec Fanny Georges et plusieurs membres du LIRMM (Mountaz Hascoët, Nancy Rodriguez, Bérenger Arnaud, Guillaume Artignan). La coopération mise en place a permis de soulever des hypothèses concernant la nature des relations entre membres d'un réseau, et de proposer une visualisation à base de graphes noeuds-liens s'appuyant sur les développements de thèse de Guillaume Artignan [[Artignan et Hascoët, 2010](#)] qui simplifie la lecture et l'analyse des données collectées, ici, dans Facebook.

5.4.1 Contexte et dispositif

Dans son travail d'étude et d'analyse de l'identité numérique, Fanny Georges s'est notamment intéressée à l'observation de comportements dans Facebook et Myspace et à la façon dont les indices de notre identité nous étaient renvoyés dans ces applications de réseau sociaux (SocNet).

Facebook et Myspace permettent de créer un réseau d'amis, et relèvent tous deux d'un mélange catégoriel : l'utilisateur peut bloguer, partager des albums, de la musique et des vidéos. Comportant donc de nombreux points communs, ces deux logiciels diffèrent par leur tendance : Facebook opte pour la diversité (applications de rencontre, notation, jeux, musique etc.), proposant de fédérer toutes les facettes de la vie quotidienne, tandis que Myspace se concentre sur les centres d'intérêt musicaux (fonctionnalités avancées de partage de médias).

Plus particulièrement, nous nous sommes intéressés dans ce travail à proposer une visualisation de la structuration du lien social entre utilisateurs par les éléments calculés : médias partagés, amis communs et tags photo. L'idée que nous défendons dans cette thèse concernant les réseaux sociaux en ligne actuels, est qu'il manque souvent de diversités dans les types de liens et d'entités représentés dans les graphes de réseaux sociaux. Facebook ne relie que des utilisateurs par des liens d'amitié. Pourtant, les visualisations que nous présentons ici montrent bien que tous ces amis n'interagissent pas de la même façon les uns avec les autres. Conformément à la théorie de Nadel [[Nadel, 1957](#)], nous supposons que ces relations sont influencées, voir déterminées, par les rôles ou classes des individus dans le réseau.

Les données utilisées sont obtenues par le biais d'une application Facebook que nous

avons développé pour cette expérimentation. Cette application nous permettant de collecter les informations de l'utilisateur relatives à ses échanges avec les membres de son réseau. Les données sont ensuite anonymisées. Les utilisateurs nous ont cédé leurs données en totale conscience de l'usage que nous en faisions. Nous n'avons utilisé que des comptes Facebook de proches.

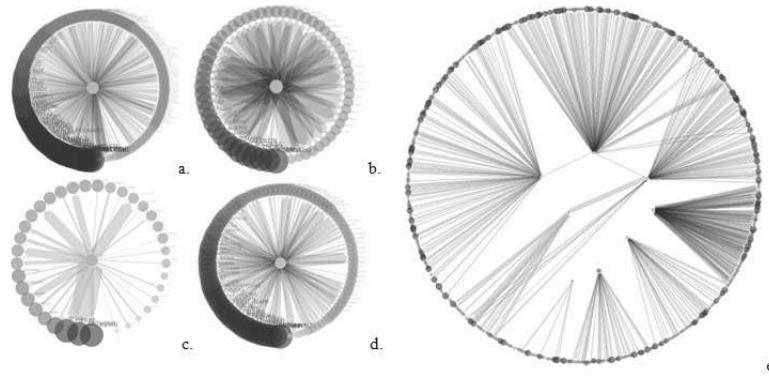


FIGURE 5.11 : Vues de profils d'utilisateurs individuels (a. à d.) et vue collective de 8 utilisateurs dans Facebook (e.). Les liens entre l'utilisateur X et ses amis Y. sont de trois couleurs : les liens bleus représentent le nombre d'amis en commun entre X et Y. Les liens jaunes représentent le nombre de groupes partagés entre X et Y. Les liens roses représentent le nombre de photos de X taggées au nom de Y. L'épaisseur des liens traduit l'intensité des relations.

L'objectif de cette représentation est de permettre d'évaluer la quantité d'informations de profil déclarées par les utilisateurs ainsi que les relations qu'ils entretiennent avec leurs amis via Facebook. Etant donnée la nature des informations étudiées, une visualisation à base de graphes (vue noeuds-liens) nous a semblé légitime.

La visualisation de graphes repose sur la théorie des graphes [Berge, 1958]. Cette théorie introduit entre autre la notion d'acteurs (nommés noeuds) liés par des relations (nommées arêtes ou liens). Les graphes sont largement utilisés dans le domaine de la visualisation de connaissances [Herman *et al.*, 2000].

5.4.2 Description des graphes

Les graphes de 5.11 présentent des vues de profils d'utilisateurs Facebook. Le graphe de 5.11 présente une vision globale de 8 utilisateurs. L'indice de l'identité déclarative est retracé par la taille et la couleur de chaque noeud. Plus l'utilisateur a rempli d'informations dans sa page de profil Facebook, plus le noeud le représentant occupe d'espace et plus il est coloré en rouge. Les trois relations entre utilisateurs sont représentées par des liens orientés entre les noeuds utilisateurs.

La question sous-jacente de la structuration du lien social par le partage est celle de la nature du lien social qui unit les participants. Est-il fondé sur une communauté d'intérêts (musiques et vidéos partagées), sur une communauté réelle (les photographies) ou une communauté d'activités communes (les groupes) ? Nous avons retenu 3 critères de l'identité calculée : le nombre de groupes partagés, le nombre de photos taggées avec le nom de l'autre, le nombre d'amis communs.

5.4.3 Résultats

Les nœuds représentant les amis ont été classés sur la ligne périphérique des amis sur chaque graphe en fonction de leur profil déclaratif (des "hypervisibles" aux "cachés"). Les nœuds représentant les amis se superposent en fonction du nombre d'amis total. Par exemple, dans 5.11, les graphes a, b et d montrent des profils d'utilisateurs présentant un nombre d'amis élevé, comparé au graphe e dont le voisinage des nœuds est plus marqué.

Intensité du lien social

Cette visualisation permet de mettre en évidence les dominantes du lien social entre les utilisateurs. Les liens étant rendus en transparence, une couleur intense signifie l'intensité du lien donné. Par exemple, le graphe b présente une dominante d'amis communs répartis de façon très homogène sur l'ensemble de la population des amis.

Les hyper-visibles et les activités communautaires

L'analyse de Facebook montrait que les utilisateurs à fort indice déclaratif déployaient plus d'activités communautaires collectives (appartenance à des groupes) ; ce phénomène apparaît sur les graphes : on distingue une dominante d'arêtes jaunes vers des profils d'amis "hypervisibles" et une zone calculée plus transparente vers des profils à indice faible dans le graphe a de la figure 4 et le graphe c. Le graphe c présente une dominante de groupes communs et un faible nombre d'amis : on peut supposer que le profil a été récemment créé et que l'initiative de cette création est l'appartenance à un groupe par lequel l'utilisateur a retrouvé des connaissances.

Les tags

Les graphes font davantage ressortir le lien intime qui unit les personnes : les tags correspondent à des personnes qui se voient "en vrai", de nombreux amis communs le suggèrent aussi. Les tags dans Facebook sont utilisés pour indiquer la présence d'une personne dans une photo et sont donc relatifs à des individus. Les tags sont des outils utilisés majoritairement par les adolescents [Lenhart et Madden, 2007], de fait, c'est sur le graphe b, d'un profil de lycéenne, que les tags apparaissent plus nombreux ; les amis représentés font partie de sa classe qui a créé un groupe auquel tous appartient : cet état de

fait est très manifeste dans le graphe dans lequel on voit que les liens sont de constitution homogène et répartis de façon régulière vers l'ensemble des amis.

5.4.4 Conclusion

Nous écrivions dans l'article les conclusions suivantes :

- Les graphes "nœud-lien" permettent à la fois de visualiser quantitativement le phénomène et de révéler la nature du lien social, fondé sur une création de liens comme les médias partagés, les tags, ou à une appartenance commune comme les groupes communs. Nous avons ainsi montré les enjeux très forts de l'identité agissante dans Facebook comme outil de régulation du lien social.
- L'identité calculée incite les utilisateurs à établir des comparaisons entre eux et à nourrir leur représentation, mais il ne s'agit pas pour autant de jouer à accroître démesurément un critère comme le nombre d'amis pour jouer le jeu social : les graphes nœud-lien montrent que le réseau s'entretient par le partage. Partager des photographies, participer à des groupes communs fait partie des échanges sociaux réguliers qui forment ainsi une sorte de nouveau langage intersubjectif.

Ce travail réalisé à la marge du projet Intermed nous a permis de souligner et préciser cet aspect de la structuration de réseaux sociaux : la définition de relations typées entre individus en fonction des comportements des individus les uns par rapport aux autres.

Notre jeu de données dans cette expérimentation est centré sur quelques utilisateurs. Ce choix est lié au fonctionnement du moissonneur d'identités Facebook. Ce jeu de données permet d'étudier le profil d'un utilisateur particulier et ses relations avec ses amis mais ne nous fournit pas les relations entretenues entre ses amis. Le graphe 3.e. présente les relations entre 8 utilisateurs et leurs amis mais peu de relations entre leurs amis, des liens relatifs à des groupes et des amis communs.

A terme, nous souhaitons caractériser l'ensemble des relations entre un ensemble d'utilisateurs dans un outil de débat. A partir de ce nouveau jeu de données, nous pensons appliquer un algorithme de fouille de données afin de pouvoir caractériser des types de profil et ainsi regrouper sous forme de clusters de comportement les utilisateurs ayant un même type de profil.

Nous avons opté pour une visualisation 2D : les informations sont retranscrites en utilisant les indices visuels de taille et de couleur. Un passage à une visualisation 3D et l'utilisation d'autres indices visuels (distance, saturation, forme, volume, texture) permettrait de représenter plus d'informations sur l'identité déclarative des utilisateurs.

5.5 Argumentea : Prototype actuel

Nous présentons dans cette section la version actuelle de l'outil d'annotation développé durant la thèse. Nous avons choisi de nommer cet outil Argumentea en référence au

protocole Annotea.

Nous avons travaillé dans le cadre de l'ANR Intermed avec plusieurs sociologues qui sont demandeur de données rapidement. Nous avons donc fourni des exports dans des formats leur correspondant et mettant en avant les indicateurs qu'ils souhaitaient observer. En parallèle, nous avons entamé une réflexion profonde de la représentation des données dans leur intégralité facilitant l'interopérabilité. Les données demandaient concernent le contenu des discussions : Qu'est-ce qui est dit ? Qui est auteur ? Quand est écrit le Post ? A quels autres Posts est-il relié ? Nous discutons de ces données et plus dans la section traitant des données des débats 2.0.

5.5.1 Choix d'implémentation

Le projet ANR Intermed définit lui-même un choix de technologies Web reposant sur l'usage de technologies JAVA Web. Les composants développés par les différents partenaires doivent s'intégrer aisément dans une plateforme commune. Le choix a donc été d'adopter la technologie Portlets et Servlets ainsi que le conteneur JBoss Portal. Mais aussi, afin de faciliter l'interfaçage entre les différents composants, il était décidé de mettre en place des exports de données par le biais de Web services. Les formats d'échanges étant à définir en fonction des différents composants et services.

Nous avons développé l'ensemble de nos composants sous forme de Portlets en respectant la JSR 286 permettant la communication entre Portlets par le biais d'un système d'événements en AJAX. Cette solution permet de rapidement développer des interfaces Web dynamiques utilisant la technologie AJAX de façon transparente pour le développeur. La plupart des données sont stockées en base de données. Nous avons utilisé le framework Hibernate de gestion de la persistance d'objets Java en base de données. Là encore, le choix d'Hibernate a été fait pour simplifier le déploiement de l'application sur des architectures variables. Hibernate permet notamment de faciliter le déploiement sur n'importe quel gestionnaire de base de données par le biais de fichiers de configuration. Les données de l'application sont intégralement disponibles par le biais de servlets assurant les exports RDF. Nous proposons aussi d'interfacer certains services comme la création d'annotation par le biais de Web Services SOAP, nous utilisons l'implémentation Axis. Nous nous inspirons du protocole Annotea qui définit un protocole de communication client-serveur basé sur HTTP pour créer, stocker et accéder à des annotations.

5.5.2 Plateforme évolutive

La définition des formulaires de saisie et le choix de la technologie employée est le résultat d'une démarche expérimentale incrémentale. Nous nous sommes pour l'instant concentrés sur l'annotation socio-sémantique de segments de textes. L'enjeu primordial est pour nous de trouver une solution non contraignante pour les utilisateurs et qui nous permette de structurer la discussion.

Les différents prototypes permettent aux utilisateurs de préciser leur opinion par des formulaires à choix multiples. L'utilisateur dans nos expérimentations peut choisir parmi des formulaires d'annotation d'exprimer sa position en regard du débat : "commentaire", "confirmation", "opposition", "interrogation", "vote" et "synthèse". A l'intérieur des quatre premiers, l'utilisateur est invité, s'il le souhaite, à utiliser deux types d'annotation à choix multiples : l'annotation de l'humeur ("furieux", "indifférent", "satisfait", "sceptique") et l'annotation du jugement ("je m'interroge", "je ne suis pas d'accord", "je suis d'accord"), la seconde étant partiellement redondante avec le choix du formulaire.

Nous avons donc entrepris une série d'expérimentations visant à évaluer l'influence de l'annotation sur la participation au débat, et la pertinence de l'annotation avec des champs prédéfinis : est-elle utilisée ? Lorsqu'elle l'est, les utilisateurs y ont-ils recours de façon judicieuse et cohérente ? Une autre source de questionnements concerne le type d'annotations à choix multiples ou semi-structurées, pertinents et utiles pour l'expression de son positionnement dans le débat. Quels types d'annotations à choix multiples sont les plus utiles ? Sont-elles utilisées pour exprimer des postures nuancées ? Une troisième source de questionnements concerne la richesse de la discussion. Permet-elle aux utilisateurs de participer plus facilement aux discussions ? L'utilisation d'annotations à choix multiples est-elle compatible avec la production d'annotations en texte libre, de fait plus argumentées ?

Ces expérimentations ont été conduites auprès d'élèves de l'enseignement secondaire (lycée, collège) et d'étudiants de l'enseignement supérieur (école d'ingénieur ENST-ParisTech, universités). L'expérimentation présentée ci-après rend compte de l'expérimentation conduite en école d'ingénieurs, sur un échantillon composé de 120 étudiants.

5.5.3 Expérimentations Argumentea

Méthodologie de débat

Notre méthodologie de débat est simple. Elle consiste à définir un ensemble de documents de références servant de base aux annotations. Ces textes doivent de préférence être courts et favoriser l'émission d'annotations. En d'autres termes, les textes doivent exprimer des opinions bien marquées et différentes ou amener des éléments d'information pouvant servir de références dans une argumentation. L'étape suivante consiste en la création de formulaires d'annotation permettant de définir les modalités d'expression dans le débat.

Dans le cadre de nos expérimentations, nous choisissons de préférence des utilisateurs habitués à l'usage du Web pour l'expression d'opinions. Nous avons essentiellement expérimenté auprès de lycéens ou d'étudiants d'université. Nous souhaitons mesurer la qualité des débats en terme de retour dans l'outil d'annotation. Nous espérons que la solution de l'annotation permette de mieux structurer le débat. Les utilisateurs devraient donc plus facilement lire les participations des autres que dans les classiques blogs, forums ou wikis.

Description d'une expérimentation

Le débat s'est déroulé en deux phases manifestant deux degrés d'engagement. Lors d'une première phase, les participants ont été invités à annoter trois textes portant sur le thème "Nanotechnologies et société", après avoir été initiés à cette thématique par une conférence destinée à sensibiliser les étudiants aux questions soulevées par les textes et à mettre en perspective la nécessité de réfléchir à un moratoire sur les nanotechnologies. A la fin de la séance de débat, les étudiants sont invités à participer à la seconde phase de débat en ligne à distance s'ils le souhaitent, selon une modalité non contrainte (les étudiants peuvent intervenir quand ils le souhaitent et d'où ils le souhaitent pendant un mois). Un message électronique de rappel leur est envoyé.

Trois textes ont été proposés à annotation et à discussion, présentant des tonalités et des points de vue différents sur la thématique. En 1h30, les étudiants doivent prendre connaissance du texte et l'annoter.

Nayla Farouki: Quelques réflexions sur l'homme augmenté...

4 mars 2010
Par nanosciences

On pourrait déjà dans un premier temps se demander ce que signifie un être humain augmenté. En effet, l'outil - quel qu'il soit - augmente forcément la personne qui l'utilise en ce sens qu'il lui donne des moyens qu'elle n'aurait pas sans lui. Quels moyens? Prenons quelques exemples :

- a - suis-je augmenté(e) grâce à mes lunettes sans lesquelles je ne vois rien?
- b-suis-je augmenté(e) grâce à mon vélo, ma voiture, mon bateau, mon avion qui me permettent de me déplacer d'une manière par ailleurs impossible?
- c - Suis-je augmenté(e) en raison de la prothèse qui vient remplacer mon col du fémur cassé?
- d - ma mémoire est-elle augmentée grâce à mon carnet de notes?
- d-plus hasardeux : mon intelligence est-elle augmentée du fait que je peux comprendre des mathématiques que Thalès ou Pythagore n'avaient jamais imaginées? ou encore - ma liberté est-elle augmentée du fait que j'ai de loin plus de choix dans les Grands Magasins qu'au temps de Zola?

Auteur: Nayla Farouki


Je m'interroge
Il semble que deux choses soient présupposées ici :
1- la liberté de l'homme est augmentée par plus de choix dans les Grands Magasins...
2- un homme dont la liberté est augmentée est un homme augmenté

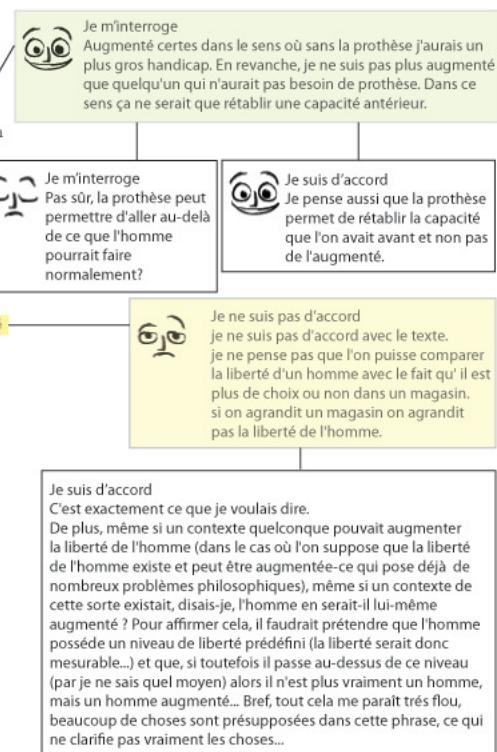


FIGURE 5.12 : Un exemple de discussions via l'annotation

5.12 présente un extrait de discussion produit au cours d'une de nos expérimentations. Il est possible d'utiliser les humeurs, les termes ("je suis d'accord", "je m'interroge", ...), et

les champs de saisie de façon complémentaire ou indépendante. Un message peut avoir plusieurs réponses. Le premier message d'une discussion est ancré dans un texte.

Analyses

Nous présentons ici plus en détail l'une des expérimentations réalisées auprès d'étudiants. Nous reprenons les résultats présentés dans [Seilles *et al.*, 2010a].

Participation au débat : 3 phases de défection

La procédure d'annotation présente plusieurs phases successives d'engagement. Lors de la phase 1, les 120 étudiants sont réunis dans une salle de cours (échantillon complet). Pour débattre, ils sont invités à s'inscrire (1er stade d'engagement), produire une annotation (2ème stade d'engagement) et à participer de façon non contrainte pendant 1 mois (3ème stade d'engagement).

Les résultats montrent en premier lieu une défection de 20,9% des étudiants au débat en ligne avant même l'inscription. Sur 120 étudiants, 95 seulement se sont identifiés sur le site. Cette participation, jugée faible compte tenu de la dimension confinée et contrainte de la première phase et du contexte (les étudiants étant invités à débattre dans le cadre d'un cours) peut s'expliquer par un faible intérêt pour la thématique. Ce faible taux d'inscription peut également s'expliquer par les problèmes d'ergonomie. Pour favoriser la participation lors de cette première phase de défection attendue, il est donc nécessaire de stimuler l'intérêt des participants, par exemple en développant le dispositif de formation préalable, en rendant l'interface plus attrayante.

Une défection de 50,6% est observée au second stade d'engagement : sur les 95 étudiants ayant créé un compte, 47 produisent une annotation. Une défection de 73,7% est observée au troisième stade d'engagement : 13 étudiants seulement sur les 47 se reconnectent pour poursuivre le débat en ligne. La moyenne d'annotations par personne, qui double du 2ème au 3ème stade d'engagement, manifeste que les participants qui se sont reconnectés librement sont plus engagés dans le débat que les participants en présentiel, conformément à ce qui était attendu. 5.13 présente les taux de participation selon les différentes phases d'engagement.

On peut supposer que le motif le plus répandu de la défection des participants est l'absence de remarque ou commentaire à exprimer sur les textes. Hormis l'amélioration ergonomique et graphique du dispositif destinée à le rendre plus attrayant, une solution pour améliorer le taux de participation est donc de permettre aux utilisateurs d'exprimer à moindre coût leur opinion et arguments, par l'annotation à choix multiples.

Des formulaires d'annotation utiles

L'annotation à choix multiples est-elle utilisée et comment ? On peut juger de l'utilité d'un type d'annotation en observant si elle est ou non utilisée. Comme exposé précédemment, trois types de formulaires ou champs d'annotation à choix multiples étaient propo-

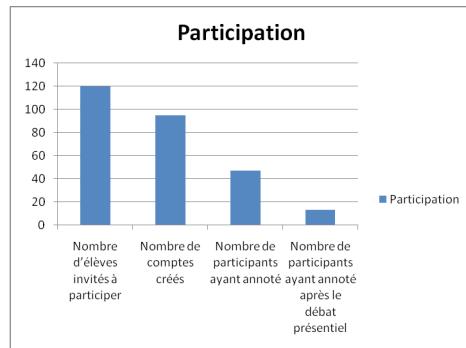


FIGURE 5.13 : Taux de participation au débat à chaque phase d'engagement

sés : un formulaire de prise de position, un champ mentionnant l'humeur et un champ mentionnant le jugement, volontairement redondant avec le formulaire.

Sur 151 annotations, 138 (soit 91,4%) recourent à un formulaire de prise de position. 116 annotations mentionnent l'humeur (76,8%), 136 mentionnent le jugement (90,1%). Ces trois annotations à choix multiples sont donc utilisées dans la grande majorité des annotations, ce qui tend à montrer leur utilité.

On peut s'étonner que le formulaire exprimant le jugement, étant redondant avec les formulaires de prise de position, ait été plus utilisé que la mention de l'humeur (utilisée dans les trois-quarts des annotations). Neuf dixièmes des annotations mentionnent de façon redondante le jugement dans un formulaire de prise de position qui exprime lui-même un jugement. Cette redondance manifeste l'utilité de cette information, au point que sa redondance ne soit pas jugée inappropriée.

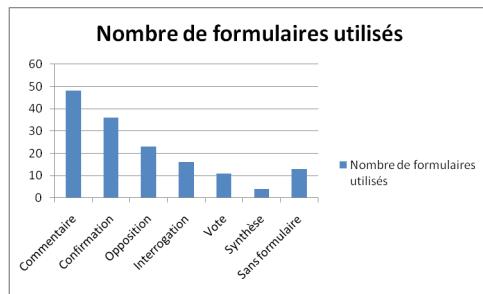


FIGURE 5.14 : Nombre de formulaires d'annotations utilisés par catégorie

La figure 5.14 présente le nombre de formulaires d'annotations utilisés pour chaque type de formulaire.

Une utilisation des champs judicieuse

Lorsqu'elle est utilisée, les utilisateurs ont-ils recours à l'annotation à champs multiples de façon judicieuse et cohérente ? Il n'est jamais observé d'annotation utilisant le jugement "je suis d'accord" et faisant mention d'un désaccord par le biais d'un texte libre. C'est plutôt l'adéquation entre le commentaire saisi et le jugement qui prévaut comme règle d'utilisation des formulaires. Cette observation est valable pour tous les jugements proposés. Ce type d'annotations n'a donc été pour aucun des utilisateurs l'opportunité de participer en remplissant les champs au hasard et de participer de ce fait au débat de façon injustifiée.

L'expression des nuances par des champs à choix multiples

Les résultats présentés dans le paragraphe précédent s'appuient sur la correspondance sémantique des annotations à choix multiples redondantes pour montrer leur utilisation judicieuse par les participants ; à l'inverse de cette démarche, l'annotation de l'humeur exprime une posture différente et plus affective du participant : sa mention vise à différencier le raisonnement de l'affect.

Toutefois, on peut se demander si cette différence est perçue par les participants, et s'ils l'utilisent pour exprimer des postures contrastées. On pourrait attendre de l'expression nuancée d'une posture dans le débat que le dispositif permette, par exemple, d'exprimer une opposition sereine ou une confirmation colérique. La capacité de l'annotation à choix multiples à exprimer une posture nuancée peut donc être évaluée par observation de l'utilisation conjointe d'une annotation du jugement et d'une annotation de l'humeur contrastée. L'examen de la correspondance entre l'annotation de l'humeur et du jugement montre qu'à 36,9% seulement en moyenne, les mentions de l'humeur et du jugement s'accordent ("je suis d'accord" et "satisfait", "je ne suis pas d'accord" et "furieux") : en ce cas leur utilisation conjointe n'est pas mise à profit pour exprimer une posture contrastée, ce qui ne déprécie en rien leur utilité. Par contre, 5,4% des annotations expriment une humeur en contraste avec le jugement ("je suis d'accord" et "furieux" et "je ne suis pas d'accord" et "satisfait") : cette utilisation, même résiduelle, à l'image de la fréquence des postures contrastées en général, montre que l'expression des nuances est possible. L'annotation à choix multiples n'est pas contradictoire avec l'expression de postures et de jugements nuancés : elle peut être utilisée au besoin pour exprimer une posture contrastée.

L'annotation à choix multiples comme entrée dans les débats

La possibilité d'annoter un texte et de débattre via des annotations à choix multiples incite-t-elle les utilisateurs à participer au débat, qu'ils aient des réticences ou des difficultés à s'exprimer ou n'aient pas d'argumentation en tête qui justifie leur saisie en texte libre ?

Sur 151 annotations, 44 ne présentent aucun commentaire en texte libre, ce qui représente 29,1% des annotations. Parmi ces 44 annotations, 14 présentent un usage de l'humeur, du jugement, ou des deux (soit 31,8%) : ces fonctionnalités pourraient permettre à

l'utilisateur de s'exprimer de façon embryonnaire ? Pourtant, 19 de ces annotations n'utilisent aucune de ces informations : erreur de manipulation ou utilisation non prévue de l'annotation ? Le fait d'insérer une annotation, même totalement vide, a pour effet de mettre en surbrillance le passage sélectionné : les utilisateurs auraient-ils utilisé ce procédé pour attirer l'attention sur un passage du texte comme ils auraient surligné sur une feuille au stabilo les passages qui retiennent leur attention ? Il serait intéressant d'examiner cette fonctionnalité plus précisément dans une autre expérimentation.

Parmi les 44 annotations sans commentaire en texte libre, 9 annotations présentent une mention du jugement et l'on peut observer un phénomène intéressant : l'ensemble de ces annotations mentionnent "je suis d'accord", à l'exclusion des deux autres jugements proposés ("je m'interroge" ou "je ne suis pas d'accord"), complétés par des annotations en texte libre. On peut donc en conclure que l'annotation du jugement permet d'exprimer une posture de soutien à moindre coût, sans doute parce qu'elle ne fait qu'appuyer ce qui a été exprimé dans l'annotation précédente ou dans le texte annoté : pour exprimer son accord, l'annotation à choix multiples est jugée suffisante. En revanche, pour exprimer son désaccord ou son interrogation, les annotations à choix multiples ne sont pas jugées suffisantes mais complétées par des commentaires en texte libre qui motivent cette posture de désaccord ou d'interrogation.

Choisir des jugements pré-remplis et produire un argumentaire

Les utilisateurs qui utilisent les champs libres utilisent-ils les champs d'annotation à choix multiples ? Sur 107 annotations avec commentaire à texte libre, 104 utilisent soit la mention de l'humeur soit celle du jugement, et pour renforcer ce résultat, sur les 3 annotations qui n'utilisent ni l'un ni l'autre, une est une insulte, et les deux autres ne contiennent pas une confirmation simple par texte libre ("exact, je suis d'accord") : ce chiffre montre que l'utilisation d'annotations à choix multiples est pleinement conciliable et complémentaire de l'argumentation en texte libre.

Cette expérimentation a montré que l'annotation à choix multiples est utile dans le débat en ligne et peut même être jugée suffisante par les utilisateurs pour appuyer ou confirmer un propos préalablement exprimé. Tandis que le désaccord ou l'interrogation ont systématiquement fait l'objet de commentaires en texte libre, les points du débat qui font l'objet d'accord ou de consensus ne donnent que rarement lieu à l'expression de cet accord en texte libre : l'annotation à choix multiples permet alors d'exprimer son accord comme un participant opinerait de la tête. Elle s'apparente à un vote positif pour le point sur lequel elle porte.

Un autre résultat est la non-contradiction entre annotation à choix multiples et annotation en texte libre : l'annotation à choix multiples n'est pas seulement une façon de participer à moindre coût mais entre dans une dynamique de construction du jugement, de sorte que les argumentaires construits utilisent tous les choix multiples en complément du texte libre.

5.5.4 Utilisation d'Argumentea par des tiers

Le site jeux2debat¹⁴ répertorie un ensemble d'expériences menées à la marge d'Inter-Med dont le débat mené au LIRMM préparant le changement de direction¹⁵. Les débats présentés sur ce site nous ont permis de tester plusieurs modalités de débats en présence complétés par des débats en ligne. Ils nous ont permis d'utiliser plusieurs outils comme LogiLogi, TalkMap et bien sur Argumentea.

En particulier, le débat sur la télé médecine à Madagascar a été l'occasion de laisser un groupe utiliser Argumentea sans notre présence. Il y a eu quelques échanges de mail pour l'installation des Portlets mais aucune intervention de notre part. Les utilisateurs, 25 inscrits ont produit 102 annotations en 6 heures d'utilisation (30 minutes de présentation par la personne ayant installé l'outil auprès des étudiants et une utilisation libre le reste de la journée).

Nous pensons que Argumentea est aujourd'hui dans un état suffisamment avancé pour pouvoir supporter des expérimentations de tiers. L'interface a été évaluée par Intactile Design et une proposition d'amélioration a été faite. Il faudrait mettre en place cette nouvelle interface et il y aurait probablement encore des améliorations à apporter aux exports de données, mais la fonctionnalité d'annotation de ressources est fonctionnelle.

14. <http://www.jeux2debat.net/>

15. http://www.jeux2debat.net/debat_lirmm.html

Préambule

Ce chapitre présente les conclusions et perspectives du travail de thèse.

Sommaire

6.1	Reformulation de la problématique	177
6.2	Perspectives	180

6.1 Reformulation de la problématique

Le travail de thèse ici présenté porte sur l'usage d'annotations socio-sémantiques afin de structurer des débats en ligne. La conception du système développé prend en compte l'importance de la dimension sociale de l'application afin de faciliter une analyse des réseaux sociaux centrés sur les interactions prenant place dans les débats entre les citoyens. Cette dimension impacte notamment sur la définition d'indices pouvant permettre de qualifier des comportements et d'organiser les citoyens en groupes d'opinion.

Ce travail fait face aux verrous suivants identifiés dans les débats 2.0 :

- Les débats 2.0 étant des débats à grande échelle faisant intervenir un grand nombre de participants et de sujets produisant un grand nombre de données, la problématique héritée du Web 2.0 de la navigation dans un grand nombre de données est présente dans une application de débats 2.0.
- Les débats, et plus largement la concertation citoyenne, ont pour objectif d'éclairer des prises de décision. Il est même défendu par certains que si ces débats ne produisent pas de prise de décision clairement explicitées aux citoyens, la concertation

aura pour effet négatif de désintéresser les citoyens et de les rendre plus difficilement mobilisables dans les prochains processus de concertations. Il s'agit donc de considérer une application de débats 2.0 comme étant un élément d'un processus, élément donc à l'interface d'autres applications. Il faut donc penser les données des débats 2.0 comme pouvant être consommées par des applications tierces. Il s'agit d'identifier les bonnes données, les organiser et les représenter dans des formats facilement intégrables par d'autres applications.

- L'insularisation des groupes d'opinion est un fait avéré. L'adage "qui se ressemble s'assemble" semble particulièrement vrai dès qu'il s'agit d'opinions politiques. L'un des phénomènes observés sur le Web est une forme de durcissement des opinions au sein des groupes qui ont tendance à s'isoler des autres, à se fermer à toute discussion avec des avis divergents. Ce phénomène tend à durcir les débats et les rendre stériles. Ajoutons que le choix d'un outil de discussion basé sur l'annotation implique qu'il a fallu tester son acceptabilité auprès des utilisateurs.

Notre approche a consisté en la mise en place d'une boucle de conception Agile d'une application de débats 2.0. Ce processus itératif de conception et d'implémentation fait intervenir des phases de conception ou le recueil d'avis d'experts est primordial, des phases d'implémentation de prototypes et enfin des phases de tests ou expérimentations sur le terrain permettant de valider ou invalider des choix de fonctionnalités, d'ergonomie et de design. Cette approche a permis de tester l'acceptabilité de l'annotation auprès d'un public particulier, les jeunes utilisateurs d'Internet. Aujourd'hui, nous sommes convaincus de l'acceptabilité de l'annotation au moins en tant qu'outil de discussion. Cette conviction vient de l'observation sur le terrain du bon usage de l'annotation mais aussi du fait que la fonctionnalité d'annotation, peu présente sur le Web il y a 3 ans, est aujourd'hui une fonctionnalité bien installée dans plusieurs outils grand public.

L'expression d'opinions, ou le recueil d'avis d'internautes est aujourd'hui une pratique courante. Le "*i like*" de facebook ou les systèmes de modération au mérite en sont probablement des exemples pionniers. L'absence d'expression d'opinions négatives dans la plupart des applications grand public tient probablement à la visée grand public des applications et à la volonté donc de créer une communauté ou la bonne entente est de mise. Il est plus difficile par exemple dans Facebook de dire que l'on aime pas un commentaire que de dire qu'on l'aime. Soit on aime soit on signale un contenu malveillant à un modérateur. Pour exprimer son désaccord, il faudrait alors rédiger un commentaire textuel soi-même... L'effort est bien plus important que de cliquer sur un bouton. Cette considération de bonne entente n'a pas sa place dans les débats en ligne, tout du moins pas à l'origine d'une discussion. Il est important que des avis tranchés et opposés puissent être énoncés. Il ne s'agit pas de semer la zizanie mais bien de permettre aux participants d'un débat de se positionner. Ce mécanisme est classique dans les formes de concertation grand public adopté jusqu'ici. Il nous semblait naturel d'en faire l'élément de structuration de base des débats 2.0. Nous avons pu mesurer non seulement l'acceptation et le bon usage d'un certain nombre d'indicateurs d'opinions (nous parlons de tags de jugements ou de tags d'opinions ou en-

core de modalités) mais aussi l'existence d'indicateurs à l'usage plus marginal, réservé à des groupes plus restreints d'utilisateurs. Il est aujourd'hui possible dans notre système de créer manuellement de nouveaux indicateurs d'opinion. Cette fonctionnalité nous semble primordial dans de futurs usages ou la gestion des communautés sera d'avantage au cœur de l'application. Il serait par exemple possible pour un groupe d'avoir son propre jeu d'indicateurs d'opinions.

Nous avons donc intégré l'expression d'opinions comme élément structurant les débats. Mais les jugements et opinions ne sont pas la seule donnée structurante des débats. Les liens entre les ressources, leurs dates de création, le lien à l'auteur et son activité ainsi que son réseau d'accointances, ou encore toutes les données de catégorisation sont également des données structurantes qu'il faut pouvoir échanger entre une application de débats 2.0 et les autres applications du processus démocratique. Nous avons pu nous baser sur des formats déjà existants de représentation de données en RDF concernant le *tagging* et la catégorisation, la création de contenus par une communauté, ou encore la représentation de l'identité de l'utilisateur et de son réseau social. A partir de quelques vocabulaires déjà existants et quelques extensions que nous avons développé selon nos besoins plus spécifiques, le format de données dédiés aux débats 2.0 est posé. De plus, il est facilement adaptable en se basant sur les mêmes mécanismes d'extension que ceux que nous utilisons et qui font la force de RDF. Non seulement la définition d'un tel format permet de garantir l'interopérabilité, mais il permet de mettre en place des indicateurs permettant de mieux indexer l'information. L'utilisation de RDF s'adresse donc à la problématique de l'interopérabilité mais également à la problématique de la navigation dans la grande quantité de données.

Concernant l'insularisation des groupes d'opinion, nous avons commencé à déblayer le terrain pour une analyse du réseau social intrinsèque au débat. Nous avons adopté un format de données permettant d'accorder une part importante au réseau et à l'activité, les interactions sociales, dans la représentation des données.

Notre approche concerne donc la mise en place d'outils de concertations venant après une phase d'information et précédent une prise de décision. Notre démarche s'est essentiellement concentrée autour des points suivants : l'interopérabilité, la structuration par le biais d'opinions, la dynamique du réseau social intrinsèque aux débats. L'interopérabilité est fondamentale pour qu'une application s'insère aisément dans un processus pouvant faire intervenir plusieurs applications, pour faciliter la réutilisabilité des données. L'expression d'opinions est un mécanisme naturel et classique dès qu'il s'agit d'argumenter. Ne pas considérer le réseau social dans le cadre d'un débat en ligne serait contraire aux bonnes pratiques du Web actuel et c'est se priver d'informations primordiales pour évaluer la qualité des débats. L'application développée s'appuie sur l'utilisation d'annotations socio-sémantiques permettant aux utilisateurs de discuter, d'exprimer leurs opinions, de lier des fragments de ressources et de produire un ensemble de métadonnées riches facilitant l'indexation et donc la navigation dans la masse des données. Les formats adoptés pour représenter ces métadonnées se basent sur des outils standards du Web séman-

tique et permettent de faciliter la compréhension et donc la réutilisabilité des données par des programmes. Ces standards étant déjà utilisés par ailleurs, ceci permet d'imaginer des usages ou l'application d'annotation serait couplé à d'autres outils produisant des données dans ces mêmes formats. Enfin, notre réflexion sur l'analyse des réseaux sociaux a fortement guidé la définition d'indices de l'activité qui nous semblent primordiaux pour mettre en place une analyse fine du réseau ou les acteurs et ressources stratégiques pourront être plus aisément identifiés. Notre corpus argumentatif est fortement structuré par les opinions exprimées par les utilisateurs et la structuration du réseau social repose fortement sur des relations de support et d'opposition.

Nous avons donc proposé un outil de débats permettant une structuration des discussions facilitant l'indexation des contenus mais aussi facilitant la définition de relations d'opposition et de soutien entre les participants aux débats. Cet outil, même s'il reste un prototype expérimental, nous a permis de valider des formats de représentation des données des débats afin de favoriser la réutilisation des données de la concertation dans d'autres outils.

6.2 Perspectives

6.2.1 Perspectives théoriques

Analyse des débats

Nous avons manqué au cours de nos expérimentation d'un critère de qualité des discussions. La littérature en propose plusieurs, encore faut-il pouvoir l'implémenter. Le taux de *flame* ou insultes est un indice fréquent mais contesté (certains défendent qu'il faut des débats polis pour avoir des argumentations de qualité, d'autres au contraire estiment qu'il faut que ça chauffe...). La variabilité des arguments émis en est un autre mais que l'on peut difficilement implémenter.

Nous proposons comme méthode d'analyse de considérer l'ensemble des lectures des ressources produites dans les débats. Nous mesurons le ratio entre le nombre moyen de personnes ayant lu chaque ressource et le nombre de participants au débat. Soit la_i le nombre de personnes ayant lu l'annotation i . A l'ensemble des ressources produites dans l'application de débats, ou ensemble des annotations. P l'ensemble des participants au dé-

bat. Notre fonction de calcul de la qualité des débats est la suivante :
$$\frac{\sum_{i=0}^n la_i / |A|}{|P|}$$

Ce ratio permet de mesurer la couverture de la lecture par les participants des productions des autres participants. En d'autres termes, on mesure si tout le monde prend le temps de lire la participation de tout le monde. Plus le ratio tend vers 1 plus la qualité du débat est bonne selon ce critère de lecture. Cette mesure est tout de même à nuancer en prenant en compte la quantité de production et la quantité de ressources malveillantes (*flame*, *troll* ...) produites.

6.2.2 Perspectives applicatives

L'utilisation d'Argumentea pour des concertations publiques

Aujourd'hui notre application d'annotation socio-sémantiques pour des débats 2.0 que nous nommons Argumentea est distribuée en licence libre (LGPL plus précisément). Elle a fait l'objet d'une analyse ergonomique et une proposition d'interface plus adaptée a été réalisé par la société Intactile Design (cette proposition est en annexe). A peu de frais, l'application Argumentea telle qu'elle est aujourd'hui pourrait atteindre un niveau d'ergonomie de bonne qualité. Son usage simplifié pourrait en faire une bonne solution pour des concertations publiques de grande échelle. Cependant, il reste nécessaire d'implémenter certaines fonctionnalités de base qui sont hors du cadre de cette thèse. La gestion des groupes par exemple serait une fonctionnalité incontournable pour une exploitation de l'outil. La définition des profils d'utilisateur est dans l'état actuel très basique... Et bien entendu, la mise en place de services annexes comme un système de recommandations, ou encore de visualisation de la dynamique du réseau social seraient des éléments forts de motivation à l'usage de l'application.

L'expression d'opinions et les formats de données dans l'existant

Maintenant, sans vouloir vendre à tout prix l'application Argumentea, il reste incontestable que l'adoption de systèmes simple d'expression d'opinion au cœur d'outils de discussions "classiques" comme des blogs ou des forums permettrait de produire des synthèses, des solutions de modération efficaces. L'adoption de formats de données standards, des technologies du web sémantique, permettrait de croiser les données et de développer des services d'agrégation et de filtre des concertations en ligne déjà mis en place facilitant la tâche des collectivités territoriales.



Glossaire

Préambule

*Dans ce glossaire, par soucis de clarté, nous allons définir les termes et acronymes aux-
quels nous faisons référence dans ce document.*

ANR

ANR : L'Agence Nationale de la Recherche¹. L'ANR propose un moyen de financement de projets de recherche sur présentation de dossier.

Dublin Core

Le vocabulaire Dublin Core est écrit en RDF, il est consacré à la représentation de métadonnées concernant le contexte de production d'un document. Par exemple, le vocabulaire permet de représenter la relation entre un document et ses auteurs, la date de publication ...

FAI

FAI est l'acronyme signifiant Fournisseur d'Accès à Internet.

1. <http://www.agence-nationale-recherche.fr/>

FOAF

FOAF pour Friend Of A Friend est un vocabulaire RDF dédié à la représentation d'un individu et de son réseau social².

Folksonomie

Le mot Folksonomie est né de la contraction des termes Folks et Taxonomie. Une folksonomie est une taxonomie créée par un ensemble d'individus non nécessairement experts du domaine. C'est une forme d'indexation partagée par un groupe d'individus.

Insularisation

On parle d'insularisation des groupes d'opinion pour souligner une tendance de ces groupes à s'isoler et se refermer sur eux-mêmes plutôt que de s'ouvrir aux débats en discutant avec des individus d'opinions différentes.

Intermed

Le projet Intermed est le projet ANR ayant permis de financer nos travaux de recherche. Celui-ci regroupe des chercheurs et des entreprises de plusieurs domaines de l'informatique et des sciences sociales. Son objectif est la production d'outils permettant d'améliorer la concertation publique dans le cadre de gestion de biens communs.

LIRMM

Le LIRMM est le Laboratoire d'Informatique de Robotique et de Micro-électronique de Montpellier.

Mashup

Un Mashup est l'action de récolter des données provenant de plusieurs sources, de les croiser et les exploiter pour proposer une nouvelle application. Il s'agit bien souvent de réutiliser des données existantes dans un contexte non prévu à l'origine de la création de ces données. Le vocabulaire RDF et la philosophie d'ouverture des données prônée par de nombreux acteurs du Web Sémantique va dans ce sens : faciliter la réutilisation de données pour favoriser la création de nouveaux usages.

2. <http://www.foaf-project.org/>

Méta Données

Une métadonnée est une donnée ajoutée. C'est une donnée sur une donnée comme par exemple le titre d'un document ou sa date de production. Ce n'est pas normalement l'information primordiale mais c'est une information qui peut être particulièrement utile pour indexer un document par exemple.

Méta Langage

Un métalangage est un langage fournissant toutes les primitives nécessaires pour se décrire lui-même et pour écrire d'autres langages. RDF est un métalangage car toutes les primitives de RDF sont écrites en RDF et RDF permet d'écrire des vocabulaires, ou ontologies. Un vocabulaire ou ontologie est une description de classes et de relations entre classes qui permet de décrire des données et relations entre données. Par exemple le vocabulaire FOAF définit la classe Personne et la relation know pour lier des personnes. A partir de ce vocabulaire on peut par exemple représenter la collection d'amis d'un utilisateur Facebook. FOAF est donc un format de données et FOAF est écrit en RDF.

MOAT

MOAT pour Meaning Of A Tag est un vocabulaire RDF dédié à la représentation des Tags et de leurs définitions. Ce projet est initié lors du travail de thèse de A. Passant.

NiceTag

NiceTag est un vocabulaire RDF dédié à la représentation de la portée, du sens, des actions de Tagging. Les auteurs de NiceTag l'ont conçu en complémentarité au vocabulaire MOAT dans le but de pallier aux limites de MOAT.

PageRank

PageRank est l'algorithme à l'origine de Google permettant d'indexer les résultats d'une recherche. A la base cet algorithme est basé sur le nombre de liens pointant vers une ressource pour mesurer l'importance de cette ressource.

Post

Un post ou billet ou commentaire est un message laissé sur un site de discussion type blog ou forum.

SemSNA

SemSNA pour Semantic Social Network Analysis est un vocabulaire RDF dédié à la représentation des résultats d'analyse d'un réseau social. Ce projet est initié durant la thèse de Guillaume Erétéo.

SIOC

SIOC pour Semantically Interlinked Online Communities est un vocabulaire RDF dédié à la représentation de l'activité d'une communauté en ligne, notamment la production de Post dans des Blogs ou Forum.

Socio-sémantique

Le Web socio-sémantique est l'alliance des pratiques Web 2.0 dites sociales et des technologies du Web sémantique. Il s'agit de faire produire des méta-données par les utilisateurs lors de la création de contenus par ceux-ci.

SocNet

Le terme est la contraction de Social Network. Il désigne les réseaux sociaux numériques (Facebook par exemple).

STIC

STIC est l'acronyme pour Sciences et Technologies de l'Information et de la Communication.

TIC

TIC est l'acronyme pour Technologies de l'Information et de la Communication.

Tweet

Un tweet est un message court, 140 caractères maximum émis dans l'application Twitter.

W3C

Le W3C pour World Wide Web Consortium est un groupe constitué pour promouvoir les technologies du Web et améliorer le Web. Le W3C supporte notamment HTML et RDF.



Bibliographie

- M. AGOSTI et N. FERRO : A formal model of annotations of digital content. *ACM Trans. Inf. Syst.*, 26, November 2007. ISSN 1046-8188. URL <http://doi.acm.org/10.1145/1292591.1292594>. Cité pages 8, 98, 99, 126 et 127.
- G. ANTONIOU, D. BILLINGTON, G. GOVERNATORI et M.J. MAHER : A flexible framework for defeasible logic. In *American National Conference on Artificial Intelligence*. AAAI, 2000. Cité page 151.
- G. ARTIGNAN et M. HASCOËT : From coding to automatic generation of legends in visual analytics. In *ICEIS*, pages 404–417, 2010. Cité page 165.
- J.L. AUSTIN : *Quand dire c'est faire*. Paris : Seuil, 1970. Cité page 32.
- P. BAGGIA, F. BURKHARDT, C. PELACHAUD, C. PETER et E. ZOVATO : Emotion markup language (emotionml) 1.0. Rapport technique, W3C, 2011. URL <http://www.w3.org/TR/emotionml/>. Cité page 90.
- J. A. BARNES : Class and committees in a norwegian island parish. In *Human Relations*, volume 7, pages 39–58. 1958. Cité page 61.
- C. BERGE : *Théorie des graphes et ses applications*. Dunod, 1958. Cité page 166.
- T. BERNERS-LEE : Information management : A proposal. Rapport technique, 1989. URL <http://www.w3c.org/History/1989/proposal.html>. Cité page 42.
- T. BERNERS-LEE : Basic http as defined in 1992. Rapport technique, 1992. URL <http://www.w3.org/Protocols/HTTP/HTTP2.html>. Cité page 43.

- T. BERNERS-LEE : Uniform resource identifiers in www. rfc 1630. Rapport technique, CERN, 1994. URL <http://tools.ietf.org/html/rfc1630>. Cité page 100.
- T. BERNERS-LEE, R. FIELDING, U.C. IRVINE et U.C. MASINTER : Ietf rfc 2396 : Uniform resource identifiers (uri) : Generic syntax. Rapport technique, 1998. URL <http://www.ietf.org/rfc/rfc2396.txt>. Cité page 44.
- T. BERNERS-LEE, W. HALL, J. HENDLER, N. SHADBOLT et D. J. WEITZNER : Creating a Science of the Web. *Science*, 313(5788):769–771, août 2006. ISSN 1095-9203. URL <http://dx.doi.org/10.1126/science.1126902>. Cité page 60.
- J.G. BIDIMA : *La palabre, une juridiction de la parole*. Michalon, le bien commun, 1997. Cité pages 19 et 72.
- Y. BREINDL et P. FRANCQ : Can web 2.0 applications save e-democracy? a study of how new internet applications may enhance citizen participation in the political process online. *Electronic Democracy*, 1(1):14–31, 2008. Cité pages 28 et 73.
- J. BRESLIN, A. PASSANT et S. DECKER : *The Social Semantic Web*. Springer-Verlag, Heidelberg, 2009. ISBN 978-3-642-01171-9. Cité pages 52, 76 et 126.
- M. BUFFA : Du web aux wikis : une histoire des outils collaboratifs, 05 2008. URL http://interstices.info/jcms/c_37151/du-web-aux-wikis-une-histoire-des-outils-collaboratifs. Cité pages 37, 46 et 47.
- V. BUSH : As We May Think. *Atlantic Monthly*, 176(1):641–649, March 1945. ISSN 1072-5520. URL <http://www.theatlantic.com/doc/194507/bush>. Cité pages 35 et 43.
- A. S. BÉLIARD : Pseudos, avatars et bannières : la mise en scène des fans. *Terrains et Travaux*, 15:191–212, 2009. Cité page 112.
- Y. CASEAU : Efficacité de la communication 2.0 dans l’entreprise, 03 2009. URL <http://organisationarchitecture.blogspot.com/2009/03/efficacite-de-la-communication-20-dans.html>. Cité pages 77 et 78.
- F. CAVAZZA : Qu'est-ce que l'entreprise 2.0?, 07 2007. URL <http://www.entreprise20.fr/2007/07/24/qu'est-ce-que-l'entreprise-20>. Cité page 50.
- C. CHABROL et M. BROMBERG : Préalable à une classification des actes de parole. *Psychologie française*, 44:291–306, 1999. URL <http://cat.inist.fr/?aModele=afficheN&cpsidt=1234225>. Cité pages 21 et 33.
- P. CHARAUDEAU : Le dialogue dans un modèle de discours. In *Cahiers de Linguistique Française*, numéro 17, pages 141–178. Université de Genève, 1995. Cité page 32.

- C. CHESÑEVAR, J. MCGINNIS, S. MODGIL, I. RAHWAN, C. REED, G. SIMARI, M. SOUTH, G. VREESWIJK et S. WILLMOTT : Towards an argument interchange format. *Knowl. Eng. Rev.*, 21:293–316, December 2006. ISSN 0269-8889. URL <http://portal.acm.org/citation.cfm?id=1187936.1187940>. Cité page 89.
- J. CLÉMENT : L'hypertexte de fiction : naissance d'un nouveau genre ? In *Communication au colloque de l'ALLC*. Sorbonne, Avril 1994. URL <http://hypermedia.univ-paris8.fr/jean/articles/allc.htm>. Cité page 43.
- I. COMPIÈGNE : *La société numérique en question(s)*. Editions Sciences Humaines, 2011. Cité pages 16, 27, 29, 35, 47 et 75.
- F. CONFINO : *Blogs territoriaux, réseaux sociaux et nouveaux enjeux du web2.0 pour les collectivités*. Territorial Editions, octobre 2009. ISBN 978-2-35295-825-3. Cité pages 27 et 53.
- D. CONNOLLY : A little history of the world wide web, 2000. URL <http://www.w3.org/History.html>. Cité page 42.
- J-M. CORNU : *la coopération, nouvelles approches*. décembre 2004. URL http://www.cornu.eu.org/files/cooperation1_2.pdf. Cité page 46.
- S. COSTE-MARQUIS, C. DEVRED, S. KONIECZNY, M. C. LAGASQUIE-SCHIEX et P. MARQUIS : On the merging of dung's argumentation systems. *Artif. Intell.*, 171(10-15):730–753, 2007. Cité page 89.
- J. COTRET et A. GOUAÏCH : Modalités minimales d'interactions pour la pratique du débat en ligne. Rapport technique, novembre 2011. URL <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00644038/en/>. Cité page 128.
- J. DE ROSNAY : L'ordinateur du futur fera communiquer les objets et les humains, december 2006. URL <http://www.cite-sciences.fr/derosnay/articles/Chathtml>. Cité page 18.
- J. DELACROIX : *Les wikis : les espaces de l'intelligence collective*. M2 éditions, 2005. ISBN 2-9520514-4-5. Cité page 45.
- J-F. DELANNOY : Argumentation mark-up : A proposal. Workshop On Towards Standards And Tools For Discourse Tagging, 1999. URL <http://www.aclweb.org/anthology-new/W/W99/W99-0303.pdf>. Cité pages 90 et 91.
- J. DENOUËL-GRANJON : Nouvelles approches de la coprésence médiatisée par messagerie instantanée. In *Doctoriales du GDR TIC et Société*, juin 2008. URL http://gdrtics.u-paris10.fr/pdf/doctoriales/2008/DENOUEL_GRANJON.pdf. Cité page 112.

- N. DESQUINABO : *Caractéristiques et contraintes des genres interlocutifs dans les émissions de plateaux télévisées : analyses pragma-discursives et en réception.* Thèse de doctorat, Université Paris III-Sorbonne Nouvelle, 2005. Cité page 32.
- N. DESQUINABO : Dynamiques et impacts des propositions politiques dans les webforums partisans. *Réseaux*, 26(150), (2008). Cité page 125.
- N. DESQUINABO et N. FERRAND : Participative e-governance with stakeholders : the camargue park management plan case. In *Electronic Government, Proceedings of EGOV 09, 8th International Conference*, pages 107–114. H.J Scholl, M. Janssen, R. Traunmüller and M.A. Wimmer, august-september 2009. Cité page 164.
- N. DESQUINABO, N. FERRAND et J. MARLIER : E-participation benefits for local planning : an experimental study. In *Electronic Participation, Proceedings of ePart 2009, 1st International Conference*, pages 176–185. E. Tambouris and A. Macintosh, september 2009. Cité page 163.
- N. DESQUINABO, N. FERRAND et J. MARLIER : Stakeholder e-participation in local planning : The camargue park case. In *Next Generation Society Technological and Legal Issues, eDemocracy 2009, LNCS*, volume 26, pages 128–137. A.B. Sideridis and Ch.Z. Patrikakis, 2010. Cité page 164.
- C. DEVRED et S. DOUTRE : Preuve dialectique dans les systèmes d'argumentation à contrainte. URL <http://gdri3iaf.info.univ-angers.fr/IMG/pdf/devred-doutre.pdf>. 2008. Cité page 89.
- P. M. DUNG : On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77:321–357, September 1995. ISSN 0004-3702. Cité page 89.
- J. ENGESTRÖM : Why some social network services work and others don't — or : the case for object-centered sociality, 2005. URL <http://www.zengestrom.com/blog/2005/04/why-some-social-network-services-work-and-others-dont-or-the-case-for-object-centered-sociality.html>. Cité page 69.
- G. ERÉTÉO : *Semantic Social Network Analysis.* Thèse de doctorat, INRIA Sophia Antipolis - Méditerranée, April 2011. URL <http://tel.archives-ouvertes.fr/tel-00586677/fr/>. Cité pages 113 et 142.
- G. ERÉTÉO, M. BUFFA, F. GANDON et O. CORBY : Analysis of a Real Online Social Network using Semantic Web Frameworks. In *Lecture Notes in Computer Science ISWC 2009*, volume 5823/2009, pages 180–195, Washington États-Unis, 10 2009a. URL <http://hal.archives-ouvertes.fr/hal-00562035/en/>. ANR. Cité page 88.

- G. ERÉTÉO, M. BUFFA, F. GANDON, P. GROHAN, M. LEITZELMAN et P. SANDER : A State of the Art on Social Network Analysis and its Applications on a Semantic Web. In *Proc. SDoW2008 (Social Data on the Web), Workshop held with the 7th International Semantic Web Conference*, Karlsruhe, Germany, October 2008. Cité page 88.
- G. ERÉTÉO, F. GANDON, M. BUFFA et O. CORBY : Semantic social network analysis. In *Web Science WebSci'09, Athens, Greece*, march 2009b. URL http://journal.webscience.org/141/2/websci09_submission_43.pdf. Cité page 88.
- P. FLICHY : Internet, un outil de la démocratie ? la vie des idées, janvier 2008. URL <http://www.laviedesidees.fr/Internet-un-outil-de-la-democratie.html>. Cité pages 28 et 76.
- M. FORSÉ : *La sociologie de Georg Simmel*, chapitre Les réseaux sociaux chez Simmel : les fondements d'un modèle individualiste et structural, pages 63–107. Presses Universitaires de France, Paris, 2002. Cité page 61.
- S. FORTUNATO : Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010. ISSN 0370-1573. URL <http://www.sciencedirect.com/science/article/B6TVP-4XPYXF1-1/2/99061fac6435db4343b2374d26e64ac1>. Cité pages 66 et 67.
- A. FUNG : Varieties of Participation in Complex Governance. *Public administration review*, 66:66, décembre 2006. URL <http://proquest.umi.com/pqdweb?did=1174438121&Fmt=7&clientId=65345&RQT=309&VName=PQD>. Cité page 75.
- B. J. GAINES et J. J. MONDAK : Typing together? clustering of ideological types in online social networks. *Journal of Information Technology and Politics*, 6(3 et 4):216–231, juillet 2009. Cité page 85.
- J. F. GANTZ, D. REINSEL, C. CHUTE, W. SCHLICHTING, J. MCARTHUR, S. MINTON, I. XHENETI, A. TONCHEVA et A. MANFREDIZ : The diverse and exploding digital universe : An updated forecast of worldwide information growth through 2011. *IDC white paper*, 2007. URL <http://www.mendeley.com/research/the-expanding-digital-universe-a-forecast-of-worldwide-information-growth-through-2010/>. Cité page 20.
- F. GEORGES : *Sémioptique de la représentation de soi dans les dispositifs interactifs. L'Hexis numérique*. Thèse de doctorat, Université Panthéon-Sorbonne - Paris I, 2007. Cité pages 115 et 137.
- F. GEORGES, A. SEILLES, G. ARTIGNAN, B. ARNAUD, M. HASCOËT, N. RODRIGUEZ, J. SAL-LANTIN et B. DRESP-LANGLEY : Sémiotique et visualisation de l'identité numérique : une étude comparée de facebook et myspace. In HERMÈS, éditeur : *Actes d'H2PTM'09 (Hypermedias, Hypertexts, Products, Tools and Methods) Rétrospective et Perspective 1989-2009*, pages 257–268. Hermès, 2009. Cité pages 113 et 134.

- F. GEORGES, A. SEILLES et J. SALLANTIN : Des illusions de l'anonymat : les stratégies de préservation des données personnelles à l'épreuve du web 2.0. *Revue CREIS Terminal : Technologies et usages de l'anonymat à l'heure de l'Internet*, (105):97–107, 2010a. Cité page 113.
- F. GEORGES, A. SEILLES et J. SALLANTIN : Sic et informatique : vers une approche quantifiée de la médiation de soi dans la communication médiée par ordinateur. In *Actes du 17ème congrès de la SFSIC, Société Française des Sciences de l'Information et de la Communication : Au cœur et aux lisières des SIC*. SFSIC, Juin 2010b. Cité page 113.
- A. GIBOIN et F. GANDON : Vers des ontologies à l'état sauvage. In *Atelier IC 2.0, IC2008*, June 2008. Cité page 60.
- S. GOLDER et B. A. HUBERMAN : Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, April 2006. URL <http://www.isrl.uiuc.edu/~amag/langev/paper/golder05taggingSystems.html>. Cité page 139.
- A. GÓMEZ-PÉREZ et O. CORCHO : Ontology specification languages for the semantic web. *IEEE Intelligent Systems*, 17:54–60, 2002. ISSN 1541-1672. Cité page 98.
- M. S. GRANOVETTER : The Strength of Weak Ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973. Cité pages 68 et 135.
- M. S. GRANOVETTER : The strength of weak ties : A network theory revisited. In *Sociological Theory*, pages 105–130, 1982. Cité page 68.
- T. GRUBER : Ontology. In *Encyclopedia of Database Systems*, pages 1963–1965. 2009. Cité page 59.
- T. R. GRUBER : A translation approach to portable ontology specifications. *KNOWLEDGE ACQUISITION*, 5:199–220, 1993. Cité page 59.
- H HALPIN et V PRESUTTI : An ontology of resources for linked data. *World*, 2009. URL <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:An+Ontology+of+Resources+for+Linked+Data>. Cité page 44.
- T. HEATH, C. BIZER et J. HENDLER : *Linked Data*. Morgan & Claypool Publishers, 1st édition, 2011. ISBN 1608454304, 9781608454303. Cité pages 55 et 56.
- I. HERMAN, Ieee Cs SOCIETY, G. MELANÇON et M. S. MARSHALL : Graph visualization and navigation in information visualization : a survey. *IEEE Transactions on Visualization and Computer Graphics*, 6:24–43, 2000. Cité page 166.
- A. ISKOLD : The art, science and business of recommendation engines, 2007. URL http://www.readwriteweb.com/archives/recommendation_engines.php. Cité page 65.

- R. KVAVIK : *Educating the Net Generation*, chapitre Convenience, Communications, and Control : How Students Use Technology. EDUCAUSE, 2005. URL <http://www.educause.edu/Resources/EducatingtheNetGeneration/ConvenienceCommunicationsandCo/6070>. Cité page 16.
- A. LALANDE : *Vocabulaire Technique et Critique de la Philosophie*. Puf, 1900. Cité page 95.
- B. LATOUR : *Changer de société, refaire de la sociologie*. Editions La Découverte, septembre 2007. Cité page 69.
- S. LAVELLE : Politique des artefacts. *Cités*, 39:39–51, 2009. URL <http://www.cairn.info/revue-cites-2009-3-page-39.htm>. Cité page 35.
- A. LENHART et M. MADDEN : Teens, Privacy and Online Social Networks : How teens manage their online identities and personal information in the age of MySpace. Rapport technique, Pew Internet and American Life Project, Washington DC, 2007. Cité page 167.
- F. LEVIE : *L'homme qui voulait classer le monde : Paul Otlet et le Mundaneum*. Les impressions nouvelles, 2006. ISBN 2-87449-022-9. Cité page 35.
- J. C. R. LICKLIDER : Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1:4–11, march 1960. URL <http://memex.org/licklider.html>. Cité page 18.
- J. C. R. LICKLIDER et R. W. TAYLOR : The computer as a communication device. *Science and Technology*, 76:21–31, 1968. Cité pages 4, 35 et 46.
- G. LORTAL : *Médiatiser l'annotation pour une herméneutique numérique : AnT&CoW, un collecticiel pour une coopération via l'annotation de documents numériques*. Thèse de doctorat, Université de Technologie de Troyes, 11 2006. URL <http://tel.archives-ouvertes.fr/tel-00136042/en/>. Cité pages 99 et 126.
- D. LUZEAUX, J. SALLANTIN et C. DARTNELL : Logical extensions of aristotle's square. In *Handbook of the Square of Opposition*, 2007. Cité pages 93 et 94.
- C. C. MARSHALL : Toward an ecology of hypertext annotation. In *HYPERTEXT '98 : Proceedings of the ninth ACM conference on Hypertext and hypermedia : links, objects, time and space—structure in hypermedia systems*, pages 40–49, New York, NY, USA, 1998. ACM. ISBN 0-89791-972-6. URL <http://dx.doi.org/10.1145/276627.276632>. Cité page 99.
- A. McAFFEE : Enterprise 2.0 : The dawn of emergent collaboration, april 2006a. URL <http://sloanreview.mit.edu/the-magazine/2006-spring/47306/enterprise-the-dawn-of-emergent-collaboration/>. Cité page 54.

- A. McAfee : Enterprise 2.0, version 2.0, may 2006b. URL http://andrewmcafee.org/2006/05/enterprise_20_version_20/. Cité pages 50 et 53.
- A. McAfee : Enterprise 2.0 vs. soa, may 2006c. URL http://andrewmcafee.org/2006/05/enterprise_20_vs_soa/. Cité pages 50 et 53.
- P. MERCKLÉ : les réseaux sociaux, les origines de l'analyse des réseaux sociaux, 2004. URL http://eco.ens-lsh.fr/sociales/reseaux_merckle_03_origines.pdf. Cité page 61.
- J. MOESCHLER et A. REBOUL : *Dictionnaire Encyclopédique de Pragmatique*. Seuil, 1994. ISBN 2-02-013042-4. Cité page 58.
- A. MONNIN, F. LIMPENS, F. GANDON et D. LANIADO : Speech acts meet tagging : Nicetag ontology. In *I-SEMANTICS '10 : Proceedings of the 6th International Conference on Semantic Systems*, pages 1–10, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0014-8. Cité pages 131 et 140.
- A. MONNIN, F. LIMPENS, D. LANIADO et F. GANDON : Nice tag ontology. URL <http://ns.inria.fr/nicetag/2010/09/09/voc.html>. Cité page 90.
- J. L. MORENO : *Who shall survive ? : Foundations of sociometry, group psychotherapy and sociodrama*. Bacon House, Beacon, N. Y. :, [rev. ed.]. édition, 1953. URL <http://www.asgpp.org/docs/WSS/WSS.html>. Cité page 62.
- A. MORETTI : Geometry for modalities ? yes : through n-opposition theory. In Elsevier BV, éditeur : *Aspects of Universal Logic*, volume 6. J-Y.BEZIAU, A.COSTA-LEITE, A.FACCHINI, 2005. Cité page 92.
- P. MUTTON : Inferring and visualizing social networks on internet relay chat. In *Proceedings of the Information Visualisation, Eighth International Conference*, pages 35–43, Washington, DC, USA, 2004. IEEE Computer Society. ISBN 0-7695-2177-0. URL <http://portal.acm.org/citation.cfm?id=1018435.1021596>. Cité pages 115 et 134.
- S. F. NADEL : *The Theory of social Structure*. Cohen et West, London, 1957. Cité pages 70 et 165.
- T. H. NELSON : *Computer Lib : You can and must understand computers now/Dream Machines : New freedoms through computer screens—a minority report*. South Bend, 1974. ISBN 0-914845-49-7. Cité page 43.
- T. H. NELSON : *Literary Machines*. Mindful Press, 1982. ISBN 0-89347-062-7. Cité page 35.
- S. H. NIELSEN et S. PARSONS : A generalization of dung's abstract framework for argumentation : Arguing with sets of attacking arguments. In *Argumentation in Multi-Agent Systems*, pages 54–73, 2006. Cité page 89.

- I. NONAKA et H. TAKEUCHI : *The Knowledge-Creating Company : How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, mai 1995. ISBN 0195092694. URL <http://www.worldcat.org/isbn/0195092694>. Cité page 46.
- D. NUTE : Defeasible logic. In Oxford University PRESS, éditeur : *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3, pages 353–395. D.M.GABBAY, C.J.HOGGER, J.A.ROBINSON, 1994. Cité pages 90 et 152.
- T. O'REILLY : What is web 2.0. design patterns and business models for the next generation of software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, September 2005. URL <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>. Stand 12.5.2009. Cité page 51.
- G. PALLA, A. BARABASI et T. VICSEK : Quantifying social group evolution. *Nature*, 446:664–667, avril 2007. URL <http://dx.doi.org/doi:10.1038/nature05670>. Cité page 67.
- A. PASSANT : *Technologies du Web Sémantique pour l'Entreprise 2.0*. Thèse de doctorat, Université Paris-Sorbonne, 06 2009. Cité pages 50 et 54.
- M. PLANTIÉ, M. ROCHE, G. DRAY et P. PONCELET : Is a voting approach accurate for opinion mining? In *Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery*, DaWaK '08, pages 413–422, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-85835-5. URL http://dx.doi.org/10.1007/978-3-540-85836-2_39. Cité page 86.
- P. PONS : *Détection de communautés dans les grands graphes de terrain*. Thèse de doctorat, Université Paris 7 - Denis Diderot, juillet 2007. URL http://psl.pons.free.fr/publi/these_pascal_pons.pdf. Cité pages 65 et 67.
- V. PRICE : Citizen deliberating online : Theory and some evidence. 2006. Cité pages 28 et 75.
- I. RAHWAN et P. V. SAKEER : Towards representing and querying arguments on the semantic web. In *Proceeding of the 2006 conference on Computational Models of Argument : Proceedings of COMMA 2006*, pages 3–14, Amsterdam, The Netherlands, The Netherlands, 2006. IOS Press. ISBN 1-58603-652-1. URL <http://portal.acm.org/citation.cfm?id=1565233.1565236>. Cité pages 89, 90 et 91.
- A. REVILLARD : Les interactions sur l'internet. *Terrains et Travaux*, 1:108–129, 2000. Cité pages 28, 115 et 134.
- H. RHEINGOLD : *The Virtual Community : Homesteading on the Electronic Frontier*. Reading, Massachusetts : Addison-Wesley, 1993. ISBN 0-201-60870-7. URL <http://www.rheingold.com/vc/book/>. Cité pages 115 et 134.

- J. SALLANTIN et A. SEILLES : n-opposition theory to structure debates. Rapport technique, September 2009. URL http://hal.archives-ouvertes.fr/docs/00/41/22/80/PDF/article_jyb.pdf. Cité page 151.
- S. SCHAEFFER : Graph clustering. *Computer Science Review*, 1(1):27–64, août 2007. ISSN 15740137. URL <http://dx.doi.org/10.1016/j.cosrev.2007.05.001>. Cité page 66.
- J.R. SEARLE : *Sens et expression*. Paris : Editions de Minuit, 1982. Cité page 32.
- A. SEILLES : Modèles et outils de raisonnement argumentatif dans les communautés et organisations virtuelles. Mémoire de D.E.A., Université Montpellier 2, juin 2007. Cité pages 150 et 151.
- A. SEILLES, J. COTRET, F. GEORGES, N. RODRIGUEZ et J. SALLANTIN : Co-conception multidisciplinaire d'une plateforme de débat en ligne. *Document numérique*, 14 (n°1/2011):011–030, avril 2011. ISSN 1279-5127. URL <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00599213/en/>. Cité page 157.
- A. SEILLES, J. COTRET, F. GEORGES, J. SALLANTIN, N. RODRIGUEZ, A. GOUAICH et C. FAGOT : L'annotation discursive et sémantique pour la pratique de débats 2.0. *Document numérique*, 13(3):153–177, 12 2010a. ISSN 1279-5127. URL <http://hal-lirmm.ccsd.cnrs.fr/lirmm-00572868/en/>. Cité page 172.
- A. SEILLES, N. RODRIGUEZ et J. SALLANTIN : Stakeholder detection for online debates. In *@WAS Masters and Doctoral Colloquium*. IIWAS, 2010b. Cité page 135.
- J. SUROWIECKI : *The Wisdom of Crowds*. Random House, New York, 2004. Cité pages 17 et 47.
- TERRANOVA : Moderniser la vie politique : Innovations Américaines, Leçons pour la France. Rapport de mission, Fondation Terra Nova, janvier 2009. URL <http://www.tnova.fr/sites/default/files/terranova-rapportmissionus.pdf>. Cité pages 48 et 49.
- P. TESTARD-VAILLANT : Internet, un outil au service de la démocratie ? *Le journal du CNRS*, 219, 04 2008. Cité pages 20 et 73.
- S. THORNE : Data analysis in qualitative research. *Evid Based Nurs*, 3:68–70, 2000. URL <http://ebn.bmj.com/content/3/3/68.extract>. Cité page 95.
- S. E. TOULMIN : *The Uses of Argument*. Cambridge University Press, July 1958. ISBN 0521534836. URL <http://www.amazon.com/exec/obidos/redirect?tag=citeulike-20&path=ASIN/0521534836>. Cité page 89.
- J. TRAVERS et S. MILGRAM : An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969. Cité page 65.

- J. R. TYLER, D. M. WILKINSON et B. A. HUBERMAN : E-Mail as Spectroscopy : Automated Discovery of Community Structure within Organizations. *The Information Society*, 21(2):143–153, avril 2005. URL <http://taylorandfrancis.metapress.com/index/L92R14R72473713X.pdf>. Cité pages 115 et 134.
- T. VANDER WAL : Folksonomy coinage and definition, february 2007. URL <http://vanderwal.net/folksonomy.html>. Cité pages 81 et 97.
- T. VEDEL : L'idée de démocratie électronique : Origines, visions, questions. 2003. Cité pages 20, 27, 73, 74 et 76.
- J. VERONIS : Annotation automatique de corpus : panorama et état de la technique. pages 111–129, 2000. Cité page 99.
- G. VIGNAUX : *Le discours acteur du monde*. Ophrys, 1988. Cité page 21.
- S. VILLATA : *Meta-Argumentation for Multiagent Systems : Coalition Formation, Merging Views, Subsumption Relation and Dependence Networks*. Thèse de doctorat, Università degli Studi di Torino, january 2010. Cité page 89.
- B. WELLMAN, J. SALAFF, D. DIMITROVA, L. GARTON, M. GULIA et C. HAYTHORNTHWAITE : Computer networks as social networks : Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22(1):213–238, 1996. URL <http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.soc.22.1.213>. Cité page 134.
- H. C. WHITE : *Identity and control : how social formations emerge*. Princeton University Press, 2008. ISBN 0691137153. Cité page 69.
- X. WU, L. ZHANG et Y. YU : Exploring social annotations for the semantic web. In *WWW '06 : Proceedings of the 15th international conference on World Wide Web*, pages 417–426, New York, NY, USA, 2006. ACM. ISBN 1-59593-323-9. Cité page 97.

Abstract

Keywords: *annotations, social network analysis, social web, semantic web, e-democracy*

Résumé

Cette thèse traite de l'usage de l'annotation socio-sémantique dans le cadre de la démocratie électronique et plus particulièrement des débats en ligne. L'annotation socio-sémantique est utilisée ici comme solution de structuration des débats. La représentation des données des débats est pensée pour faciliter la mise en place de méthodes d'extraction et d'analyses du réseau social des utilisateurs, en particulier pour faciliter l'extraction de groupes d'opinions. La thèse est réalisée dans le contexte de l'ANR Intermed qui vise à produire des outils d'aide à la concertation en ligne, en particulier pour la gestion de zones côtières.

En nous appuyant sur la tendance 2.0, nous définissons la notion de débat 2.0. Débat à grande échelle, au moins l'échelle d'une collectivité territoriale, s'appuyant sur l'usage de technologies du Web 2.0 pour faciliter les interactions entre les citoyens. Dans ce contexte, l'interopérabilité est un enjeu crucial. Si les annotations discursives s'inscrivent dans la tendance 2.0 et permettent aux citoyens de discuter en ligne, le traitement des données produites en vue de structurer les débats, de synthétiser les discussions, de modérer, d'évaluer la représentativité ... devient une tâche de plus en plus complexe avec l'augmentation de la quantité de données produites. Nous proposons d'utiliser les technologies du web sémantique, et donc des annotations à la fois discursives et sémantiques (appelées annotations socio-sémantique), pour représenter les données produites par les citoyens dans un outil de débat 2.0 et pour faciliter l'interopérabilité de ces données, faciliter la création d'autres services comme par exemple un service d'analyse du réseau social, un service de recommandation, un service de visualisation des débats ... Nous présentons donc un mécanisme d'annotation structurant les discussions, fruit d'un processus incrémental d'implémentation et d'expérimentation sur le terrain.

Mots clefs : *annotations, analyse de réseaux sociaux, web social, web sémantique, démocratie électronique*
