



**HAL**  
open science

# La gestion de la recherche collaborative d'information dans le cadre du processus d'intelligence économique

Victor Odumuyiwa

► **To cite this version:**

Victor Odumuyiwa. La gestion de la recherche collaborative d'information dans le cadre du processus d'intelligence économique. Sciences de l'information et de la communication. Université Nancy 2, 2010. Français. NNT : 2010NAN21014 . tel-01748886

**HAL Id: tel-01748886**

**<https://hal.univ-lorraine.fr/tel-01748886>**

Submitted on 29 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

**LA GESTION DE LA RECHERCHE COLLABORATIVE  
D'INFORMATION DANS LE CADRE DU PROCESSUS  
D'INTELLIGENCE ECONOMIQUE**

**THÈSE**

Présentée et soutenue publiquement le 13 décembre 2010 pour l'obtention du

**Doctorat de l'Université Nancy 2**

(Spécialité : Sciences de l'Information et de la Communication)

par

Victor ODUMUYIWA

**Composition du jury**

Directeur : *Amos DAVID : Professeur en SIC à l'Université Nancy2*

Rapporteurs :

*Laïd BOUZIDI : Professeur en SIC à IAE Lyon*

*Madjid IHADJADENE : Professeur en SIC à l'Université Paris 8*

Examineurs :

*Anne BOYER : Professeur en Informatique à l'Université Nancy2*

*Luc QUONIAM : Professeur en SIC à l'Université du Sud Toulon-Var*

*Odile THIERY : Professeur en Informatique à l'Université Nancy2*





*La faculté n'entend donner ni approbation ni improbation  
aux opinions émises dans cette thèse. Ces dernières doivent être  
considérées comme propres à leurs auteurs.*



# Remerciements

---

*Je tiens à remercier Amos David, responsable de l'équipe SITE, professeur à l'université Nancy 2, de m'avoir accueilli dans son équipe et d'avoir encadré mes travaux de recherche. Je le remercie vivement pour sa patience, ses conseils, ses corrections, ses soutiens, ses encouragements tout au long de ces années de recherche. Il a été plus qu'un directeur de thèse pour moi. Ses expériences à la fois scientifiques et sociales m'ont beaucoup apporté.*

*Je remercie Mme la Professeure Odile Thiery pour son soutien tout au long de ces quatre années et pour le temps qu'elle a consacré pour relire ma thèse.*

*Je remercie M. le Professeur Laïd BOUZIDI et M. le Professeur Madjid IHADJADENE pour m'avoir fait l'honneur de rapporter cette thèse et pour le temps qu'ils ont consacré à sa lecture, sa correction et sa soutenance. Je les remercie également pour l'intérêt qu'ils ont manifesté à l'égard de mon travail.*

*Je remercie M. le Professeur Luc QUONIAM et Mme la Professeure Anne BOYER pour avoir accepté de participer au jury de ma thèse. Je les remercie pour leur temps et leur disponibilité.*

*Je remercie tout particulièrement la famille Wilhelm pour m'avoir accueilli comme l'un des leurs. Merci à Betty Wilhelm qui est devenu une maman pour moi en France. Un grand merci à Henri Wilhelm, mon grand frère, pour le temps qu'il a consacré à corriger une partie de ce mémoire.*

*Je tiens à remercier très chaleureusement Alain Mattieu pour sa gentillesse et ses soutiens tout au long de mes années en France. Tu es vraiment un père pour les étrangers. Merci beaucoup pour ta patience et pour les corrections que tu m'as apportées.*

*Je remercie tous les membres de l'équipe SITE et d'autres collègues au LORIA. Merci à Hanène Maghrebi pour m'avoir supporté au bureau. Merci à mes grands frères, Charles Robert et Babajide Afolabi pour leur soutien au début de ma thèse. Merci à Stéphane Gorla, Audrey Knauf, Pierre Humbert, Philippe Kislin, Frédérique Peguiron, Chedia Dhaoui, Ahmed. Tous mes remerciements à Soji Okunoye, Bola Oladejo, Williams Onifade, Idowu Gbolahan et Cecile Ganso.*

*Je remercie Mme la Professeure Adenike Osofisan, M. le Professeur Ojo Ayeni, M. le Professeur Ajila, M. le Professeur Chico Eboué pour leurs encouragements et leurs contributions à mon parcours scientifique.*

*Je tiens vraiment à remercier Michaël YOSASSE pour son amitié. Je le remercie pour ses sacrifices et sa gentillesse. Il m'a prouvé qu'un vrai ami est celui qui donne sans attendre une compensation ou un retour.*

*Je remercie tous mes amis qui m'ont accompagné tout au long de ces années : Stéphane Tala, Franck Moudiki, Clémentino, Gaëlle Moudiki, Andréa Moudiki, Dika Ruth-Pamella Florine Scasso, Caroline Scasso, Marc Montemont, Gabriel Amvane, Antonie Beaupage, Lydienne Tala, Claire Acampo, Charles Bourgeois, Elizabeth David etc.*

*Que serai-je devenu sans les apports immenses de ma famille ? Mes plus grands remerciements à mes parents, à mes frères : Ayodeji, Gboyega, Bankole, et à ma sœur Deola. Ce travail est dédié à eux.*

*Je tiens à remercier Sade Asaye. Ses mots d'encouragements sont comme le baume de Galaad à mon cœur.*





# Dedicace

---

Ce travail est dédié à tous ceux qui ont collaboré avec moi dès ma naissance.



# Sommaire

---

Remerciements .....	vi
Dedicace .....	viii
Sommaire .....	x
Table des illustrations .....	xvi
Tableaux .....	xix
Introduction générale .....	1
Contexte de l'étude .....	1
Hypothèse et problématique .....	5
Structure de la thèse .....	7
Chapitre 1 .....	11
Contexte d'étude et vue globale de la collaboration .....	11
1.1 Introduction .....	11
1.2 A propos du concept d'information .....	11
1.2.1 Formes fondamentales de l'information selon Bates .....	12
1.2.2 Trois perspectives de l'information .....	14
1.2.2.1 Information selon Shannon .....	14
1.2.2.2 Point de vue cognitif de l'information .....	17
1.2.2.3 Point de vue social de l'information .....	19
1.2.3 Manifestations de l'information .....	20
1.2.3.2 L'information en tant que connaissance .....	20
1.2.3.3 L'information en tant qu'objet .....	20
1.3 A propos du concept de connaissance .....	21
1.3.1 La connaissance tacite et explicite .....	21
1.3.2 La transformation de connaissance entre types tacites ou explicites .....	24
1.3.3 Donnée, information et connaissance : quelle différence ? .....	25
1.4 La communication .....	27
1.4.1 La communication vue sous quatre ordres de réalité .....	27
1.4.1.1 Les pratiques de communication .....	27
1.4.1.2 Techniques de communication .....	29
1.4.1.3 Théories de la communication .....	29
1.4.1.4 Enjeux associés à la communication .....	30
1.4.2 La cybernétique .....	30
1.4.3 La communication interpersonnelle .....	33
1.4.3.1 La relation interpersonnelle vue sous trois dimensions .....	35
Le contexte .....	35
La distance .....	36
Le temps .....	36

1.5 A propos du concept de collaboration .....	37
1.5.1 Collaborer ou coopérer ou coordonner ? .....	38
1.5.2 Les dimensions de la collaboration .....	42
L'intention .....	42
La profondeur .....	42
Le synchronisme.....	43
La localisation.....	43
1.5.3 Les facteurs sociocognitifs associés à la collaboration .....	43
1.5.3.1 Modèle de Freshman.....	45
1.5.4 Quelques concepts associés à la collaboration .....	47
1.6 Conclusion du premier chapitre.....	49
Chapitre 2 .....	51
L'intelligence économique : domaine d'application.....	51
2.1 Introduction .....	51
2.2 Définitions et caractéristiques générales de l'intelligence économique .....	52
2.2.1 Les acteurs de l'IE.....	58
2.3 Processus de l'IE selon SITE.....	59
2.3.1 Les axes de recherche de l'équipe SITE.....	60
2.3.1.1 Modélisation de l'utilisateur-acteur.....	60
2.3.1.2 Modélisation de l'interaction entre l'utilisateur et le médiateur .....	60
2.3.1.3 Conception et exploitation d'un entrepôt de données.....	61
2.3.2 Processus de l'IE .....	61
2.3.3 Le paradigme des abeilles .....	63
2.3.4 Les modèles de l'IE au sein de l'équipe SITE .....	66
2.3.4.1 Définition d'un problème décisionnel.....	66
2.3.4.2 Passage du problème décisionnel au problème informationnel .....	68
2.3.4.3 D'autres modèles développés au sein de l'équipe SITE .....	69
2.4 La recherche collaborative d'information dans le processus de l'IE.....	70
2.5 Conclusion du deuxième chapitre.....	71
Chapitre 3 .....	75
De la recherche d'information à la recherche sociale d'information .....	75
3.1 Introduction .....	75
3.2 Le comportement informationnel de l'utilisateur .....	76
3.2.1 L'accès à l'information .....	80
3.2.2 Le besoin informationnel.....	82
3.3 Système de recherche d'information .....	84
3.3.1 Introduction .....	84
3.3.2 Le mode de fonctionnement d'un SRI.....	86
3.3.2.1 La collection, l'indexation et la classification des documents .....	89

3.3.2.1.1 La collection .....	89
3.3.2.1.2 L'indexation.....	89
3.3.2.1.3 La classification .....	90
3.3.2.2 La représentation du besoin informationnel utilisateur sous forme de requêtes .....	92
3.3.2.3 La fonction d'appariement requête – document.....	92
3.3.3 Les modèles de système de recherche d'information.....	93
3.3.3.1 Le modèle booléen.....	93
3.3.3.2 Le modèle vectoriel.....	95
3.3.3.3 Le modèle probabiliste.....	97
3.3.4 La notion de pertinence en recherche d'information.....	101
3.3.5 Les limitations de SRI classique .....	103
3.4 L'intégration de la représentation de l'utilisateur dans les SRI.....	105
3.4.1 Système de recherche d'information personnalisé .....	108
3.5 La recherche sociale d'information .....	110
3.5.1 Le filtrage collaboratif.....	113
3.5.2 L'indexation sociale .....	117
3.6 Conclusion du troisième chapitre .....	119
Chapitre 4 .....	123
Du comportement informationnel collaboratif à la recherche collaborative d'information.....	123
4.1 Introduction .....	123
4.2 Comportement informationnel collaboratif .....	123
4.2.1 Modèle pour la compréhension du CIB selon Reddy et Jassen.....	126
4.2.2 Modèle métacognitif de Vivian et Dinet .....	128
4.2.3 Cadre d'analyse de travail cognitif pour la recherche collaborative d'information .....	130
4.3 La recherche collaborative d'information (RCI) .....	131
4.4 Systèmes de recherche collaborative d'information existants .....	134
4.4.1 Systèmes permettant la navigation collaborative .....	134
4.4.2 Les systèmes de recherche collaborative.....	139
4.4.2.1 C-TORI .....	139
4.4.2.2 ARIADNE.....	140
4.4.2.3 CIRE .....	141
4.4.2.4 METIORE («Multimedia coopErative informaTIOOn Retrieval systEm »).....	142
4.4.2.5 SearchTogether .....	144
4.4.2.6 Le système de Foley.....	147
4.4.2.6 Le système RCI Web .....	148
4.5 Analyse des SRCI existants .....	150
4.6 Le cadre conceptuel pour la gestion de la recherche collaborative d'information .....	152
4.6.1 La communication.....	153
4.6.2 Le mode de collaboration.....	154

4.6.3 La coordination des interactions entre les utilisateurs.....	155
4.6.4 La gestion des connaissances impliquées dans la collaboration.....	157
4.7 Conclusion du quatrième chapitre.....	158
Chapitre 5.....	161
Propositions en terme de modèles.....	161
5.1 Introduction.....	161
5.2 Le modèle de pyramide de collaboration.....	161
5.2.1 La phase de confiance de départ.....	163
5.2.2 La phase de compréhension partagée du problème informationnel.....	164
5.2.3 La phase de communication.....	165
5.2.4 La phase de partage de connaissances.....	166
5.2.4.1 Connaissances et compétences à partager.....	168
5.2.4.2 Activités de recherche d'information qui nécessite le partage.....	169
5.2.5 La phase de conscience de groupe.....	170
5.2.6 La phase de répartition des tâches.....	172
5.3 Le modèle de communication pour la recherche collaborative d'information (COCIR).....	173
5.3.1 Utilisateur.....	175
5.3.2 Objet informationnel.....	178
5.3.3 Contexte.....	180
5.4 Conclusion du cinquième chapitre.....	184
Chapitre 6.....	187
Réalisation d'un SRCI et Expérimentation.....	187
6.1 Introduction.....	187
6.2 Développement du système MECOCIR.....	187
6.2.1 Architecture du MECOCIR.....	187
6.2.2 Langage de programmation Java.....	190
6.2.3 Les composants logiciels du système MECOCIR.....	191
6.2.3.1 L'espace de travail pour les utilisateurs.....	192
6.2.3.2 Le noyau de MECOCIR.....	193
6.2.3.2.1 Le moteur de navigation.....	193
6.2.3.2.2 Module de communication.....	194
6.2.3.2.3 Contexte manager.....	196
6.2.3.2.4 Module pour capter les activités.....	197
6.2.3.2.5 Moteur d'annotation.....	197
6.2.3.2.6 Module de visualisation.....	197
6.2.3.2.7 Module pour la répartition des tâches.....	197
6.2.3.2.8 Analyseur d'indicateur.....	198
6.2.3.2.9 Moteur de filtrage collaboratif et système de recommandation.....	198
6.2.3.3 La couche de données et de connaissances.....	199

6.2.3.3.1 MySQL .....	200
6.2.3.3.2 JDBC (Java Database Connectivity) .....	200
6.3 Le cadre d'application .....	202
6.4 Utilisation de MECOCIR .....	205
6.4.1 Inscription.....	206
6.4.2 L'interface graphique principale de MECOCIR .....	207
6.4.3 Définition du contexte de recherche.....	210
6.4.4 Demande de collaboration.....	213
6.4.5 Clarification du problème informationnel .....	215
6.4.6 La recherche d'information .....	217
6.4.7 La conscience de groupe et le partage de connaissances.....	218
6.4.8 Synchronisation de l'interface de navigation .....	221
6.4.9 La répartition des tâches.....	223
6.4.9 Paradigme de collaboration.....	225
6.5 Expérimentation.....	226
6.5.1 Première série d'expérimentation.....	227
6.5.2 Deuxième série d'expérimentation.....	232
6.5.3 Bilan des deux séries d'expérimentation.....	236
6.6 Comparaison du MECOCIR avec certains SRCI existants .....	237
6.7 Conclusion du sixième chapitre.....	238
Conclusion générale .....	240
Bibliographies .....	<b>Erreur ! Signet non défini.</b>
Annexe 1 : Exemplaires des questionnaires utilisés lors d'expérimentation.....	247





## Table des illustrations

Figure 1.1 : Un schéma général du système de la communication selon Shannon (1948, p.380).....	15
Figure 1.2 : Le modèle de Shannon avec les modifications de Weaver.....	16
Figure 1.3 : Le système de communication selon Belkin (1978, p.81).....	18
Figure 1.4 : Un simple schéma de modèle de rétroaction selon Nobert Wiener.....	31
Figure 1.5 : Le modèle de rétroaction de Nobert Wiener pour la prise de décision.....	32
Figure 1.6 : Le modèle de parties prenantes d'un système (Brenda Freshman 2009, p.352).....	46
Figure 1.7 : Les « forces retenues » influant sur la collaboration.....	46
Figure 1.8 : Les « forces motrices » influant sur la collaboration.....	47
Figure 2.1 : Importance stratégique du management de l'information, adapté de (Menendez et coll. 2002, p.16).....	54
Figure 2.2 : Concepts associés à l'IE (Menendez et coll. 2002, p.19).....	55
Figure 2.3 : le cadre d'intelligence économique.....	63
Figure 3.1 : Modèle imbriqué du comportement informationnel, adapté de (Wilson 1999).....	78
Figure 3.2 : RIN, PIN, demande et requête (Mizarro 1998, p.308).....	84
Figure 3.3 : Un modèle simple de la RI, adapté de (Ingwersen 1992, p49).....	87
Figure 3.4 : Un SRI typique, adapté de (Rijsbergen 1979).....	88
Figure 3.5 Un schéma global regroupant le processus de RI et une architecture générale de SRI.....	88
Figure 3.6 : Un exemple de requête booléenne.....	94
Figure 3.7 : Activation contextuelle des préférences sémantiques de l'utilisateur (Vallet et coll. 2006, p.4).....	110
Figure 3.8 : RI traditionnelle et analyse des réseaux sociaux (Kirsch 2005, p.35).....	111
Figure 3.9 : <i>Modèle de domaine</i> pour la RSI (Kirsch 2005, p.35).....	112
Figure 3.10 : Une carte de différenciation entre la taxonomie, l'ontologie et la folksonomie.....	119
Figure 4.1 : Modèle imbriqué de la pratique informationnelle collaborative, adapté de (Wilson 1999).....	126
Figure 4.2 : Comportements impliqués lors des recherches individuelle et collaborative d'information, d'après Reddy et Jansen (2008, p.266).....	128
Figure 4.3 : Connaissances et métaconnaissances impliquées lors d'une recherche collaborative d'information (Dinet 2007, p.7).....	129
Figure 4.4: Cadre d'analyse de travail cognitif pour la RCI (Fidel et coll. 2004, p.943).....	130
Figure 4.5 : Classification temps-lieu de collecticiels.....	133
Figure 4.6 : Copie d'écran du système de cabri et coll. (1999, p.143).....	135
Figure 4.7 : Interface de visualisation de recherche dans ARIADNE.....	140
Figure 4.8 : Schéma global de l'architecture d'un SRCI (David 1999, p.50).....	143

Figure 4.9 : Interface de STREEMS (David 1999, p.108) .....	144
Figure 4.10 : Interface de SearchTogether .....	146
La figure 4.11 Microsoft SearchTogether .....	147
Figure 4.12 : Vue d'ensemble conceptuelle d'une session collaborative entre deux utilisateurs selon Foley (2009a, p.7).....	148
Figure 4.13 : Historique des pages visitées et notes (Vivian et Dinét 2008, p.95).....	149
Figure 4.14 : Enrichissement de résultats sous Google (Vivian et dinét 2008, p.96).....	150
Figure 4.15 : Notre schéma global représentatif de la RCI .....	154
Figure 4.16 : Graphe d'observation (Odumuyiwa 2009, p.8) .....	155
Figure 4.17 : Graphe d'interaction (Odumuyiwa 2009, p.8).....	155
Figure 4.18 : Dépendance de flux .....	156
Figure 4.19 : Dépendance de partage .....	156
Figure 4.20 : Dépendance d'ajustement .....	156
Figure 4.21: Cadre 3CM.....	158
Figure 5.1 : La pyramide de collaboration .....	163
Figure 5.2 : Modèle COCIR .....	174
Figure 5.3 : Trois éléments du COCIR et leurs sous-éléments .....	175
Figure 5.4 : Représentation de l'utilisateur .....	177
Figure 5.5 : Structure d'annotation sur un document web .....	179
Figure 5.6 : Structure de l'annotation sur un problème informationnel .....	179
Figure 5.7 : Structure d'un message envoyé par la messagerie instantanée.....	180
Figure 5.8 : Représentation d'un problème décisionnel.....	181
Figure 5.9 : Représentation d'un problème informationnel .....	184
Figure 6.1 : L'architecture générale de MECOCIR .....	189
Figure 6.2 : Schéma des composants fonctionnels du système MECOCIR.....	192
Figure 6.3 : protocole de communication pour MECOCIR .....	196
Figure 6.4 : Graphe montrant l'association entre les utilisateurs et les problèmes .....	199
Figure 6.5 : Architecture JDBC ( <a href="http://java.sun.com/products/jdbc/overview.html">http://java.sun.com/products/jdbc/overview.html</a> ) .....	202
Figure 6.6 : Fenêtre d'authentification .....	205
Figure 6.7 : Formulaire d'inscription (saisir des informations personnelles).....	206
Figure 6.8 : Formulaire d'inscription (saisir d'adresse) .....	207
Figure 6.9 : Formulaire d'inscription (saisir des domaines de compétence de l'utilisateur).....	207
Figure 6.10 : Formulaire d'inscription captant des connaissances initiales de l'utilisateur .....	207
Figure 6.11 : L'interface principale de MECOCIR et la page d'accueil.....	208
Figure 6.12 : Menu « problème » .....	211
Figure 6.13 : Interface pour la définition du problème décisionnel .....	211
Figure 6.14 : Interface pour la définition du problème informationnel.....	212

Figure 6.15 : Interface pour définir les indicateurs .....	212
Figure 6.16 : Démarrage d'une session de recherche.....	213
Figure 6.17 : Demande de collaboration .....	214
Figure 6.18 : Notification de la demande de collaboration .....	214
Figure 6.19 : Notification d'acceptation de collaborer.....	215
Figure 6.20 : Interface pour la clarification du problème informationnel .....	216
Figure 6.21 : L'interface d'annotation du problème informationnel.....	216
Figure 6.22 : L'interface de navigation et l'interface d'analyse du document actuel .....	218
Figure 6.23 : Interface de navigation et l'interface pour la conscience des activités de groupe .....	219
Figure 6.24 : L'interface pour la conscience des activités de groupe.....	219
Figure 6.25 : L'évaluation et l'annotation d'un document.....	220
Figure 6.26 L'onglet d'évaluation dans l'interface d'analyse du document actuel.....	221
Figure 6.27 : L'interface de messagerie instantanée .....	222
Figure 6.28 : Le bouton pour prendre la main.....	222
Figure 6.29 : L'interface de l'utilisateur B synchronisé avec celle de A .....	223
Figure 6.30 Le champ « SplitSearch » .....	223
Figure 6.31 : Le résultat de la recherche affiché sur l'écran de l'utilisateur A .....	224
Figure 6.32 : Le résultat de la recherche affiché sur l'écran de l'utilisateur B.....	224
Figure 6.34 : Les moteurs de recherche utilisés par les participants .....	228
Figure 6.35 : la définition du problème informationnel faite par U1 .....	230
Figure 6.36 : les annotations ajoutées par U2 pour clarifier le problème informationnel.....	231
Figure 6.37 : Les moteurs de recherche utilisés par les participants de la deuxième série d'expérimentation.....	234



# Tableaux

Tableau 1.1 : Les quatre processus de conversion de connaissances (Nonaka 1994, p.19) .....	25
Tableau 1.2 : Des exemples de technologies pouvant faciliter la transformation de connaissance (Marwick 2001, p.816) .....	25
Tableau 1.3 : Coopérer ou collaborer ? .....	41
Tableau 2.2 : Modèles développés au sein de l'équipe SITE (Oladejo 2008).....	70
Tableau 2.2 : Notre travail de recherche par rapport au processus de l'IE selon l'équipe SITE.....	71
Tableau 3.1 : Modèle de Khulthau sur le processus de recherche d'information (Khulthau 1999, p.367 ; Khulthau 2004, p.82) .....	82
Tableau 4.1 : Dimensions de l'analyse de travail cognitif et exemples des questions à poser lors d'analyse de chaque dimension, adapté de (Fidel et coll. 2004, p.943) .....	131
Tableau 4.2 : Classification des SRCI existants .....	151
Tableau 5.1 : Les activités de la RCI facilitant la conversion de connaissance .....	167
Tableau 6.1 : Tableau comparatif des SRCI.....	237



# Introduction générale

---

## Contexte de l'étude

Dans un monde en perpétuelle transformation, la prise de décision dans tous les secteurs socio-économiques demande une réactivité permanente surtout avec la globalisation qui s'installe et qui ne fait qu'augmenter les problématiques liées à la prise de décision. Cette dernière est étroitement liée à la maîtrise d'informations externes et internes à l'organisation. On constate qu'au cœur même du fonctionnement des organisations, le flux d'informations ne cesse de s'accroître. L'augmentation exponentielle du volume d'informations pose un défi incontournable à la gestion des activités des organisations et vient impacter directement la question de la prise de décisions.

Entre la période de 1947 à 1991 qui a opposé les deux superpuissances qu'étaient les Etats-Unis et l'Union des républiques socialistes soviétiques, période que l'on a qualifiée de Guerre Froide, l'information était devenue un enjeu capital et un bien précieux. Aujourd'hui même après la détente entre ces deux protagonistes, les enjeux restent tout aussi importants et s'orientent vers les domaines de l'économie, du social, du politique, de la production industrielle. De ce fait l'on peut considérer que l'information reste une matière première qui détermine grandement les orientations du monde qui nous entoure. Il n'est donc pas rare d'entendre parler de la société de l'information et de celle de la connaissance.

La démocratisation de l'internet articulée avec le grand développement des technologies de l'information et de la communication a contribué énormément à l'essor informationnel. Ces évolutions selon Kislin « *ont favorisé l'explosion de l'offre et principalement de la demande en information, essentielle à toute activité de l'Homme, autant pour son adaptation à son environnement que pour la prise de décision* » (Kislin 2007, p.16). Néanmoins, il est à noter que le défi incessant d'obtenir des informations pertinentes susceptibles de répondre aux besoins informationnels d'utilisateurs augmente proportionnellement à l'augmentation du volume d'informations. La question aujourd'hui porte donc sur comment retrouver les informations pertinentes à la résolution d'un problème décisionnel dans les meilleures conditions de qualité, de délai et de coût compte tenu de l'hétérogénéité et la disparité des sources d'informations.



Plusieurs approches ont été développées dans le cadre des études sur la recherche d'information pour traiter le problème de la pertinence. Différentes disciplines ont contribué à l'amélioration des méthodes et systèmes de recherche d'information parmi lesquelles s'inscrivent la science de l'information et de la communication, l'informatique, la linguistique et la psychologie cognitive. Certains auteurs tels que Wilson (1981, p.3-15) et Khulthau (1991, p.361-371), se focalisent sur le *comportement informationnel d'utilisateurs*<sup>1</sup> pour pouvoir donner des indicateurs<sup>2</sup> aux développeurs du système de recherche d'information. D'autres tels que Salton (1970, p.335-343 ; 1983, p.1022-1036) et Maron et Kuhns (1960, p.216-244) ont travaillé sur le développement des algorithmes d'appariement, de classement, d'indexation etc. Tous ces travaux avaient pour objectif de contribuer à la résolution des problèmes liés à la pertinence. L'aspect cognitif de la recherche d'information a été aussi le sujet d'intérêt de plusieurs chercheurs (Belkin et coll. 1982a, p.61-71 ; Ingwersen 1992, p.15-47). Nous remarquons que ces diverses contributions appartenant à des disciplines différentes se complètent les unes les autres.

Dans les différents secteurs socio-économiques où la démarche d'intelligence économique est appliquée dans la résolution des problèmes décisionnels, on identifie clairement l'importance du processus décisionnel et informationnel. La démarche englobe l'identification du problème décisionnel, sa traduction en problème informationnel, l'identification des sources, la collecte et l'analyse d'informations et la prise de décision. Ces différentes phases comportant la recherche d'information ne peuvent être dissociées de la question récurrente de la pertinence et de celle de la validation des sources et de l'information collectée.

Des évolutions ont vu le jour dans les travaux de recherche concernant le domaine de la recherche d'information. Ces dernières s'appliquent aussi bien dans l'élaboration des concepts que dans le développement des algorithmes. Traditionnellement ces travaux de recherche se fondent sur un modèle de l'interaction individuelle entre l'homme et le système de recherche d'information.

---

<sup>1</sup> Le comportement informationnel intègre toutes les pratiques informationnelles des utilisateurs recherchant de l'information, généralement dans des environnements plus ou moins numériques. Il comporte également la recherche active et passive d'information (Wilson 1999) ainsi que l'évitement d'information (Case 2002, p.6-7). Il recouvre la totalité du comportement humain en relation avec des sources ou canaux d'information.

<sup>2</sup> Les indicateurs dans ce contexte comportent tout ce qui permet à comprendre le processus de recherche d'information d'utilisateur. Par exemple, le modèle de Khulthau explique six étapes constituant le processus de l'accès à l'information. Ces étapes peuvent aider un développeur à mieux concevoir et développer un système de recherche d'information.

Au début des années 90, les chercheurs ont repris les études, d'une part, sur les méthodes de recherche d'information et, d'autre part, sur l'architecture de système de recherche d'information. C'est ainsi que l'on observe l'introduction de la dimension sociale, qui prend en compte l'utilisateur et son environnement mais également son désir de collaborer avec d'autres pour résoudre son problème informationnel. De cette réflexion a émané le concept de filtrage collaboratif.

Le terme « filtrage collaboratif » a été utilisé en premier par Goldberg et coll. (1992, p.1) dans le développement de leur système « Tapestry » au centre de recherche Xerox Palo Alto. Ce concept est fondé sur l'utilisation des interactions des utilisateurs sur des objets informationnels afin d'en recommander aux autres. L'hypothèse sur laquelle se fonde ce concept suppose que ce qui est déjà jugé pertinent par des utilisateurs pourrait l'être pour un autre utilisateur ayant un problème informationnel similaire (Bueno et coll. 2001, p.310 ; Berrut et Denos 2003, p.247). Le processus de collaboration dans ce cas n'est pas ouvert aux utilisateurs car ils ne se sont pas mis en relation directe. Il s'agit d'une collaboration implicite ayant pour but d'améliorer le processus de recherche individuelle. Dans ces conditions nous pensons que ce genre d'interaction relève davantage de la recherche sociale d'information plutôt que de la collaboration.

Les chercheurs dans le domaine des sciences de l'information ont poussé plus loin leurs réflexions sur la recherche d'information. Ils constatent que les utilisateurs ont plutôt une tendance à coopérer ou à collaborer, (ces deux termes seront distingués à leur tour dans 1.5.1), dans la résolution de leur problème informationnel. Le cas d'un utilisateur qui demande des clarifications d'un bibliothécaire ou d'un autre expert dans sa démarche de recherche d'informations a été donné à titre d'exemple dans plusieurs travaux. Romano et coll l'ont présenté de cette manière « *D'une part, les recherches et développements dans la RI se sont focalisés quasi exclusivement sur des individus. D'autre part, l'observation des individus effectuant une recherche d'informations montre que quels que soient l'endroit et l'heure de la recherche, soit dans une bibliothèque, une organisation, ou par un groupe de touristes visitant une ville, les ressources d'informations sont souvent utilisées de manière collaborative pour rechercher des informations et pour la prise de décision* »<sup>3</sup> (Romano et

---

<sup>3</sup> « *On the one hand, IR research and development has focused almost exclusively on individuals. On the other hand, observation of people performing searches reveals that, regardless of where and whenever a search occurs, whether in a library, in business organizations or in groups of tourists wandering around cities,*

coll. 1999, p.2). Ayant pris connaissance de cela, ils ont pu développer un environnement de recherche d'information collaborative en fusionnant les paradigmes du collectifiel (*groupware*) et de la RI.

En pratique la recherche collaborative d'information est directement liée à l'intelligence économique et plus particulièrement à l'activité de veille. Par notre propre observation du comportement des utilisateurs en situation de recherche d'informations, nous pouvons faire ressortir les comportements suivantes :

- les utilisateurs se trouvent toujours dans un fonctionnement en boucle qui articule les problèmes informationnels, les sources d'information et les autres utilisateurs ;
- les utilisateurs se servent de plusieurs systèmes d'information dans la résolution de leurs problèmes informationnels ;
- les utilisateurs accordent plus de crédit à l'information donnée par un expert qu'à celle retrouvée dans un système d'information ;
- les utilisateurs suivent l'évolution d'un expert afin d'alimenter leur réflexion et leur pratique ;
- les utilisateurs exploitent les réseaux d'autres utilisateurs qu'ils considèrent plus expérimentés dans leurs domaines ;
- les utilisateurs suivent l'évolution d'un système d'informations qu'ils considèrent comme une source d'information pertinente ;
- les utilisateurs dépendent beaucoup de leurs réseaux soient sociaux ou professionnels dans la résolution de leurs problèmes informationnels.

Les études sur le comportement informationnel d'utilisateurs ont montré que ces derniers recherchent généralement de l'information d'une manière collaborative. Cependant, nous remarquons que certaines pratiques informationnelles des utilisateurs ne sont pas prises en compte dans le développement de système de recherche d'information qu'il soit collaboratif ou pas. Par exemple, la plupart des systèmes de recherche collaborative d'information existants sont développés autour d'un moteur de recherche ou d'une base de données bibliographique. Toutefois l'on remarque que les utilisateurs ne se contentent pas d'un seul

---

*information resources are often used collaboratively to seek in formation and to make decisions* » (Romano et coll. 1999, p.2).

système, il est même fréquent qu'ils se servent de plusieurs sources et systèmes d'information au cours de leur démarche informationnelle. Ils partagent alors entre eux les informations retrouvées, leurs processus de recherche et les sources d'informations.

La RI est un processus à la fois social et cognitif. Cela s'explique par le fait que le problème lié à la satisfaction d'un besoin informationnel est en lien direct avec la compréhension et l'élucidation du problème de l'utilisateur. Bien évidemment la clarification et la résolution du problème sont tributaires de la capacité cognitive de ce dernier. Toutefois, ce manque cognitif peut se compenser en collaborant avec d'autres utilisateurs.

C'est dans cette optique de partage de connaissances que nous définissons la recherche collaborative d'information comme **méthodes** et **systèmes** permettant de gérer les **activités collectives** des utilisateurs dans un processus de recherche d'informations. Dans ce cas, il s'agit ici d'une collaboration explicite.

### **Hypothèse et problématique**

Il y a une croyance générale que l'efficacité d'un groupe ne correspond pas à la somme des efficacités individuelles des membres du groupe. Ceci sous-entend qu'un groupe d'individus collaborant autour d'un problème aura une meilleure solution que la simple agrégation des solutions individuelles des membres du groupe. Transposant cette croyance à la RI, nous pouvons supposer que les problèmes liés à la RI, surtout dans le cadre de l'intelligence économique, peuvent être mieux et plus rapidement résolus si l'approche de la recherche collaborative d'information est utilisée. En effet, les avancées récentes dans la technologie du web et les systèmes et les technologies de l'information et de la communication (STIC) ont permis à plusieurs utilisateurs de travailler ensemble ou de coopérer à l'exécution des tâches assistées par ordinateur. Ces avancées ont considérablement transformé la conception et le développement des méthodes et systèmes de recherche d'informations.

L'importance de la collaboration dans la résolution d'un problème informationnel individuel ou organisationnel nous a guidé, dans le cadre de cette thèse, à mener une réflexion qui vise à **faciliter la gestion** des activités de la recherche collaborative d'information (RCI).

Nous formulons l'hypothèse qu'une bonne gestion des connaissances devrait faciliter dans un premier temps la RCI et dans un deuxième temps, permettre d'améliorer le résultat global du processus d'IE dans le cadre de la résolution d'un problème décisionnel.

Bien que plusieurs auteurs aient tenté d'éclairer le concept de la RCI et de développer des systèmes de RCI, nous remarquons qu'ils n'abordent ce sujet que du point de vue d'un groupe d'utilisateurs collaborant sur un problème informationnel partagé. Ces auteurs n'ont pas étudié le problème de compréhension du problème par le groupe. En réalité, un problème partagé n'implique pas forcément une compréhension partagée du problème. Le niveau de compréhension d'un problème informationnel détermine grandement la qualité et la rapidité de la résolution quelle que soit la nature de l'intervention qui peut être individuelle ou collective.

Notre hypothèse nous mène à nous poser les questions suivantes :

- Comment passer d'un problème informationnel partagé à une compréhension partagée du problème ? Sachant que le point de départ dans une collaboration devrait toujours passer par la question : pourquoi collabore t-on et pour atteindre quel objectif?
- Comment gérer les différents types de connaissances mises en œuvre dans la recherche collaborative d'information ?
  - Quelles sont les connaissances à partager lors de la RCI ?
  - Comment acquérir ces connaissances ?
  - Comment coordonner les interactions entre les acteurs impliqués dans la collaboration afin de faciliter le partage de connaissances entre eux ?
- Comment assurer la conscience d'appartenance à un groupe<sup>4</sup> (*Awareness*) et la répartition des tâches (*division of labour* ) dans une activité de recherche d'information ?
- Comment garantir ou assurer la sécurité et la protection du patrimoine dans un système de recherche collaborative d'information ?

---

<sup>4</sup> « La conscience de groupe permet la compréhension de l'activité des autres, ce qui fournit un contexte pour sa propre activité. Ce contexte permet d'assurer que les actions individuelles sont situées dans l'activité globale du groupe [ . . . ] Cette information permet au groupe de gérer le processus de collaboration » (Dourish 1992, p.1). Traduction reprise de la thèse de doctorat de madame Sawsan Alshattawi (2008, p.23).

## **Structure de la thèse**

Chapitre 1 : Notre première préoccupation dans ce chapitre sera de présenter les trois concepts principaux qui se réunissent dans la collaboration à savoir : l'information, la connaissance et la communication. Nous expliquerons également le concept de collaboration dans son sens large et les facteurs sociocognitifs associés à la collaboration.

Chapitre 2 : Ce chapitre sera consacré à notre domaine d'application qui est l'Intelligence Economique (IE). Nous présenterons ce concept ainsi que son évolution. Ensuite, nous aborderons les processus d'IE comme définis par l'équipe de recherche SITE du LORIA. Le concept de la veille sera également abordé.

Chapitre 3: Dans ce chapitre, nous évoquerons le concept de la recherche d'information (RI). Nous nous attellerons à présenter les aspects historiques afin de retracer les évolutions dans ce domaine. Puis nous évoquerons les problèmes liés aux systèmes traditionnels de recherche d'information et leur évolution vers un système de recherche sociale d'information.

Chapitre 4 : Ce chapitre sera consacré à la RCI. Nous commencerons par l'étude du comportement informationnel collaboratif puis nous expliquerons la RCI. En nous appuyant sur la présentation des travaux qui ont été effectués dans ce domaine, nous tenterons de développer un cadre conceptuel pour la compréhension du concept de la RCI.

Chapitre 5 : Ce chapitre présentera nos propositions pour la résolution de la problématique de départ. Nos propositions consistent en le modèle de pyramide de collaboration et le modèle de communication. La pyramide de la collaboration montre les différents composants et phases nécessaires pour la réussite d'un processus de collaboration surtout dans le cadre de la RI. Le modèle de communication est une modélisation du contexte collaboratif pour le partage de connaissances.

Chapitre 6 : Ce chapitre expliquera notre démarche pour la conception et la réalisation d'un système de RCI. Nous présenterons notre protocole expérimental et les conclusions tirées de l'expérimentation



*“The lure of the next bend behind which new sights might appear distracts us from the labor of taking stock and the effect of this distraction is reinforced by the anxiety that our theories might be defeated at the next turn” (Polanyi 1966 p. ix).*





# Chapitre 1

---

## Contexte d'étude et vue globale de la collaboration

### 1.1 Introduction

Ce travail de thèse s'inscrit dans le domaine des sciences de l'information et de la communication, domaine à caractère pluridisciplinaire. Deux des principaux concepts du domaine sont l'information et la communication. Ces deux concepts sont à la base de nos préoccupations dans le cadre de cette thèse. Nous rejoignons la position de Pickens et coll (2008, p.1)<sup>5</sup>, et pensons que la recherche collaborative d'information se situe à l'intersection de la recherche d'information et de la communication qui sont les utilisations les plus répandues de la technologie de l'internet. Nous expliquerons en premier lieu le concept d'information et le concept de connaissance. Ensuite nous aborderons la notion de la communication sous quatre ordres de réalité. Enfin nous nous appuierons sur quelques modèles de communication pour illustrer les concepts d'information et de communication qui avec la connaissance s'articulent dans l'acte de collaboration. Ceci nous conduira à distinguer la collaboration de la coopération afin de pouvoir aborder par la suite les facteurs sociocognitifs influant sur la collaboration.

### 1.2 A propos du concept d'information

Le mot information, forme nominale du verbe « informer », est apparu dans la langue française au XIII<sup>e</sup> siècle. Le mot latin *informatio* signifiait « conception, explication (d'un mot) », « dessein, esquisse » mais ces sens-là ne furent pas attestés en français. En français, le mot information a été d'abord attesté au sens juridique « enquête judiciaire », puis « renseignement » et par extension désigne, au pluriel (informations), l'ensemble réuni sur un sujet donné. À partir du XIX<sup>e</sup> siècle, avec le développement de la presse, le mot information a pris un autre sens celui de: « événements communiqués au public, nouvelles » (Dictionnaire

---

<sup>5</sup> « ...considering that the two most common uses of Internet technologies are communication and search; collaborative information retrieval systems exist at the intersection of these two trends » (Pickens et coll 2008, p.1).

étymologie et historique de la langue française, 1996 ; Le Robert : Dictionnaire historique de la langue française, 1995).

Selon le travail de Capurro et Hjørland (2003, p.343-411), repris en français par Leleu-Merviel et Useille, deux contextes principaux se détachent de l'étymologie latine du terme information : « soit l'information correspond à l'acte de donner une forme à l'esprit, soit à celui de communiquer des connaissances<sup>6</sup> » (Leleu-Merviel et Useille 2008, p.27). Nous commençons par une explication du concept d'information selon le travail récent de Bates dans lequel l'information est considérée comme un phénomène.

### 1.2.1 Formes fondamentales de l'information selon Bates

L'une des définitions d'information qui nous paraît intéressante est celle donnée par Parker (1974, p.10) mais élaborée par Bates (2006, p.1). Selon cette définition, « l'information est un *pattern* d'organisation de matière et d'énergie »<sup>7</sup>. Cette définition peut apparaître comme très provocante surtout pour les acteurs dans le domaine des sciences de l'information mais à la réflexion on y trouve du sens. En effet, il nous semble important de préciser que Bates en se fondant sur cette définition voulait simplement donner une définition fondamentale au phénomène « information » sur laquelle d'autres définitions pourraient s'appuyer. Il explique que tout ce qui existe dans l'univers et pouvant venir à la portée de l'homme ou les autres êtres vivants peut être vécu comme de l'information, exemple : le chant d'un oiseau, « halo » dit par un ami, l'intuition que nous avons sur l'honnêteté d'un ami avec qui l'on parle, un livre que nous lisons, etc.

Par un « *pattern* d'organisation », il entend, en premier lieu, toute sorte d'arrangements ou combinaisons qui n'équivalent pas au chaos ou à une désorganisation. En deuxième lieu, cela implique une combinaison des qualités, des actes, des tendances, etc., qui aboutissent à un mode régulier ou une combinaison typique – ex. le mode de comportement des adolescents. Quand il parle de l'information en tant que *pattern* d'organisation de la matière, il s'agit de différentes matières par exemple, la roche, la terre, la matière du cerveau, etc. En parlant de l'information en tant que *pattern* d'organisation de l'énergie, il fait allusion au mode d'organisation de l'énergie d'une parole quand elle fait vibrer l'air. Selon lui la seule chose

---

<sup>6</sup> « *The act of molding the mind and the act of communicating knowledge* » (Capurro et Hjørland 2003, p.351)

<sup>7</sup> « *Information is the pattern of organisation of matter and energy* » (Parker 1974, p.10) (Bates 2006, p.1)

dans l'univers qui ne contient pas d'information est l'entropie totale ; elle est sans modèle. Sa définition de l'information porte un sens très large, mais il a appuyé son raisonnement sur le fait que les êtres humains peuvent être influencés et agir à partir de toutes sortes d'informations contenues dans leur environnement.

Bates ne s'arrête pas sur la définition ci-dessus mais à partir de celle-ci, il propose trois points de vue :

- Information 1 : *Le pattern* d'organisation de matière et d'énergie.
- Information 2 : Un *pattern* d'organisation de matière et d'énergie auquel les êtres humains donnent un sens<sup>8</sup>.
- Connaissance : De l'information dotée de significations et intégrant d'autres éléments de compréhension<sup>9</sup>. Notons que cette définition récursive montre la difficulté associée à la définition de l'information ainsi que celle associée à la dissociation de l'information de la connaissance.

Il rapproche « Information 1 » au fondement sur laquelle la théorie de sémiotique<sup>10</sup> est fondée ainsi que le fondement sur lequel les actes individuels d'interprétation de signe s'opèrent. Il explique la différence entre « Information 1 » et « Information 2 » selon cet exemple : « Dans certaines sociétés, la façon de faire un geste de la main peut se traduire en « viens ici » »<sup>11</sup>. Si un ami me fait ce type de geste auquel je donne la même signification, c'est-à-dire: « viens ici », je suis informé que mon ami veut que je le rejoigne. Dans ce cas, ce mouvement de main est considéré comme « Information 1 ». L'interprétation que j'en fais « viens ici » est considérée comme « Information 2 » (Bates 2006, p.9).

---

<sup>8</sup> « *Information 2: some pattern of organization of matter and energy given meaning by a living being (or its constituent parts)* » (Bates 2006, p.10)

<sup>9</sup> « *Knowledge: Information given meaning and integrated with other contents of understanding* » (Bates 2006, p.10)

<sup>10</sup> La sémiotique est « la science générale des modes de production, de fonctionnement et de réception des différents systèmes de signes qui assurent et permettent une communication entre individus et/ou collectivités d'individus » (Larousse.fr 2009). Elle « étudie le processus de signification c'est-à-dire la production, la codification et la communication des signes » (Wikipédia consulté le 14 Juin 2010).

<sup>11</sup> « *For example, in various cultures over the years, waving a hand toward oneself has come to mean, "come here" Thus, a longterm semiotic process has resulted in that association being present in many human beings' minds. My culture shares that association. When, at a party, I interpret that hand motion from a friend across the room as a sign to come join him, I am selecting out a certain subset of all the information around me* » (Bates 2006, p.9).

« Information 2 », porteuse du sens, construite à partir des connaissances de l'émetteur, est propagée via la parole, l'écriture ou d'autres moyens. Un destinataire reçoit d'abord l'information comme « Information 1 » soit en forme de sons ou de signes d'écriture. Cette information ne porte aucun sens jusqu'à ce qu'elle soit interprétée par ce dernier. L'interprétation donnée par le destinataire peut différer de l'intention de l'émetteur.

Cela implique que « Information 2 » construite par le destinataire peut être différente de « Information 2 » produite par l'émetteur. Éventuellement, « Information 2 » construite par le destinataire s'intègre plus ou moins à ses propres connaissances. Ceci explique ce que Bates entend par information en tant que connaissance dans les définitions ci-dessus. Selon lui une information devient une connaissance chez le récepteur (destinataire) quand il lui donne une interprétation qui est par la suite intégrée dans ses connaissances déjà acquises.

Dans la littérature, le mot « information » a fait l'objet de plusieurs définitions par plusieurs auteurs. Il existe encore aujourd'hui des désaccords sur sa définition. Nous sommes d'accord avec Belkin pour ne pas nous focaliser sur les définitions de l'information, mais plutôt sur les concepts d'information. La distinction est qu'une définition nous donne vraisemblablement ce qu'est un phénomène, alors qu'un concept est à la fois une représentation intellectuelle, mais également une façon d'appréhender le phénomène. L'idée de concept nous permet de chercher une conception utile de l'information au lieu de nous disperser en cherchant sa définition universellement vraie (Belkin 1978, p.58).

Pour aborder le concept d'information, il nous semble important de commencer par une définition donnée par Belkin. Selon lui, le problème fondamental de sciences de l'information est de faciliter la communication efficace de l'information entre une personne productrice et une personne réceptrice. En nous fondant sur cette définition, nous élaborons le concept d'information sous deux angles : trois perspectives de l'information et trois manifestations de l'information.

## **1.2.2 Trois perspectives de l'information**

### ***1.2.2.1 Information selon Shannon***

Le modèle de Shannon (1948) est parmi les concepts de l'information les plus cités. Son modèle est régulièrement expliqué en conjonction avec la proposition de Weaver. En fait, ce

modèle est fréquemment traité dans la théorie de la communication ainsi que celle de l'information. Le concept d'information selon Shannon à son origine, était proposé pour le domaine de télécommunication. Grâce à sa formalisation et son application conséquente, il a été appliqué dans plusieurs autres domaines (Belkin 1978, p.66). Shannon ne s'occupait pas du contenu du message ou autrement dit de la sémantique du message. Il se focalise plutôt sur la probabilité de choisir un message désiré parmi un ensemble de messages possibles. Il précise que « *le problème fondamental de la communication est de reproduire, d'une manière exacte ou approximative en un point, un message venant d'un autre point* »<sup>12</sup> (Shannon 1948, p.379). Il propose un schéma général du système de communication comme le montre la figure 1.1.

Ce système est composé de cinq éléments :

- Source d'information : représente la source productrice de l'information ou d'un message à communiquer au récepteur.
- Emetteur : encode le message pour produire des signaux pouvant être transmis par le canal.
- Canal : est le moyen utilisé pour transmettre le signal de l'émetteur au récepteur.
- Récepteur : décode les signaux parvenant du transmetteur afin de reconstruire le message.
- Destinataire d'information : celui à qui le message est destiné.

Shannon introduit la notion de bruit dans son schéma. Le bruit consiste en toutes sortes d'interférences susceptibles de détériorer le signal entre le point d'entrée du canal et le point de sortie.

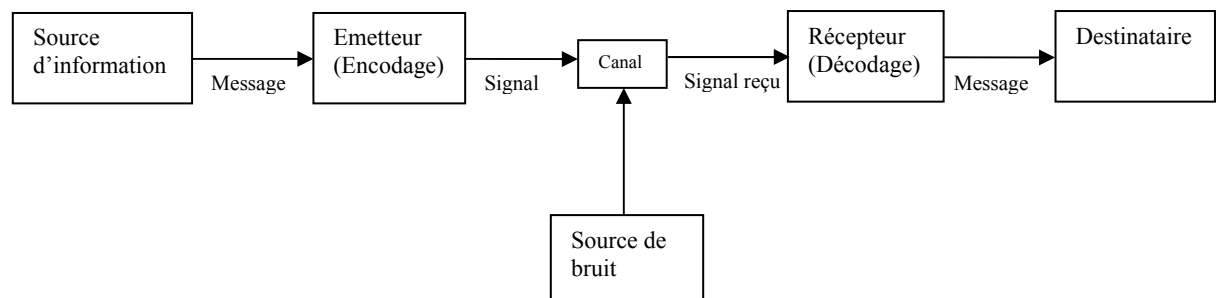


Figure 1.1 : Un schéma général du système de la communication selon Shannon (1948, p.380)

<sup>12</sup> « *The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point* » (Shannon 1948, p.379)

Plusieurs auteurs ont critiqué et analysé le concept d'information présenté par Shannon selon son modèle de communication (Artandi 1973, p.242-244 ; Belkin 1978, p.66-69). Nous considérons ce modèle pertinent dans le cadre de la communication de l'information mais le concept d'information tel qu'il est présenté dans ce modèle est très limitatif compte tenu du fait que nous sommes essentiellement concernés par le « sens » d'une information. Néanmoins, ce concept d'information peut être rapproché de la théorie de la sémiotique comme démontré dans (Artandi 1973, p.243). Artandi explique que le concept d'information de Shannon correspond à la dimension syntactique dans le contexte de la sémiotique. Les deux autres dimensions – sémantique et pragmatique – dans la sémiotique ne sont pas prises en compte par Shannon. Weaver dans un ouvrage co-écrit avec Shannon identifie deux autres niveaux dans la communication de l'information : la sémantique et l'efficacité (Shannon et Weaver 1949, p.32-35). Ces deux niveaux correspondent en gros aux niveaux sémantique et pragmatique du contexte de sémiotique (Artandi 1973, p.243).

Weaver en effet, a introduit une dimension humaine au schéma purement technique de Shannon. Selon lui le bruit dans le modèle de Shannon n'est qu'un type de bruit – le bruit technique. Il en ajoute un autre type de bruit – le bruit sémantique – entre source d'information et émetteur rendant compte de phénomènes de perturbations ou distorsions de signification. Le bruit sémantique consiste en toutes sortes d'éléments susceptibles de perturber le codage et le décodage sémantique (fatigue, distraction, maladie, ivresse, préjugés...). Un autre élément introduit par Weaver est le récepteur sémantique qui se trouve entre le récepteur technique (décodage) et le destinataire. Cet élément montre le fait que l'information a pour but d'être interprétée. Le récepteur sémantique soumet le message à un deuxième décodage. Il s'agit d'un décodage sémantique pour attribuer un sens aux mots reconstitués en accordant les caractères sémantiques des messages avec les possibilités sémantiques des destinataires.

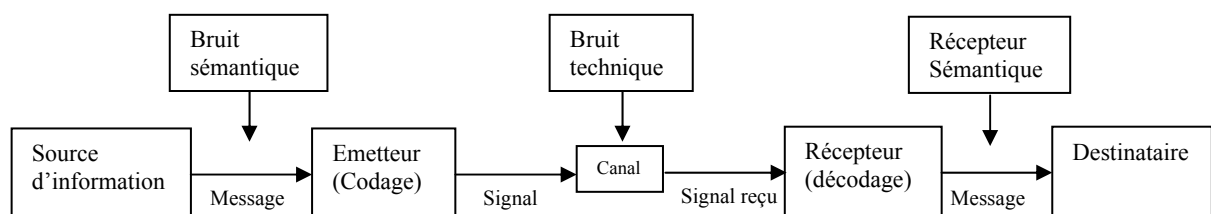


Figure 1.2 : Le modèle de Shannon avec les modifications de Weaver

Le modèle de Shannon et les modifications apportées par Weaver, malgré ses limitations, ont servi comme point de départ pour plusieurs travaux de recherche.

### **1.2.2.2 Point de vue cognitif de l'information**

La dimension cognitive d'information est très importante parce que nous nous fondons en grande partie sur la définition du problème fondamental de science de l'information selon Belkin. Dans le contexte cognitif, l'information peut être considérée comme le résultat d'un changement de la structure cognitive d'un individu. Cette façon de voir est soutenue par F.B. Thompson (1968) mais dans un sens général : « *Notre habitude subjective consiste à organiser les éléments individuels de nos expériences, à les croiser et à les corrélérer avec d'autres éléments distants en espace et temps. Nous ne serons informés qu'après que ce processus s'est réalisé* »<sup>13</sup> (Thompson 1968, p.305). De son point de vue, la structuration de ce que nous expérimentons est l'événement informatif mais, l'information peut être considérée comme la structure résultant de tel événement. Selon Belkin et Robertson (1976, p.198), « *l'information est ce qui est capable de transformer la structure mentale* »<sup>14</sup>. Ils expliquent encore ce concept d'information (en prenant information comme un « texte »), comme « *une ensemble de signes structurés délibérément par un expéditeur avec l'intention de changer la structure mentale du destinataire* »<sup>15</sup>.

Sous cet angle de réflexion, une information a pour but de changer la structure cognitive du récepteur humain. Le changement de structure ne s'effectue pas seulement chez le récepteur, mais le générateur de l'information subit aussi une transformation de sa structure cognitive lors de la construction de l'information à communiquer. Prenons un exemple d'un chercheur qui veut écrire un article, cette décision d'écrire concerne le fait de vouloir communiquer l'avancée de ses connaissances sur un sujet précis. Cette décision impose une partition et donc un tri de ses connaissances qu'il va structurer en isolant certains aspects qui ne correspondent pas à ce qu'il souhaite communiquer. Cela représente la structure conceptuelle de l'objet de sa communication. Celle-ci intègre son intention ainsi que sa connaissance sur les récepteurs potentiels (dans ce cas, une communauté scientifique).

---

<sup>13</sup> « *It is our subjective habit to organize the individual elements of our experience, to cross-correlate these elements to others distant in space and time, and it is only after this process of imposing organization that we feel informed* » (Thompson 1968, p.305)

<sup>14</sup> « *Information is that which is capable of transforming structure* » (Belkin et Robertson 1976, p.198)

<sup>15</sup> « *A collection of signs purposefully structured by a sender with the intention of changing the image structure of a recipient* » (Belkin 1978, p.80) (Belkin et Robertson 1976)



Belkin explique ce concept d'information à l'aide de deux schémas comme le montre la figure 3, en deux niveaux : le niveau linguistique et le niveau cognitif.

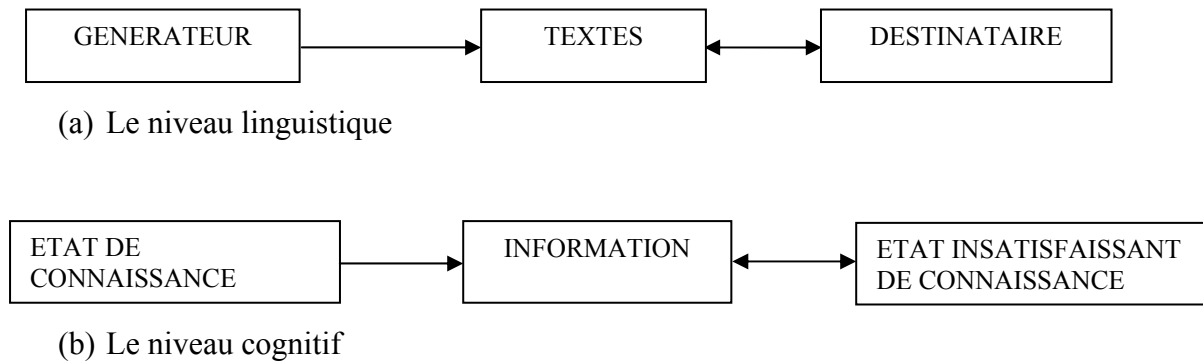


Figure 1.3 : Le système de communication selon Belkin (1978, p.81)

Utilisant les schémas de Belkin pour expliquer l'exemple ci-dessus, le chercheur représente le générateur. La structure conceptuelle de ce qu'il veut communiquer est transformée en une structure communicable (le texte au niveau linguistique et l'information au niveau cognitif). Le destinataire qui se reconnaît un manque cognitif (état insatisfaisant de connaissance<sup>16</sup>) transforme ce manque en une structure communicable (ex. requête) afin de retrouver dans le texte du générateur (l'article du chercheur) les éléments du texte qui peuvent l'aider à augmenter son état de connaissance. Le destinataire interprète le texte du chercheur pour découvrir la structure conceptuelle dont le texte est issu. Son interprétation interagit avec l'état insatisfaisant de ses connaissances. C'est par ce processus qu'il élabore sa propre interprétation pour aboutir à la résolution de son manque cognitif.

Une autre approche qui explique ce concept cognitif de l'information vient de (Bateson 1973, p.271-272) là où l'information est définie comme « une différence qui produit une autre différence ». La différence peut être positive ou négative. Cela veut dire que si une information nous aide à réduire l'incertitude, elle sera perçue comme une différence positive mais si, au contraire, une information augmente l'incertitude, alors, elle sera perçue comme une différence négative. La formalisation de Brookes (1975, p.48) explique mieux le point de vue cognitif de l'information :

$$K(S) + \Delta I = K(S + \Delta S) \text{ ----- (1)}$$

<sup>16</sup> *Anomalous state of knowledge(ASK)* (Belkin 1978, p.80-81)

$$\Delta I = K(S + \Delta S) - K(S) \text{ ----- (2)}$$

Cette formule explique qu'un état existant de connaissance,  $K(S)$ , est affecté par une augmentation d'information,  $\Delta I$ , pour produire un nouvel état de connaissance  $K(S + \Delta S)$ . En rapprochant cette formule à la définition de Bateson,  $\Delta I$  représente la différence qui produit une différence  $\Delta S$  dans l'état de connaissance  $K(S)$  d'un récepteur de l'information.

### 1.2.2.3 Point de vue social de l'information

L'homme, soit dans sa production de l'information ou son interprétation de l'information ne peut être « *séparé de la situation spécifique dans laquelle il est impliqué socialement ou de façon pragmatique* » (Capurro 1992, p.90). Ce point de vue explique qu'une information est associée au contexte culturel et social de son producteur ainsi que de son récepteur. Capurro rapproche ce concept à l'herméneutique<sup>17</sup> en expliquant que la science de l'information étudie « *les dimensions contextuelles pragmatiques dans lesquelles la connaissance est partagée positivement en tant qu'information et négativement en tant que désinformation [...]* »<sup>18</sup>. Il a d'abord expliqué que l'information dans le sens herméneutique-existentielle<sup>19</sup>, consiste à partager thématiquement et en fonction de la situation, un monde commun<sup>20</sup>. Cela veut dire que « *l'information est l'articulation d'une compréhension pragmatique d'un monde commun partagé* »<sup>21</sup>. Nous ne sommes pas des individus isolés mais nous existons dans le monde avec les autres humains. Ceci, du point de vue herméneutique, joue sur notre compréhension du monde ainsi que sur les théories que nous formons sur le monde (Capurro 1992, p.87-88). Notre façon d'interpréter une information est alors influencée par notre communauté d'appartenance et donc par notre culture. Pour nous, information du point de vue de l'approche sociale, implique que notre connaissance que nous communiquons en tant qu'information est une construction sociale, situationnelle et pragmatique.

<sup>17</sup> Herméneutique est un art de l'interprétation (Le dictionnaire des sciences humaines 2006). Elle est une théorie de l'interprétation des signes comme éléments symboliques d'une culture (Larousse en ligne 2009)

<sup>18</sup> Citation tirée de la version modifiée de (Capurro 1992). Cette version modifiée est disponible en ligne à <http://www.capurro.de/tampere91.htm> (consulté le 10 Août 2009)

<sup>19</sup> *Existential-hermeneutic*

<sup>20</sup> « *Information, in an existential-hermeneutic sense, means to thematically and situationally share a common world* » (Capurro 1992, p.89)

<sup>21</sup> « *Information is, more precisely, the articulation of a prior pragmatism of a common shared world* » (Capurro 1992, p.89)

### **1.2.3 Manifestations de l'information**

Buckland (1991, p.351) a donné une topologie de l'information par laquelle il l'aborde sous trois angles :

- L'information en tant que processus,
- L'information en tant que connaissance,
- L'information en tant qu'objet.

#### ***1.2.3.1 L'information en tant que processus***

L'information peut être considérée comme le processus ou l'action d'informer ou d'être informé. Ce processus implique la communication de l'information. Par exemple, le fait d'explicitement une connaissance à communiquer en écriture est un processus d'information. Le processus d'interprétation d'une information pour en donner le sens est un processus d'information. En résumé, tout processus qui produit de l'information ou qui résulte dans l'acquisition d'information est un processus d'information.

#### ***1.2.3.2 L'information en tant que connaissance***

Pendant le processus d'information comme expliqué ci-dessus, tout ce que nous percevons est considéré comme information. L'information en tant que connaissance correspond à la connaissance exprimée lors d'une communication qu'elle soit la connaissance sur un objet ou sur une personne ou sur un sujet. Cette notion d'information est applicable également lorsque l'on considère l'information comme un élément qui aide à réduire l'incertitude.

#### ***1.2.3.3 L'information en tant qu'objet***

Nous pouvons considérer l'information en tant qu'objet comme des données ou des documents. Pour communiquer une information d'une personne à l'autre, « l'information en tant que connaissance » doit être rendue explicite ou exprimer ou décrire en objet comme parole, texte, ou signal etc. Cette forme tangible de l'information est considérée comme « information en tant qu'objet »

### 1.3 A propos du concept de connaissance

Dans notre explication du concept d'information, nous remarquons une forte relation entre l'information et la connaissance. Il nous semble important de traiter le concept de connaissance d'une manière plus approfondie afin de bien différencier la connaissance de l'information. Nous nous attacherons à expliquer la transformation de connaissances en information et vice-versa. Comme le phénomène d'information, la connaissance a fait l'objet de plusieurs définitions par différents auteurs selon leur discipline. Ce concept est étudié par les disciplines telles que la philosophie, l'épistémologie, la psychologie, les sciences cognitives, l'anthropologie et la sociologie.

On constate que la connaissance est un objet central pour les philosophes et tout comme le phénomène de l'information, la définition de la connaissance reste un objet de débat. Néanmoins, il nous semble important d'expliquer trois types de connaissance selon le point de vue traditionnel en philosophie, à savoir :

- « La connaissance propositionnelle est le fait de savoir qu'une certaine proposition est vraie, par exemple, « savoir que la Terre est ronde » »<sup>22</sup>. Ce point de vue de la connaissance a grandement retenu l'attention des philosophes.
- La connaissance d'un objet est le fait de connaître une chose particulière, par exemple, « connaître le président actuel de la France ».
- « Le savoir-faire est le fait d'être capable de réussir une action, par exemple « savoir faire des crêpes » »<sup>23</sup>.

Dans le cadre de ce travail de thèse, comme nous l'avons fait pour le phénomène de l'information, nous n'avons pas souhaité rechercher une définition idéale de la connaissance mais plutôt expliquer certains concepts qui favorisent l'élaboration de notre propos sur la collaboration. De ce fait, nous nous sommes plus particulièrement intéressés à la notion de la connaissance tacite et explicite ainsi qu'à la transformation de l'une sur l'autre.

#### 1.3.1 La connaissance tacite et explicite

La notion de connaissance tacite fut introduite par Polanyi dans les années 1950. Dans son ouvrage « *The tacit dimension* », ses explications sur la connaissance humaine indiquent que

---

<sup>22</sup> (Wikipedia consulté le 30 Août 2009)

<sup>23</sup> (Wikipedia consulté le 30 Août 2009)

« nous pouvons connaître plus que nous ne pouvons raconter <sup>24</sup> » (Polanyi 1966, p. 4). Il appuie ce fait par un exemple : nous connaissons le visage de quelqu'un et pouvons le reconnaître parmi un millier d'individus ou même un million d'individus. Cependant, nous n'arrivons que difficilement à dire comment nous reconnaissons un visage. Sa notion de connaissance tacite couvre les deux verbes (connaître et savoir)<sup>25</sup> qui renvoie à la connaissance. *Nous pouvons donc aussi « savoir » plus que nous ne pouvons raconter.*

Polanyi ajoute que la structure basique de la connaissance tacite comporte deux choses (ce qu'il nomme à travers deux termes du *tacit knowing*) : la conscience (*awareness*) de la personne et la conscience subsidiaire d'une tâche qu'elle effectue. Il y a une relation logique entre ces deux termes de la connaissance tacite, mais les deux s'excluent mutuellement. Il justifie cette distinction en notant que : « *l'on ne peut focaliser son attention sur ce qui nous permet de réaliser la tâche que l'on tient à réaliser, puisque l'on se focalise déjà sur la tâche à réaliser. Ainsi, dans le cadre de la réalisation d'une tâche comme celle de jouer du piano, le musicien se concentre sur sa partition et non sur le mouvement de ses doigts sur le clavier. Et pourtant, il maîtrise l'art de jouer du piano, mais cette connaissance subsidiaire n'est pas l'objet de son attention particulière : sa perception focale* »<sup>26</sup>.

Nonaka, s'inspirant des travaux de Polanyi, et dans son but de formuler une théorie de l'apprentissage organisationnelle qui focalise sur la transformation de la connaissance entre ces deux formes (Nonaka 1991, p.98-99), a expliqué la connaissance tacite comme la connaissance qui est difficile à matérialiser ou à exprimer par des mots mais qui peut tout de même se transmettre par l'observation et par l'imitation. Il a introduit l'adjectif « explicite » pour décrire la connaissance appelée « codifiée » par Polanyi. Il différencie la connaissance tacite de la connaissance explicite en expliquant que cette dernière peut être formalisée ou exprimée et transférée facilement (Nonaka 1994 p.16-17).

---

<sup>24</sup> « *We can know more than we can tell* » ( Polanyi 1966, p. 4)

<sup>25</sup> Sur le plan linguistique, « le verbe savoir se construit grammaticalement avec un groupe verbal à l'infinitif (« je sais travailler », « je sais résoudre un problème », « je sais me rendre utile »), alors que le verbe connaître se construit avec un groupe nominal complément d'objet direct (« je connais le directeur », « je connais la théorie des systèmes », « je connais le sens de cette phrase ») » (Bruneau 1992, p.19). Selon Gorla, « On peut souvent dire que le savoir se forge, se construit ou se développe, tandis que la connaissance se collecte, se recueille, se cumule ou se situe » (Goria 2006, p.84)

<sup>26</sup> Traduction prise de (Goria 2006, p.87)

Pour nous, les connaissances tacites sont les connaissances qui appartiennent au monde des objets mentaux. Elles regroupent les compétences innées ou acquises, le savoir-faire et l'expérience. Elles sont généralement difficiles à «formaliser» par opposition aux connaissances explicites qui sont des connaissances clairement articulées au niveau d'un document écrit, d'un système informatique, ou d'un document électronique et qui sont facilement transférables physiquement, car elles apparaissent sous une forme tangible.

Hildreth et coll (1999, p.347-349) font une distinction entre connaissance dure « hard » et souple « soft » dans une organisation. Pour eux, la plupart des travaux sur « knowledge management » portent sur des connaissances dures dont l'accent est mis sur l'acquisition – représentation – stockage. Ils considèrent les systèmes experts comme étant une sorte de gestion de connaissances dures. La connaissance « souple » peut être considérée comme tacite tandis que la connaissance « dure » peut être considérée comme explicite.

Newman et Conrad. (2000, p.2) pensent pour leur part que la plupart des individus sont impliqués dans un flux de connaissances (knowledge stream) à travers des artefacts. Pour eux, un simple artefact de connaissance peut être explicite, implicite ou tacite. En ce qui concerne le tacite et l'explicite, ils rejoignent d'autres auteurs comme Nonaka et Marwick mais ils ont décrit les artefacts implicites de la connaissance comme des artefacts dont le sens ne peut être capturé explicitement mais inféré ; en effet leur processus de codification n'est pas complet. Ils ont proposé un modèle général de connaissance qui comprend 4 processus, à savoir : la création, la rétention, le transfert et l'utilisation de connaissances.

*La création de connaissances* : consiste en des activités associées au stockage de nouvelles connaissances dans le système. Ces activités incluent le développement, la découverte et la capture de la connaissance.

*La rétention de connaissances* : consiste en des activités qui maintiennent la viabilité de la connaissance dans le système.

*Le transfert de connaissances* : consiste en des activités associées au transfert de la connaissance.

*L'utilisation de connaissances* : consiste en des activités et événements associés à l'application de la connaissance aux processus de business (business processes).

### 1.3.2 La transformation de connaissance entre types tacites ou explicites

Nonaka (1991, p.98-99 ; 1994, p.18-19) a défini quatre processus de conversions de connaissances (cf. tableau 1.1) à savoir : la socialisation, l'extériorisation, l'intériorisation et la combinaison. Marwick (2001, p.816) dans son travail sur « knowledge management technology » a donné quelques exemples de technologies (cf. tableau 1.2) qui permettent la transformation de connaissances pour chacun de processus de Nonaka.

La socialisation permet l'acquisition de connaissances tacites chez un individu ce qui résulte d'un partage de connaissances tacites avec un autre individu. Elle comporte la construction et la communication partagée de connaissances tacites entre les personnes. Ceci se fait souvent sans la création d'une connaissance explicite. Le partage de connaissances tacites est lié à la notion de communauté et de collaboration. Il se manifeste souvent pendant les activités comme *brainstorming*, réunion d'équipe, et discussions.

L'extériorisation est un processus d'explicitation de connaissances tacites d'un individu. Cela permet une conversion de connaissances tacites en connaissances explicites. La connaissance tacite est difficile à exprimer par nature. En revanche, une partie des connaissances tacites d'un individu pourrait être capturée en forme explicite par un processus de conceptualisation, d'élucidation et même d'articulation lors d'une collaboration avec les autres (Marwick 2001, p.815).

L'intériorisation est l'inverse de l'extériorisation car elle permet l'acquisition de connaissances tacites à partir de connaissances explicites. Par exemple : on étudie et consulte plusieurs ouvrages (connaissances explicites) pour former sa propre connaissance tacite. Lors de la lecture, un individu interprète ce qu'il lit et l'intériorise pour construire sa connaissance tacite. Le processus d'intériorisation est aussi influencé par les connaissances antérieures possédées par l'individu.

Le processus de combinaison permet de transformer des connaissances explicites en d'autres connaissances explicites. Nous partageons normalement les connaissances explicites via des documents, courrier électronique, etc. Nous pourrions aussi partager ces connaissances explicites en ajoutant d'autres connaissances explicites qui les qualifient. Un exemple simple se trouve lors d'une classification d'un document ou l'ajout d'une métadonnée au document.

Le document lui-même est une connaissance explicite et la classification ou métadonnée est aussi une connaissance explicite. Cette dernière permet de retrouver facilement le document lors d'une recherche par un utilisateur.

<b>Tacite en tacite</b>	<b>Tacite en Explicite</b>
Socialisation	Extériorisation
<b>Explicite en tacite</b>	<b>Explicite en explicite</b>
Intériorisation	Combinaison

Tableau 1.1 : Les quatre processus de conversion de connaissances (Nonaka 1994, p.19)

<b>Tacite en tacite (socialisation)</b>	<b>Tacite en Explicite (Extériorisations)</b>
<ul style="list-style-type: none"> <li>• Collaboration synchrone (ex. chat)</li> <li>• <i>E-meetings</i></li> </ul>	<ul style="list-style-type: none"> <li>• Réponse aux questions</li> <li>• Annotation</li> </ul>
<b>Explicite en tacite (Intériorisation)</b>	<b>Explicite en explicite (Combinaison)</b>
<ul style="list-style-type: none"> <li>• Visualisation</li> <li>• Présentation audio ou vidéo</li> </ul>	<ul style="list-style-type: none"> <li>• Recherche textuelle</li> <li>• Catégorisation (classification) de document</li> </ul>

Tableau 1.2 : Des exemples de technologies pouvant faciliter la transformation de connaissance (Marwick 2001, p.816)

Nous expliquerons dans le chapitre 5, les activités qui facilitent ces processus de conversion de connaissance dans la recherche collaborative d'information.

### 1.3.3 Donnée, information et connaissance : quelle différence ?

Jusqu'ici, nous n'avons pas abordé le concept de donnée car ce concept n'est pas la priorité dans ce travail. Cependant, pour clarifier la différence entre ces trois concepts – donnée, information et connaissance – il nous semble important de présenter une définition du concept de donnée. Les données représentent des observations ou des faits hors du contexte et de ce fait n'ont pas de signification (Zack 1999, p.2 ; Wilson 2002). La connaissance est généralement distinguée de la donnée et de l'information. L'information est un résultat des



données mises en contexte souvent dans une forme de message. L'utilisation de ces informations dépend du contexte d'utilisation et de l'utilisateur. « En fait l'information a pour vocation d'être interprétée. Cette interprétation dépend de l'individu (son état, ses connaissances, etc.), qui intégrera, dans son interprétation, les contextes (de production et d'utilisation d'information), aussi bien que des données ou des faits » (David 2006, p.2). En résumé, l'information peut être définie comme l'ensemble des données ou des faits à partir desquels des interprétations peuvent être tirées.

Nous pouvons dire qu'une connaissance explicitée peut être considérée comme une information. Ceci veut dire qu'une connaissance formalisée et rendue explicite devient une information qui a pour but d'être interprétée par le récepteur. Donc, nous rejoignons Wilson (2002) pour dire que la connaissance existe et se construit dans l'esprit,<sup>27</sup> mais elle peut être influencée par l'interaction avec le monde extérieur ou les individus. « *La connaissance est définie comme ce que nous connaissons : elle implique le processus mental de compréhension, d'interprétation, et d'apprentissage qui se passe dans l'esprit et seulement dans l'esprit, mais néanmoins elle implique les interactions avec le monde extérieur à l'esprit ainsi les interactions avec les autres individus. Lorsque nous souhaitons exprimer ce que nous connaissons, nous le faisons à travers des messages d'une manière ou d'une autre – orale, écrite, graphique, gestuelle. Tel message ne transporte pas de connaissances mais constitue une information [...]* »<sup>28</sup> (Wilson 2002). La connaissance peut être également définie comme ce que nous croyons et valorisons (Nonaka 1994, p.16) fondée sur l'accumulation des informations (bien organisées) à travers des expériences, communication ou inférence.

Au cours de cette thèse, nous parlerons du partage de connaissances. Par le partage de connaissances dans ce contexte, nous entendons des connaissances exprimées et rendues explicites.

---

<sup>27</sup> Mind en anglais

<sup>28</sup> « *Knowledge is defined as what we know: knowledge involves the mental processes of comprehension, understanding and learning that go on in the mind and only in the mind, however much they involve interaction with the world outside the mind, and interaction with others. Whenever we wish to express what we know, we can only do so by uttering messages of one kind or another - oral, written, graphic, gestural or even through 'body language'. Such messages do not carry 'knowledge', they constitute 'information', which a knowing mind may assimilate, understand, comprehend and incorporate into its own knowledge structures. These structures are not identical for the person uttering the message and the receiver... Therefore, the knowledge built from the messages can never be exactly the same as the knowledge base from which the messages were uttered* » (Wilson 2002).

## **1.4 La communication**

Nous avons fait référence à la notion de communication depuis le tout début de ce chapitre sans vraiment élaborer ce concept. Nous essaierons dans cette section de développer ce concept qui est indispensable dans le transfert et le partage de connaissances lors d'une collaboration. Le mot communication est une forme nominale du verbe « communiquer ». Le mot communiquer du point de vue étymologique, apparu dans la langue française depuis le XIV<sup>e</sup> siècle, vient du latin « communicare » qui signifie « entrer en relation » et « être en relation, mettre en relation » et enfin « faire connaître » (Dictionnaire étymologie et historique de la langue française, 1996). Le mot communication apparu également dans la langue française depuis le XIV<sup>e</sup> siècle trouve son origine étymologique du latin « communicatio » qui signifie « mis en commun, échange de propos, action de faire-part » (Le Robert : Dictionnaire historique de la langue française, 1995). Parmi les définitions données dans le dictionnaire « Larousse », nous en sélectionnons quelques-unes pour expliquer ce concept qui est difficile à définir et qui varie en sens selon la discipline à laquelle il se rattache. La communication est définie comme « action, fait de communiquer, de transmettre quelque chose ». Elle est aussi définie comme « action de communiquer avec quelqu'un, d'être en rapport avec autrui, en général par le langage ; échange verbal entre un locuteur et un interlocuteur dont il sollicite une réponse ».

### **1.4.1 La communication vue sous quatre ordres de réalité**

Compte tenu de la multitude de sens du mot communication, nous sommes d'accord avec Bauthier (1994, p.5-7) et Breton & Proulx (2006, p.7-11) pour distinguer quatre ordres de réalité de la communication, à savoir : pratiques effectives de la communication, techniques que l'on met en œuvre dans ces pratiques, théories sur lesquelles s'appuient ces techniques et enfin les enjeux qui sont associés à la communication. Ces quatre ordres de réalité de la communication s'appliquent aussi bien dans la collaboration.

#### ***1.4.1.1 Les pratiques de communication***

Les pratiques de communication comportent les moyens de communication comme le geste, l'oralité, l'image, l'écriture que nous utilisons pour transmettre un message. Ces moyens se déploient à leur tour grâce à des multiples supports de communication comme le téléphone, le

livre, l'internet, le courrier électronique, l'espace physique commun, etc. (Breton et Proulx 2002, p.8).

Selon Robert Escarpit (1976, p.7), la communication peut être vue comme « un cas particulier du transport ». Ceci s'explique par le fait que les moyens et les supports de communication que nous utilisons permettent à un message d'être transporté vers l'autre (l'auditoire) (Breton et Proulx 2002, p.8). Le message dans ce contexte représente ce qu'on a à dire, à se dire ou à dire aux autres. « Il n'y a pas de communication s'il n'y a pas de transport d'un message » (Breton et Proulx 2002, p.35). Il est important de noter que le transport des messages, qui s'effectue grâce aux moyens et supports de communication, se distingue de la mise en forme des messages. Nous pouvons donc, nous poser des questions suivantes : quel est le sens porté par un message et comment est-ce que ce sens est exprimé ?

Selon Breton et Proulx (2002), la mise en forme de messages renvoie aux genres de la communication, à savoir : l'expressif, l'argumentatif et le descriptif. Lorsque notre but dans la communication est de convaincre, nos paroles s'étendent dans le genre argumentatif. Le genre expressif s'applique lorsque nous souhaitons exprimer un état vécu et ressenti ou « faire un récit marqué du sceau de la subjectivité et de l'imagination » (Breton et Proulx 2002, p. 9). Une communication expressive « s'appuie sur les techniques du récit et sur un certain nombre de figures pour donner accès au réel exprimé, comme par exemple la métaphore » (Breton 2006). Quant au genre informatif, ceci se déploie lorsqu'on « décrit le plus objectivement possible un fait, un événement, une opinion dont on a été témoin » (Breton et Proulx 2002, p. 40). Nous pouvons ainsi résumer qu'un message est à la fois mis en forme et transporté pour effectuer une communication.

Breton et Proulx, en expliquant « la mise en forme » de message, font référence aux théories néo-cybernétiques de la communication dont « la distinction entre « contenu » et « relation<sup>29</sup> » divise la communication en deux versants, complémentaires et indissociables, mais qui permettent de montrer qu'on insiste, en communiquant, plutôt sur la relation qui se noue avec le récepteur, tantôt sur le contenu de ce qu'on échange avec lui » (Breton et Proulx 2002, p.36). Nous rapprochons cette *relation* au concept de « contexte » dans la

---

<sup>29</sup> Il s'agit de la relation entre l'émetteur et le récepteur

communication. Ce concept sera présenté dans la section 1.4.1 portant sur la communication interpersonnelle.

#### **1.4.1.2 Techniques de communication**

La communication comporte la mise en œuvre de techniques. L'art d'argumenter est une technique indispensable quand il s'agit de la communication argumentative. Écrire est aussi une technique à apprendre pour savoir mettre des lettres de l'alphabet ensemble pour produire les sons du langage parlé. « Tous les aspects de la communication humaine sont pris dans l'univers des techniques » (Breton et Proulx 2002, p.9). Pour chaque genre de la communication, il y a tout un ensemble de techniques à apprendre pour pouvoir les déployer correctement. On constate aussi que les supports de communication ont connu beaucoup d'évolution, tout au long de l'histoire de l'humanité, grâce aux inventions techniques sophistiquées. L'usage des techniques est donc l'un des points fondamentaux qui permet de distinguer la communication humaine de la communication animale (Breton et Proulx 2002, p.9).

#### **1.4.1.3 Théories de la communication**

Plusieurs théories de la communication ont été développées depuis des années autour des pratiques et des techniques de la communication. Ces théories sont de deux ordres selon Breton et Proulx. Les premières sont centrées sur les processus de communication dans le but de les décrire et d'assurer leurs améliorations éventuelles. Dans ce cas, on parlera des *théories techniques* de la communication dont les réflexions portent sur la mise en œuvre des techniques de communication dans le but d'en accroître l'efficacité. Le modèle de Claude Shannon expliqué dans la section 1.2.2.1 s'inscrit dans cette catégorie. Le deuxième ordre de théories de la communication concerne les *théories sociales* de la communication dont les réflexions concernent l'objet « communication » dans toute sa dimension humaine. « Les *théories sociales* font appel aux champs différents des sciences humaines et sociales, qui tous rencontrent, à un moment ou à un autre, les pratiques de communication » (Breton et Proulx 2002, p.10). On constate que certaines des théories de la communication ont été inspirées par la cybernétique. Ces théories en revanche accordent à la communication une place centrale dans l'activité humaine.

#### 1.4.1.4 Enjeux associés à la communication

Breton et Proulx, en parcourant les évolutions qui ont vu le jour dans le domaine de la technologie de l'information et de la communication et en décrivant quelques-unes des grandes notions (la société de l'information, les réseaux, les communautés virtuelles, la mondialisation), ont traité le « paradigme informationnel ». Ce qu'ils entendent comme enjeux de la communication comporte le contenu utopique et idéologique de ce paradigme. Selon eux, « Nul ne doute qu'il y ait là de nouveaux outils, de nouveaux usages, [...] un élargissement de notre regard sur le monde et une augmentation de notre capacité de maîtrise du monde. Mais le problème est que ce nouveau paradigme « informationnel », depuis son émergence dans les années 1940, a d'une part une propension à « l'impérialisme épistémologique » au sein du champ de la communication et, d'autre part, est le siège d'une tentation idéologique à forte consonance utopique » (Breton et Proulx 2002, p.309). Ils remarquent que ce nouveau paradigme au sein de la communication met l'accent sur l'information et que l'idéologie de l'information tend à remplacer l'idéologie de la communication. Ils affirment, en plus, que la généalogie de ce paradigme pourrait être liée à la cybernétique et à la théorie de l'information.

Nous résumons ces quatre ordres de réalité de la communication, en reprenant le propos de Breton et Proulx (2002 p.11) : « le schéma qui s'impose est donc bien celui de la *parole humaine*<sup>30</sup>, mise en forme dans des genres distincts, qui se déploient à travers des moyens et des supports de communication sans cesse perfectionnés, dont la mise en œuvre, les usages et les conditions de réception sont théorisés, et dont le sens est pris dans des enjeux de société souvent majeurs ».

#### 1.4.2 La cybernétique

La cybernétique qui émerge dans les années 1940 grâce au mathématicien américain Norbert Wiener (1948), est un champ de savoir qui a révolutionné la communication, ses techniques et ses théories. La cybernétique présentée par son fondateur est une science « de contrôle et de communication pour l'homme, l'animal et la machine » Norbert Wiener (1948, p.11). Elle est centrée sur la modélisation de l'échange à travers l'étude de l'information et des principes d'interaction. Elle a vraiment joué un rôle important pour promouvoir le paradigme

---

<sup>30</sup> Message à communiquer

informationnel qui existe dans la communication. Par la science cybernétique, Norbert Wiener va introduire la notion de la rétroaction (feedback) dans le modèle de la communication. En revanche, certains auteurs ont démontré qu'il n'est pas l'initiateur de l'idée de feedback et que celle-ci existait depuis l'Antiquité, en tant que dispositif matériel, où la notion de feedback constituait « la partie informationnelle des horloges à eau, les clepsydres, inventées à Alexandre au IIIe siècle avant J.-C., le thermostat de Cornelius Drebbel, au XVIIe siècle, puis la régulation à boules de la machine à vapeur de James Watts au XVIIIe siècle » (Breton et Proulx 2002, p127-128). Cette idée de feedback rend le modèle de Wiener circulaire, contrairement au modèle de Shannon qui est linéaire. Ceci montre qu'émetteur et récepteur interagissent pendant un processus de communication.

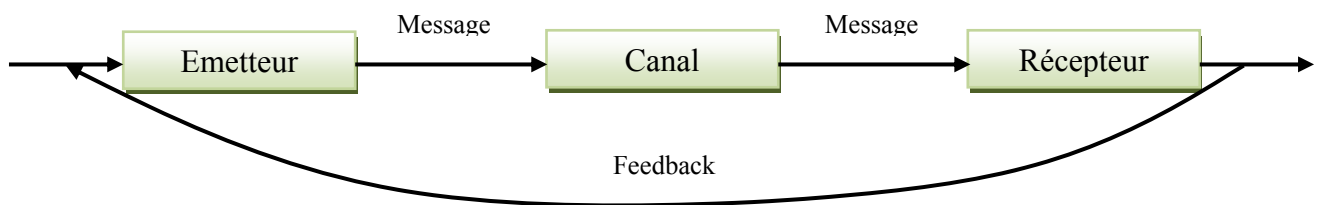


Figure 1.4 : Un simple schéma de modèle de rétroaction selon Norbert Wiener

En fait, ce modèle de Wiener fut développé pendant la Deuxième Guerre mondiale en travaillant sur le problème de la conduite de tir des canons anti-aériens (DCA). Il a tout de suite su que ce modèle, au centre de la cybernétique à l'origine, pourrait être appliqué dans plusieurs domaines du contrôle de machine (machine à penser) et qu'il est analogue au comportement de l'homme et de l'animal (Jones 2005, p.77). Prenons l'exemple, donné par Frederic Demarquet<sup>31</sup>, d'un tireur avec un canon qui vise une cible. Au premier tir, il rate la cible. Il va du coup corriger son tir pour faire mieux en estimant de combien il l'a raté. Au second tir, il se rapproche de la cible, mais la rate à nouveau. Il entre dans un processus à *causalité circulaire* par une régulation continue de son tir au moyen des informations dont il dispose (les feedbacks de ses actions) afin d'aboutir à son objectif. Ce modèle de feedback s'applique aussi dans le processus de recherche d'information là où un utilisateur formule une requête pour représenter son besoin informationnel, puis la soumet au système de recherche d'information. Le système lui fournit des résultats en réponse. En consultant les résultats, il

<sup>31</sup> [http://www.fredericdemarquet.com/feedback\\_communication.pdf](http://www.fredericdemarquet.com/feedback_communication.pdf), site consulté le 12 septembre, 2009.

trouve que son besoin informationnel n'est pas encore satisfait donc il reformule sa requête tout en intégrant ces observations sur le résultat de son premier essai. Il continue dans cet ajustement jusqu'à ce qu'il trouve des informations pertinentes.

La figure 1.5 explique mieux la notion de rétroaction selon Wiener. Elle montre la possibilité de formaliser la prise de décision dans des machines suffisamment autonomes pour observer, percevoir et analyser les informations provenant du monde extérieur et prendre en permanence des décisions afin d'atteindre un certain but fixé par avance. Cette prise de décision en permanence n'est réalisable que par le fait d'agir sur des informations reçues (observées). Cela ne concerne pas uniquement les machines, mais s'applique au comportement de tout être qui échange des informations (Breton et Proulx 2002, p.128). Nous pourrions dire également que les systèmes adaptatifs fonctionnent selon le principe cybernétique.

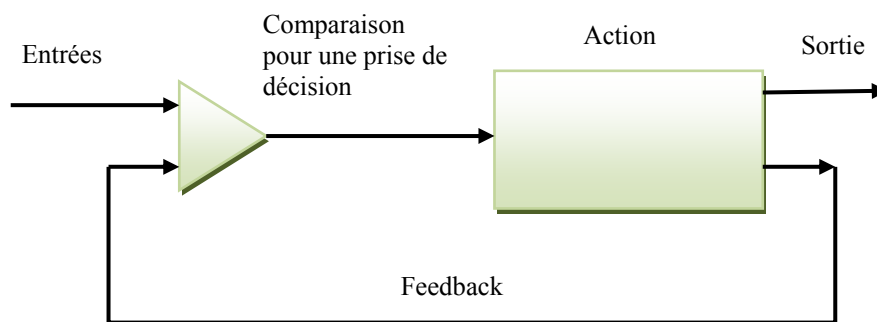


Figure 1.5 : Le modèle de rétroaction de Nobeert Wiener pour la prise de décision<sup>32</sup>

Selon Wiener, il existe deux types de feedback : (1) le feedback positif et (2) le feedback négatif.

Le feedback positif conduit à une accentuation d'un phénomène. Le feedback négatif, en revanche, conduit à une diminution d'un phénomène. Il tend donc vers la stabilité.

Nous pouvons également relier certains travaux de Gregory Bateson, anthropologue américain qui étudia de nombreuses tribus à travers le monde, à la cybernétique. Parmi ses travaux, il étudia les Thamuls et certains de leurs rites. Il se rendit compte que lors de certaines cérémonies, les hommes dansaient devant les femmes et, selon des rites précis, celles-ci réagissaient soit par des encouragements, soit par du désintérêt. Les hommes réagissaient

<sup>32</sup> Cette image est issue de <http://villemmin.gerard.free.fr/Scienmod/NotCyber.htm#site> (consulté le 7 Août, 2009).

alors à ces comportements et plus ils réagissaient, plus les femmes consolidaient leurs propres comportements jusqu'à arriver au niveau le plus intense du rite. Il se produisait bien alors une interaction entre les deux groupes qui visaient progressivement à se synchroniser par le moyen d'une communication circulaire et du feedback (Demarquet 2009, p.1). On trouve la notion de feedback positif et de feedback négatif dans ce récit. Lorsque les femmes réagissaient par du désintérêt, on pourrait dire que cette réaction est un feedback positif car elle provoque une accentuation chez les danseurs. En revanche, arrivant au niveau le plus intense du rite, les réactions des femmes deviennent un feedback négatif car elles ne provoquent plus une accentuation chez les danseurs mais plutôt tendent vers la stabilité.

### 1.4.3 La communication interpersonnelle

Il y a beaucoup de modèles issus de la théorie de la communication ainsi que de celle de l'information. Le modèle de Shannon est parmi les plus cités. On constate que dans la plupart de ces modèles, il y a trois invariants, figurant dans le modèle de Shannon – émetteur, message et récepteur. Toutes les variations apportées par chaque modèle se rapportent à ces trois invariants. Donc, en général, nous pouvons dire que la communication comporte l'échange de l'information (message) entre un émetteur et un récepteur. Elle peut être unilatérale quand il s'agit d'un échange en sens unique comme la radio, les affiches, etc. En revanche, elle peut être bilatérale quand les échanges sont à double sens.

Les recherches conduites à Palo Alto en Californie, puisant largement dans la matrice cybernétique, ont donné naissance à une conception nouvelle de la communication qui n'est plus considérée comme un schéma linéaire de l'émetteur vers le récepteur avec des relations univoques de cause à effet, mais comme une vision « orchestrale », systémique et circulaire des transactions effectuées (Josien 2004, p.92). Ces recherches, sur le plan pratique, montrent qu'en parlant de la communication, on devrait se focaliser moins sur le contenu que sur le contexte, car le contexte influence le sens du contenu (Breton et Proulx 2002) (Josien 2004, p.92). Yves Winkin a donc bien raison de dire que « c'est en terme de *niveaux de complexité*, de *contextes multiples* et de *systèmes circulaires* qu'il faut concevoir la recherche en communication » (Winkin 2000 p.25). Ceci est en concordance avec les points de vue des chercheurs de Palo Alto. Les travaux à Palo Alto ont donné naissance au concept de communication interpersonnelle ce à quoi nous nous sommes intéressés dans ce travail de thèse.



Selon Gouran et coll. (1994), il y a deux points de vue de la communication interpersonnelle – le point de vue contextuel et le point de vue relationnel. Du point de vue contextuel, la communication interpersonnelle implique un petit nombre de personnes (2 à 5), avec une proximité physique, qui communique à travers plusieurs canaux sensoriels en échangeant des feedbacks immédiats (Gouran et coll. 1994). Ce point de vue met l'accent sur le contexte dans lequel se déroule la communication, la proximité des communicants ainsi que leur nombre, mais ne prend pas en compte la relation pouvant exister entre eux. Du point de vue relationnel, la communication interpersonnelle est définie comme une communication entre les individus qui se connaissent (Gouran et coll. 1994). Ce point de vue met l'accent sur la relation existant entre les personnes. Gouran et coll. distinguent une communication entre un mari et sa femme d'une communication entre un client et un vendeur dans un magasin. Même si les deux types de communication se font entre deux personnes, ils ne sont pas tous les deux de la communication interpersonnelle selon le point de vue relationnel. Le premier type est vu comme une communication interpersonnelle car le mari et sa femme se connaissent et leur relation impacte la communication entre eux. Le deuxième est vu comme une interaction entre deux personnes dans une situation sociale : le client peut ne pas être en relation particulière avec le vendeur, il y a donc moins de possibilité d'influence de caractère relationnel sur la communication actuelle.

Dans le contexte de notre travail sur la collaboration, nous nous sommes intéressés à ces deux points de vue, mais plus particulièrement à la notion de contexte du point de vue contextuel. Nous préférons plutôt parler de la relation en tant que proximité et cognition sociale des communicants dans une communauté virtuelle ou dans un cadre traditionnel.

Dans leur ouvrage, Marc et Picard (2000) ont approfondi ce concept en distinguant la relation interpersonnelle de la communication interpersonnelle. Cela nous semble enlever la différence entre les deux points de vue de la communication interpersonnelle selon Gouran et coll. (1994), en les rassemblant. D'un point de vue terminologique, le mot « relation » se distingue du mot « communication ». D'une part, « la notion de relation désigne la forme et la nature du lien qui unit deux ou plusieurs personnes : on parle ainsi de relations familiales, professionnelles, amicales, de voisinage... » (Marc et Picard 2000, p.9). D'autre part, la communication est le rapport d'interaction qui s'établit lorsque les partenaires sont en présence (virtuel ou physique) (Marc et Picard 2000, p.9). En fait, la relation se constitue, se

développe et évolue à travers la communication. Il n'y a donc pas de relation sans communication. On pourrait dire que la communication représente la dimension dynamique du lien relationnel (Marc et Picard 2000, p.9).

#### ***1.4.3.1 La relation interpersonnelle vue sous trois dimensions***

Si la relation influe sur la communication entre les partenaires et que la communication enrichit ou appauvrit la relation, on pourrait dire que la relation et la communication s'influencent mutuellement. Cela veut dire que les facteurs affectant la relation inévitablement auront des effets sur la communication. Dans ce cas là, nous reprenons les trois niveaux pour appréhender toutes sortes de relations selon Marc et Picard – le contexte, la distance et le temps.

##### *Le contexte*

Le contexte ne peut pas être limité à un environnement dans lequel la relation ou la communication se déroule, mais il comporte plusieurs dimensions telles que le cadre, la situation et l'institution qui en effet constituent le contexte. Selon (Marc et Picard 2000, p.12, 13), le cadre est formé par les éléments physiques et temporels qui servent de « décor » à l'interaction. La deuxième dimension du contexte, qui est la *situation* d'interaction correspond au « scénario » qui définit et organise les relations (un cours, une partie de cartes, une réunion de travail, etc.). Enfin, la troisième dimension du contexte concerne l'institution dans laquelle s'inscrivent le cadre et la situation.

Prenons un exemple de deux amis dans la même promotion qui se réunissent dans la salle de cours pour travailler ensemble sur un devoir. Dans cet exemple, on trouve toutes les trois dimensions du contexte. D'abord, les deux étudiants sont des amis, cela veut dire qu'il y a une relation entre eux. Ils se réunissent dans une salle donc la salle représente le cadre. La situation d'interaction qui les réunit est une réunion de travail sur un devoir. L'institution dans laquelle s'inscrivent leurs interactions est l'institution scolaire. Un même cadre peut autoriser plusieurs situations, par exemple la salle, qui dans cet exemple dépeint une situation de travail sur un devoir, elle peut être utilisée également pour un cours. Ainsi l'on peut considérer que la situation détermine l'enjeu de la rencontre.

### *La distance*

La « proximité » ou l'« éloignement », qui se traduit en *distance*, entre les partenaires a une influence sur la relation ainsi que la communication entre eux. La distance, selon Marc et Picard (2000, p.14), a une double signification : la distance physique entre les interlocuteurs (l'espace physique qui les sépare) et la distance « psychologique » se rapportant au degré de familiarité qui les réunit.

Une simple illustration d'influence qu'une distance physique peut avoir sur une relation ainsi que sur la communication au sein de cette relation se trouve dans le cadre de deux conjoints qui vivent séparés à cause de leurs professions. Marc et Picard (2000, p.14) expliquent que ces conjoints expriment leurs difficultés à traduire leur tendresse avec la seule aide des mots écrits (une lettre) ou d'une conversation téléphonique sans l'appui de regards, de gestes et de silences que l'on partage côte à côte. En ce qui concerne la distance « psychologique », Marc et Picard (2000, p.15-21) soulignent deux ordres : une polarité « familier/inconnu » et le degré de « convergence » ou de « divergence ». Ils décrivent l'ordre « familier » comme celui qui partage notre vie quotidienne et l'« inconnu » comme celui qu'on rencontre à l'improviste ou que l'on côtoie de façon fugitive, sans rien savoir de lui. Ils expliquent que la distance en tant que « convergence » ou « divergence » entre les individus se forme à différents niveaux et traduit des réalités différentes qui, en revanche, ne jouent pas de la même façon sur l'évolution d'une relation interpersonnelle. En général, la convergence peut améliorer et fortifier une relation tandis que la divergence peut nuire la relation.

Nous remarquons qu'il y a un autre type de distance qui se manifeste surtout dans le cadre d'une collaboration. Il s'agit de la distance cognitive. Dans une interaction entre deux personnes qui collaborent, la distance cognitive qui existe entre eux vis-à-vis d'un problème à résoudre influe sur la communication. Pour expliquer un concept ou une technique, l'un des collaborateurs peut être obligé d'illustrer à l'aide d'un schéma ou d'images pour que son collaborateur puisse comprendre ce qu'il voulait dire.

### *Le temps*

Le troisième niveau d'appréhension d'une relation s'inscrit dans la « temporalité ». Selon Marc et Picard (2000, p.21-27), la relation interpersonnelle s'inscrit dans une triple temporalité :

- Temporalité propre à chaque sujet : cette dimension de la temporalité porte sur le vécu relationnel d'un individu dans le passé. On constate qu'à chaque moment de notre existence, nos expériences relationnelles, heureuses ou malheureuses, ont toujours une grande influence sur nos relations ultérieures (Marc et Picard 2000, p.21). Prenons pour l'exemple le cas de deux collègues qui se sont disputés récemment ou qui ont toujours eu des désaccords. Si les deux se retrouvent dans un groupe de travail venant d'être créé dans leur lieu de travail, évidemment leur passé influencera la relation et la communication entre eux au sein du groupe. Cette dimension de temporalité renvoie à la notion de transfert du passé dans le présent ce qui implique que l'ancienneté d'une expérience a un impact sur le présent.
- Temporalité de la relation elle-même : selon Marc et Picard (2000, p.14), chaque relation, quelle que soit sa nature, a un commencement, un déroulement et une fin. Dans notre exemple, le groupe de travail, qui vient d'être constitué, dispose d'un temps pour commencer son travail et d'un temps pour le finir. Cette temporalité peut modifier les interactions entre les deux collègues. Ils peuvent essayer de se supporter afin de ne pas nuire au fonctionnement du groupe. Ceci veut dire que la durée de l'interaction peut influencer sur la communication ainsi que sur la collaboration.
- Temporalité propre à la rencontre : dans chaque relation comme celui du groupe de travail, chaque rencontre entre les partenaires s'inscrit dans une autre dimension de temporalité - la façon dont la rencontre s'inscrit dans une structuration de la temporalité pour chaque interactant (Marc et Picard 2000, p.15).

### 1.5 A propos du concept de collaboration

Notre but dès le début de ce chapitre est d'arriver au fur et à mesure à la notion de collaboration tout en expliquant les concepts – information, communication et connaissance – qui sont rassemblés dans le concept de collaboration. La collaboration est un mot en vogue aujourd'hui en vue de toutes sortes de technologies qui sortent chaque jour. Ainsi, on entend des expressions comme travail collaboratif, apprentissage collaboratif, formation collaborative, culture collaborative, web collaboratif, intelligence collaborative<sup>33</sup>, réseaux sociaux, réseaux professionnels, intelligence collective etc. Ces expressions sont utilisées sans vraiment savoir ce qu'elles représentent en profondeur.

---

<sup>33</sup> Collaborative intelligence en anglais

Il nous semble important, en premier lieu, d'aborder le terme collaboration avant d'entreprendre les dérivées qui en découlent. Étymologiquement, le mot collaboration est emprunté du latin médiéval *collaboratio*. Ce mot est dérivé de *collaborare* au sens tardif de « travailler en commun pour gagner des bénéfices » (Le Robert : Dictionnaire historique de la langue française, 1995).

### 1.5.1 Collaborer ou coopérer ou coordonner ?

Les trois mots – collaboration, coopération et coordination – sont utilisés assez souvent pour dépeindre la même chose. Dans le dictionnaire on ne trouve pas vraiment une différence entre la collaboration et la coopération. En revanche le mot coordination dans le dictionnaire porte un sens différent. Selon le dictionnaire Larousse, la coordination est définie comme « l'action de coordonner ; harmonisation d'activités diverses dans un souci d'efficacité ». Le verbe « coordonner » est défini comme « ordonner des éléments séparés, combiner des actions, des activités distinctes en vue de constituer un ensemble cohérent ou d'atteindre un résultat déterminé ». Cependant, on y trouve une définition comme « action de coopérer, de participer à une œuvre commune » pour le mot coopération. Ces définitions montrent que la coordination est différente de la collaboration et de la coopération. En revanche elles ne donnent aucune indication d'une différence entre la collaboration et la coopération – deux mots qui sont utilisés comme des synonymes dans le dictionnaire.

Plusieurs auteurs ont alors tenté de différencier ces termes qui se manifestent tous les jours dans notre vie. Notre système physiologique est un bon exemple de manifestation de coopération et de coordination. Tous nos organes fonctionnent pour que nous restions vivants et en santé. Cela veut dire qu'ils coopèrent et qu'ils échangent des informations entre eux. Mais dans cette coopération, il y'a une coordination des activités de chaque organe afin que la coopération puisse aboutir à un but global de notre système physiologique. Chaque organe a son propre but qui est différent de celui d'un autre organe, mais tous ces buts sont orientés vers le but global. Dans notre lieu de travail, on trouve un système de coordination qui s'opère pour que l'entreprise réussisse sa mission pour aboutir aux buts fixés. Dans l'entreprise il y a une variété de hiérarchies pour bien gérer le fonctionnement de l'entreprise. Le but de cette hiérarchisation est de bien coordonner les activités de chaque employé dans l'entreprise. On

trouve des situations dans lesquelles une entreprise coopère avec une autre entreprise : ils échangent des informations, partagent des ressources sans avoir forcément le même objectif.

Selon la définition du terme *collaboration* donnée ci-dessus, une des différences majeures entre la collaboration et la coopération est que dans la collaboration, les participants partagent le même objectif et travaillent ensemble pour aboutir à ce but. En revanche, dans la coopération, il n'y a pas forcément de partage d'objectif entre les partenaires, et pas de mutualisation dans le travail. Le résultat final dans la coopération est une sommation des travaux de chaque participant. Chacun contribue pour sa part pour atteindre le but final. Chaque coopérant peut également identifier sa contribution dans la globalité. Par exemple, la construction d'un bâtiment (ou d'une maison) est un travail coopératif. L'architecte donne les plans qui illustrent le but ou le résultat final attendu. Les ingénieurs civils, les électriciens, les maçons, les menuisiers, les chauffagistes, etc. travaillent ensemble pour achever le projet de l'architecte. Chaque intervenant dispose de règles qui gouvernent le travail qu'il doit respecter. Dans le même temps, pour que la coopération aboutisse, il existe des règles de gouvernance pour contrôler les interactions entre chaque participant. Le but final est une sommation de travail de chaque intervenant. En revanche, dans la collaboration, les participants s'impliquent dans un processus de différenciation et d'intégration de leurs expertises individuelles afin d'arriver à une solution commune qui n'est pas prédéfinie, mais qui est co-construite par l'ensemble.

Nous trouvons que la coordination est un processus qui se présente dans la coopération ainsi que dans la collaboration. Certains auteurs (Winer et Ray 1994, p.22) ont tenté de différencier les trois concepts en les considérant chacun individuellement mais il nous semble important de souligner que la coordination ne se trouve pas au même niveau avec la collaboration et la coopération, car celle-ci (coordination) implique la **gestion** des interactions et des interdépendances entre des parties. Les interactions et les interdépendances existent autant dans la collaboration que dans la coopération. Nos exemples ci-dessus montrent comment la coordination s'implique au sein des activités ou travaux coopératifs. En ce qui concerne la coordination au sein d'une activité de collaboration, nous prenons l'exemple du *brainstorming* là où différents participants travaillent ensemble pour trouver une ou des solution(s) à un problème. Si tous les participants parlent en même temps, le but ne sera pas atteint, car une situation de chaos s'installera. Pour éviter cela, il faut une coordination de la prise de parole.

Ce concept de coordination est très important pour réussir une collaboration ou une coopération, nous l'élaborons dans le quatrième chapitre en présentant le modèle 3CM pour la compréhension de la recherche collaborative d'informations. Nous nous contentons donc d'expliquer dans ce chapitre la différence entre la collaboration et la coopération.

Une autre différence entre la collaboration et la coopération se manifeste au niveau de la répartition des tâches parmi les membres du groupe. Dans la coopération, les tâches sont réparties entre les partenaires, chacun exécute la tâche qui lui a été attribuée et ainsi contribue au résultat final. On trouve néanmoins, dans la collaboration, des situations qui nécessitent que les collaborateurs prennent des rôles différents tout en travaillant ensemble. Dans ce cas-là, il y a une répartition spontanée des tâches qui s'applique temporairement. Par exemple dans la recherche collaborative d'informations, un partenaire consulte les résultats issus d'une requête et un autre réfléchit sur la reformulation de la requête ou sur le besoin d'utiliser un autre système d'information. Après une suggestion pour essayer un autre système de recherche d'information, les deux partenaires continuent à travailler ensemble. Il n'y a pas de rôle fixe dans ce type de division du travail. Cette façon de répartir les tâches est désignée par Dillenbourg comme « la division horizontale du travail ». En revanche, dans le cadre de la coopération, les tâches sont généralement définies dès le début de la coopération. Les sous-tâches ainsi que les rôles sont bien distincts et fixés (Dillenbourg 1999, p.8). Chacun se focalise sur sa tâche et les résultats sont fusionnés à la fin. Dillenbourg a nommé celle-ci « la division verticale du travail ».

La collaboration se distingue aussi de la coopération au niveau de l'interactivité entre les partenaires. Intuitivement, on dirait que la collaboration doit être assez interactive (Dillenbourg 1999, p.8). Une explication de l'interactivité dans la collaboration qui la différencie de celle de la coopération a été donnée par Dillenbourg. Selon lui, le degré de l'interactivité parmi les partenaires n'est pas défini par la fréquence de leurs interactions mais plutôt par la mesure dans laquelle ces interactions influent sur les processus cognitifs de leurs pairs. Nous reprenons son exemple de deux architectes devant évaluer deux maisons, pour illustrer l'imbrication entre le raisonnement et l'interaction. Si les deux architectes coopèrent, chacun évaluera une maison et ils combineront leur résultat à la fin. Lors de l'évaluation, ils mènent chacun un processus indépendamment de l'autre : l'un peut compter l'espace de la maison en allant de chambre en chambre, l'autre peut se fonder sur la dimension générale de

la maison (en tant que cube). Ils interagissent pour rendre cohérentes leurs évaluations ; mais ces interactions, qui ne se font qu'après une production individuelle d'une solution partielle, ne peuvent que les influencer s'ils acceptent de refaire l'évaluation. Au contraire si les deux architectes entament un processus de collaboration dès le début de l'évaluation, ils entrent dans la première maison et interagissent ensemble sur la façon de procéder dans l'évaluation. L'un dit à l'autre « on compte à partir du sous-sol » et s'ils se mettent d'accord en justifiant pourquoi aller dans cette direction, ils produiront à la fin une évaluation plus valable. On trouve que la collaboration implique plus d'interactions et que celles-ci influent sur le raisonnement (processus cognitif) des partenaires.

	<b>Coopération</b>	<b>Collaboration</b>
<b>Objectif</b>	Objectif plus ou moins partagé	Objectif fortement partagé
<b>Degré d'interdépendance</b>	Moyenne	Haut
<b>Mutualisation du travail</b>	Plus ou moins présente	Haute
<b>Engagement</b>	Moins que dans la collaboration	Beaucoup d'engagement de la part des participants
<b>Interaction</b>	Moyenne	Nécessite beaucoup d'interactions entre les participants
<b>Répartition des tâches</b>	Verticale	Horizontale

Tableau 1.3 : Coopérer ou collaborer ?

Une autre dimension associée à l'interaction se traduit par le synchronisme de l'interaction. On dirait que la collaboration, intuitivement, comporte une communication synchrone mais que la coopération est souvent associée à la communication asynchrone.

Un résumé de la comparaison entre la coopération et la collaboration est montré dans le tableau 1.3. ci-dessus. Cette dichotomie peut être critiquée sur le fait que la frontière entre ces deux termes est vague. Même si cela peut paraître vrai, il est néanmoins important de savoir que la dichotomie illustre le niveau de complexité associé à chaque concept.



### 1.5.2 Les dimensions de la collaboration

Nous abordons les différentes dimensions de la collaboration en suivant le modèle de Golovchinsky et coll (2008, p.1-3). Ce modèle facilite la compréhension des technologies et des systèmes de collaboration. En fait, il permet, d'une part, de classifier les systèmes de collaboration existants et, d'autre part, d'identifier et de suggérer d'autres conceptions possibles. Ce modèle comporte quatre dimensions : l'intention (*intent*), la profondeur (*depth*), la simultanéité ou le synchronisme (*concurrency*) et la localisation (*location*).

*L'intention* : la dimension de l'intention dans les systèmes de collaboration renvoie au fait que la collaboration soit explicite ou implicite. La collaboration explicite implique le travail en commun dans un groupe d'individus ayant un objectif partagé, ces individus se connaissant virtuellement ou physiquement. L'accent est mis sur un échange explicite de l'information entre les participants et un partage explicite de leur objectif. En revanche, la collaboration implicite n'est pas vraiment une collaboration mais pour le simple but de classification, nous nous permettons d'utiliser ce terme pour désigner toutes sortes de systèmes qui permettent d'utiliser les traces d'un groupe d'utilisateurs, ou les informations concernant leurs interactions autour d'un objet pour suggérer des choix ou des objets aux autres utilisateurs. Ce type de fonctionnalité se disant collaborative s'applique en particulier dans les systèmes de filtrage et de recommandation collaborative<sup>34</sup>. Dans la collaboration implicite, un utilisateur ne connaît pas les utilisateurs dont leurs activités (traces) sont utilisées pour lui suggérer des choix. Il ne connaît pas, non plus, leurs objectifs (leurs intentions) lors de ces activités. En fait, la collaboration implicite n'est qu'une forme de coordination.

*La profondeur* : cette dimension s'applique au niveau technique du fonctionnement du système de collaboration. Elle renvoie au degré de la médiation de l'ordinateur dans la collaboration. On trouve deux cas possibles : une médiation via l'interface plutôt contrôlée par les participants ou une médiation très algorithmique. Par exemple, dans certains systèmes de collaboration, la conscience de groupe (*awareness*) et la répartition des tâches sont faites

---

<sup>34</sup> « Les systèmes de recommandation sont une forme spécifique de filtrage de l'information (SI) visant à présenter les éléments d'information (films, musique, livres, news, images, pages Web, etc) qui sont susceptibles d'intéresser l'utilisateur. Généralement, un système de recommandation permet de comparer le profil d'un utilisateur à certaines caractéristiques de référence, et cherche à prédire l'« avis » que donnerait un utilisateur » (Wikipedia, consulté le 23 Juillet, 2010).

par les participants qui communiquent par le système de chat incorporé. Il existe aussi d'autres systèmes dans lesquels ces deux dernières sont faites automatiquement par le système.

*Le synchronisme* : cette dimension se réfère à la simultanéité de l'échange entre les collaborateurs. Ainsi, la collaboration peut être de manière synchrone ou asynchrone. La collaboration synchrone implique l'échange d'information parmi les collaborateurs en même temps, par exemple un appel téléphonique, une réunion en tête à tête, la messagerie instantanée, etc. La collaboration asynchrone, en revanche, implique l'échange d'informations par lequel les collaborateurs reçoivent et envoient l'information en temps différé, par exemple le courrier électronique, le partage d'annotations électroniques, etc. Il est important de souligner qu'aujourd'hui, avec l'avancement de la technologie, on trouve de plus en plus de systèmes de collaboration synchrone.

*La localisation* : cette dimension renvoie à la localisation des collaborateurs. Les collaborateurs peuvent se trouver au même endroit ou à des endroits différents. Ainsi un système de collaboration peut être conçu pour faciliter une collaboration entre des collaborateurs se trouvant à un seul endroit ou une collaboration entre des collaborateurs dispersés géographiquement.

Ces quatre dimensions peuvent aider dans la méthodologie de conception et du développement d'un système de collaboration.

### **1.5.3 Les facteurs sociocognitifs associés à la collaboration**

Il y a plusieurs facteurs qui influent sur la collaboration au sein d'un groupe. Un groupe est constitué de deux ou plusieurs personnes - une dyade (2 personnes), un petit groupe (3-5 personnes), un groupe moyen (10 à 30 personnes), une communauté (des centaines de personnes), une société (des milliers ou des millions de personnes) (Dillenbourg 1999, p.1). Ces facteurs portent sur les aspects sociaux, cognitifs et psychologiques. Pour appréhender le concept de collaboration, il faudrait se référer aux travaux en psychologie cognitive et en psychologie sociale. En fait, ces deux domaines sont fondamentaux pour la compréhension de la résolution collaborative de problème ainsi que pour la conception de systèmes qui facilitent la collaboration (Deek et McHugh 2003, p.7-26).

Parmi les travaux qui ont traité ces facteurs s'inscrit le travail de McGrath (1984) sur l'interaction et la performance du groupe, dans lequel il a souligné quatre ensembles de propriétés influant sur le processus d'interaction de groupe. Le premier s'applique au niveau individuel des membres du groupe. Ceci comporte la propriété biologique, sociale, psychologique de chacun des membres du groupe. Le deuxième s'applique au niveau interpersonnel parmi les membres du groupe. La troisième propriété concerne l'environnement où se déroule l'interaction. Enfin, la quatrième propriété concerne la tâche à effectuer par le groupe.

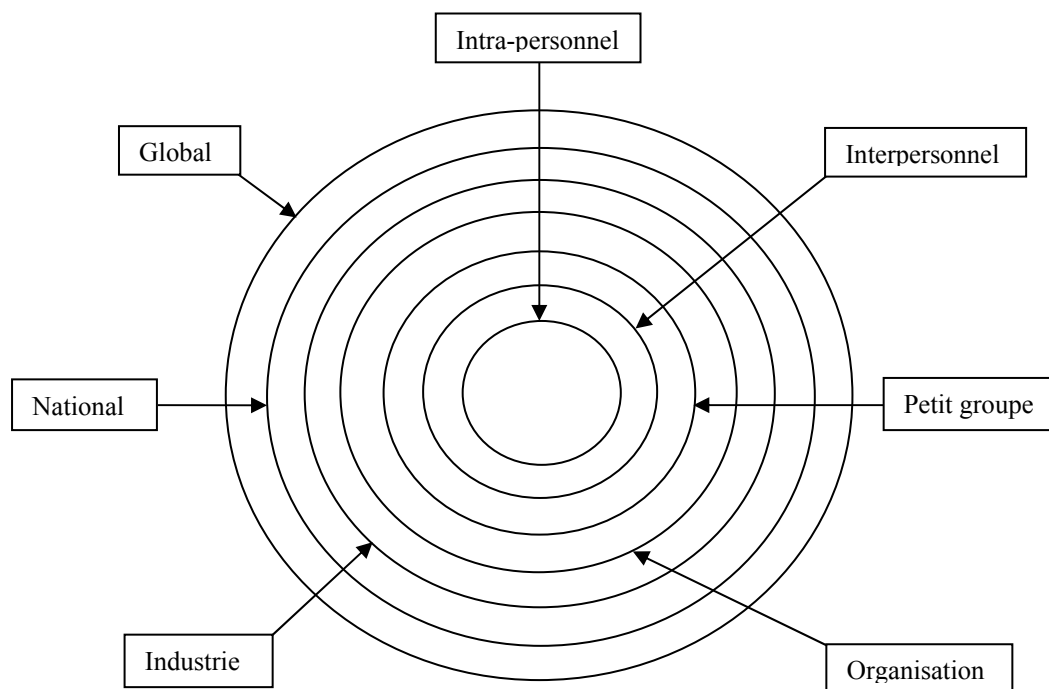
Quand deux ou plusieurs personnes collaborent, il y a un ensemble de processus qui s'effectue. Selon McGrath (1984, p.17-19), ces processus s'effectuent en trois étapes :

- La première étape est le processus de communication. Un échange entre un membre du groupe vers un autre se fait par le biais de la communication quelle que soit sa forme. La forme de communication comportant : le canal, les modalités, le temps, renvoie à un modèle (*pattern*) de communication.
- La deuxième étape est une combinaison de deux processus : le processus d'action (*action process*) et le processus d'attraction (*attraction process*). Chaque comportement interactif comporte deux composants : la tâche (processus d'action) et le composant interpersonnel (processus d'attraction). La tâche renvoie aux actions des participants qui se traduisent en le modèle de performance de tâche (*task performance pattern*). Le composant interpersonnel renvoie à la relation interpersonnelle parmi les membres du groupe.
- La troisième étape est le processus d'influence. Ceci renvoie à l'impact des trois modèles (modèle de communication, modèle de performance de tâche et modèle de relation interpersonnelle) sur les participants. Il comporte les conséquences de l'interaction sur les participants ainsi que sur leurs relations, leurs performances de tâche et leurs communications ultérieures (McGrath 1984, p.17).

Il est important de noter que ces processus, qui sont à la fois cognitifs et sociaux, impliquent un partage de « l'attention » du groupe (Deek et McHugh 2003, p.9). Cela veut dire qu'il y a plusieurs processus cognitifs qui se passent en même temps pendant une collaboration. En revanche les ressources cognitives, par exemple « l'attention », sont rares (Deek et McHugh 2003, p.9). La limite naturelle dans les exercices d'activités cognitives explique le problème de « surcharge cognitive » pouvant survenir lors d'une collaboration. Cette limite humaine explique aussi le fait que notre engagement dans une activité cognitive limite notre capacité de nous engager simultanément dans plusieurs activités cognitives (Deek et McHugh 2003, p.10) ; par exemple, l'écoute attentive entrave la capacité de traiter de façon critique de nouvelles informations (Dennis 1996 à partir de Deek et McHugh 2003 p.10). En vue de ce problème, Nunamaker et coll. (1996) ont proposé un modèle simple pour « l'attention du groupe ». Ce modèle a été repris par (Deek et McHugh 2003, p.74-75) pour montrer la façon de structurer les processus du groupe afin d'améliorer sa productivité.

### 1.5.3.1 Modèle de Freshman

Brenda Freshman (2004, p.345-368) sous l'optique de la psychologie organisationnelle et pour souligner les complexités associées aux « forces motrices<sup>35</sup> » et aux « forces de retenues<sup>36</sup> » de la collaboration interdisciplinaire, a proposé un modèle de parties prenantes d'un système (*system stakeholder model*).



<sup>35</sup> Driving force

<sup>36</sup> Restraining force

Figure 1.6 : Le modèle de parties prenantes d'un système (Brenda Freshman 2009, p.352)

Ce modèle, comportant sept niveaux, peut permettre d'identifier les pratiques, les compétences et les ressources qui facilitent la collaboration à chaque niveau. Chaque niveau intérieur dans ce modèle peut être influencé par les forces dynamiques des niveaux qui l'entourent. L'inverse est aussi valable, car les effets des forces influençant sur un niveau intérieur portent aussi bien sur les niveaux qui l'entourent. Ceci correspond à l'observation aphoristique de (Allport 1924) – « Il n'y a pas de psychologie de groupe qui n'est pas essentiellement une psychologie d'individu ». Voilà pourquoi ce modèle de Freshman est plus qu'une classification de groupe selon la taille, car le noyau de son modèle est le niveau intra-personnel. Cette façon d'étudier et d'analyser le comportement du groupe à partir du comportement individuel s'applique également dans le domaine de la recherche d'informations dont les travaux de recherche depuis une cinquantaine d'années portent sur la pratique informationnelle individuelle. Aujourd'hui on se dirige vers la pratique informationnelle collaborative. Dans le cadre de notre travail, nous ne nous intéressons qu'aux trois premiers éléments du modèle : intra-personnel, interpersonnel, un petit groupe. Donc, nous présentons dans la figure 1.7 et 1.8 les forces retenues et les forces motrices influant sur la collaboration dans la recherche d'information.

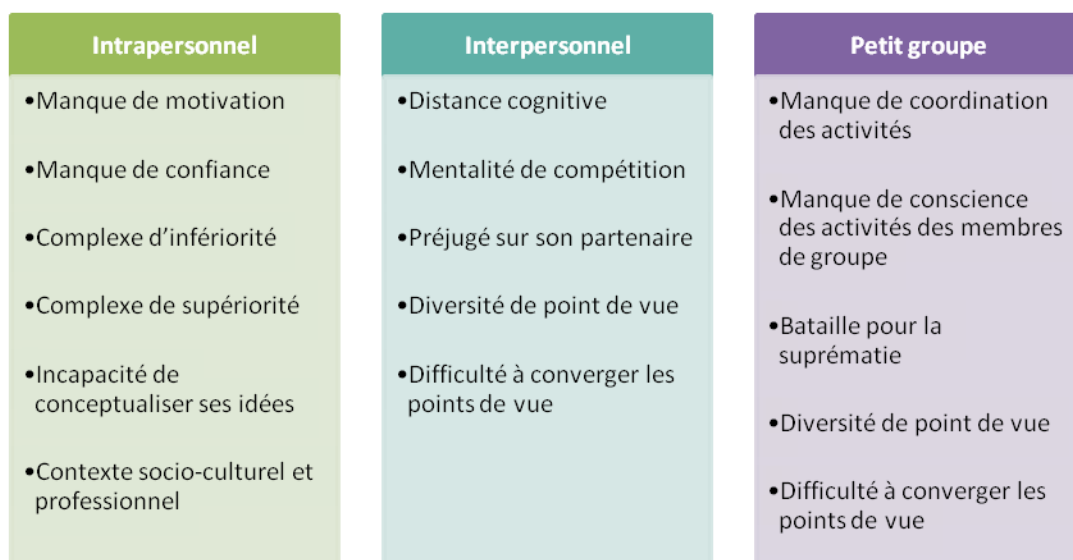


Figure 1.7 : Les « forces retenues » influant sur la collaboration

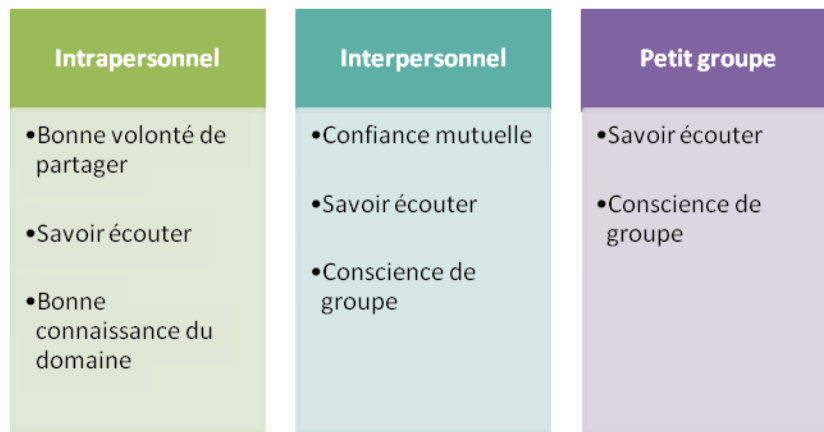


Figure 1.8 : Les « forces motrices » influant sur la collaboration

#### 1.5.4 Quelques concepts associés à la collaboration

Nous expliquons ci-après quelques notions collaboratives qui sont en vogue aujourd'hui. Notre but ici n'est pas de justifier l'adéquation de ces termes, mais plutôt de présenter les sens.

##### *Le travail collaboratif*

Aujourd'hui, le « travail collaboratif » est l'une des expressions les plus utilisées lorsqu'on parle de la collaboration assistée par la technologie. Il renvoie à une pratique de travail en commun, permettant d'échanger et de partager des compétences pour mieux réussir un projet en collectif, à travers un environnement informatisé (ou en ligne). Le collecticiel<sup>37</sup> s'inscrit dans cette dimension.

##### *L'apprentissage collaboratif*

L'apprentissage collaboratif renvoie à la co-construction des connaissances par un groupe d'individus. Selon Depover et Marchand (2002, p.85), il prend la forme de discussions dans les webconférences, les forums de discussions ou d'échanges privés par courrier électronique.

##### *L'édition collaborative*

L'édition collaborative se dit d'une pratique du groupe d'individus travaillant ensemble, via des contributions individuelles, pour produire un document. Cette pratique se manifeste souvent dans la coédition de sources code et dans la production de documents textes. Un

<sup>37</sup> groupware

exemple très répandu est le Wikipedia<sup>38</sup>. Ce type de collaboration est souvent asynchrone, mais aujourd'hui on trouve de nouveaux outils d'édition collaborative qui sont synchrones, par exemple Collabedit<sup>39</sup>.

### ***L'intelligence collaborative***

Selon Mark Elliot de MetaCollab, l'intelligence collaborative (aussi appelé le quotient d'intelligence collaborative) est vue comme une mesure de la capacité collaborative d'un groupe ou d'une entité. Elle implique que la capacité d'un groupe à résoudre un problème dépasse largement la connaissance individuelle d'un membre du groupe.

### ***Le web collaboratif***

Le web collaboratif regroupe toutes les technologies du web (surtout web 2.0) qui permettent de relier les gens et de faciliter des interactions entre eux.

### ***L'intelligence collective***

L'intelligence collective renvoie à une intelligence partagée ou à l'intelligence du groupe ou à l'intelligence des connexions et des relations qui résultent d'une coopération ou d'une compétition de plusieurs individus (Zara 2004, p.8). Elle représente les capacités cognitives d'une communauté résultant des interactions multiples entre les membres (ou agents)<sup>40</sup>.

### ***Les réseaux sociaux***

Un réseau social est une structure sociale constituant un ensemble d'entités sociales telles que des individus ou des organisations reliées entre elles par des liens (ou des relations) telles que l'amitié, les relations familiales, les goûts, les croyances, les professions, les relations affectives, etc. Le réseautage social est une structure très ancienne, qui aujourd'hui grâce à l'internet a vu une explosion et une exploitation énormes. On trouve à présent plusieurs applications Internet (services de réseautage social) favorisant la création de réseaux d'amis, aidant à trouver un emploi, des partenaires commerciaux, des amants etc. Dans l'analyse des réseaux sociaux, les relations sociales sont analysées par la théorie des réseaux comportant des graphes de relation entre des nœuds. Les sommets (nœuds) désignent, en général, les individus ou des organisations constituant le réseau et les arcs (arêtes) sont les relations entre les nœuds. Le réseautage social nous permet d'élargir ou agrandir notre propre réseau en

---

<sup>38</sup> [www.wikipedia.com](http://www.wikipedia.com)

<sup>39</sup> <http://collabedit.com>

<sup>40</sup> Wikipedia, consulté le 30 septembre, 2009

créant des réseaux sociaux (réseaux de réseaux). Les services ou systèmes de réseautage sociaux ne sont pas des systèmes de collaboration en soi, mais sont plutôt des structures permettant de retrouver des collaborateurs potentiels.

## **1.6 Conclusion du premier chapitre**

Nous avons tenté de démontrer dans ce chapitre le fait que la collaboration n'est pas une activité que l'on peut confier à une machine ou un système quelconque. Bien sûr, certaines fonctionnalités peuvent être automatisées. Mais globalement, la collaboration est un processus communicationnel à la fois social et cognitif qui pourrait être facilité par un système ou un dispositif technologique. Il est donc possible d'affirmer que le succès d'une collaboration dépend, à la fois de la culture d'ouverture des participants et des technologies de collaboration employées.

Comme le montre le sujet de cette thèse, nous avons comme domaine d'application l'intelligence économique. Notre but est de proposer une approche générique pour la gestion de la recherche collaborative d'information qui peut être utilisée dans n'importe quel domaine. Néanmoins, il est important de présenter ce domaine d'application qui est aussi une préoccupation majeure de l'équipe SITE au sein de laquelle ce travail de thèse a été effectué. Ainsi, le deuxième chapitre portera sur l'intelligence économique.





# Chapitre 2

---

## L'intelligence économique : domaine d'application

### 2.1 Introduction

Comme le titre de cette thèse l'indique, nous nous sommes impliqués dans le cadre du processus de l'intelligence économique (IE). Mais avant de décrire le processus de l'IE et le niveau d'implication de la recherche collaborative d'information (RCI), il apparaît indispensable de comprendre ce concept « IE » qui gagne du terrain aujourd'hui dans le monde de l'entreprise. Il faut savoir que des pratiques que l'on peut qualifier de démarches de l'IE existent depuis fort longtemps contrairement aux idées couramment répandues. L'IE a toujours existé en tant que pratique mais elle est devenue récemment un domaine de recherche voire une discipline d'enseignement dans certaines filières de formation. D'une manière générale, on peut considérer l'IE comme étant *la capacité à comprendre notre environnement et à anticiper le changement* (Levet 2001, p.vii). Comment peut-on comprendre son environnement, autrement dit comment maîtriser les informations afférentes ? Comment peut-on extraire des connaissances qui mènent à l'action à partir des informations dont on dispose ? Ces deux questions sont au cœur de l'IE.

Comprendre son environnement consiste à rechercher et à collecter des informations pertinentes, à les interpréter de façon à pouvoir identifier les opportunités et les menaces afin de pouvoir décider ou agir. En revanche, anticiper le changement consiste à identifier et étudier les signaux faibles<sup>41</sup> (*weak signals*) (Ansoff 1975) qui sont les ruptures susceptibles de se produire dans l'environnement (économique, social, technologique, culturel, etc.) (Lesca et Castagnos 2000, p.3). Cette écoute anticipative de l'environnement constitue le concept que l'on appelle la veille.

---

<sup>41</sup> En intelligence économique, les signaux faibles sont les éléments de perception de l'environnement, opportunités ou menaces, qui doivent faire l'objet d'une écoute anticipative, appelée veille, dans le but de participer à l'élaboration de choix prospectifs en vue d'établir une stratégie, et de réduire l'incertitude. (Wikipédia consulté le 10 Mai 2010)

Dans ce chapitre nous expliquerons le concept de l'IE ainsi que différents concepts associés. Ensuite nous décrirons le processus de l'IE tel qu'il est présenté par l'équipe de recherche SITE au sein du laboratoire LORIA<sup>42</sup>. Les acteurs de l'IE feront également l'objet d'une présentation ainsi que la place de la RCI dans le processus de l'IE.

## **2.2 Définitions et caractéristiques générales de l'intelligence économique**

La maîtrise de l'information, comme souligné dans l'introduction, est au cœur de l'intelligence économique. La collecte et l'utilisation de l'information à des fins utiles a toujours servi les desseins des hommes et des nations en quête de la prospérité (Levet 2001 p.1) et de la supériorité. Tout au long de l'histoire des sociétés, le besoin de rechercher de l'information et le besoin de protéger les informations jugées « sensibles » a toujours été présent (Menendez et coll. 2002 p.15). Dans tous les secteurs socio-économiques, l'on peut constater que la bonne gestion d'une entreprise ou d'une société ou d'une institution est caractérisée par un besoin de mettre en place des bonnes stratégies. La mise en place des bonnes stratégies n'est pas aisée. Elle est étroitement liée à la qualité d'informations externes et internes dont dispose l'entreprise. L'information en elle-même et quel que soit son volume, ne résout pas un problème. Ce qui est important est le traitement et l'interprétation de l'information qui permet d'extraire des connaissances. Si deux entreprises détiennent la même information, ceci n'implique pas qu'elles s'en serviront de la même façon. La capacité de donner du sens à l'information et de la transformer en connaissance qui mène à l'action est une condition essentielle pour la qualité de la démarche de l'IE.

Selon la définition proposée par Revelli, l'IE est un « processus de collecte, traitement et de diffusion de l'information qui a pour objet la réduction de la part d'incertitude dans la prise de toute décision stratégique... avec la volonté de mener les actions d'influence »<sup>43</sup> (Revelli 2000, p.18). Remarquons que dans cette définition, l'IE concerne la gestion d'informations et la prise de décision stratégique. On peut dire que l'IE est au cœur de la stratégie. En ce qui concerne la gestion d'informations, nous ne contestons pas l'importance de l'information et nous n'ignorons pas le besoin de les collecter. Cependant, nous souhaitons mettre l'accent sur le fait que ce sont les connaissances élaborées à partir de l'analyse des informations collectées et leur utilisation qui déterminent la réussite du processus décisionnel. Dans cette optique,

---

<sup>42</sup> Laboratoire lorrain en informatique et ses applications

<sup>43</sup> Selon Revelli, si l'on n'ajoute pas la volonté de « mener les actions d'influence » à cette définition, il ne s'agit donc que de l'intelligence stratégique. (Revelli 2000, p.18).

nous pouvons dire que « l'IE ne se réduit pas à la gestion des flux d'informations, mais s'intéresse fondamentalement à son interprétation, à son utilisation et donc à sa transformation en connaissances et en avantage concurrentiel durable » (Levet 2001 p.22).

La stratégie, selon (Menendez et coll. 2002, p.16) est un processus dialectique entre la situation externe et interne d'une entreprise. Ce processus comporte les techniques de benchmarking telles que la confrontation, la comparaison et le calibrage afin d'aider le décideur à développer un cadre d'actions vers la stratégie à long terme et vers le management à court terme.

La gestion de l'information concerne à la fois l'information interne et externe de l'entreprise. Cette gestion de l'information comporte deux concepts différents :

- La veille : dans toutes ses formes est indispensable pour maîtriser l'information externe de l'entreprise. Elle consiste en la surveillance active et organisée de l'environnement. On parle de la veille stratégique, technologique, marketing, juridique, normative, concurrentielle, brevet, etc.
- Le knowledge management (KM) : concerne la gestion des informations internes de l'entreprise. Ces informations comportent les savoirs faire collectifs et individuels au sein de l'entreprise, le capital humain, les clients, les fournisseurs, etc. (voir figure 2.1).

La veille constitue un élément crucial de l'IE mais elle n'en représente qu'une partie. L'IE englobe la veille et rejoint le KM comme le montre la figure 2.2 (Notons également que ces trois concepts sont rassemblés dans l'intelligence stratégique<sup>44</sup>). De ce point de vue il est possible de dire que la maîtrise de l'information et la capacité d'en tirer des connaissances constituent le socle de la prise de décision dans tous les secteurs socio-économiques.

L'IE est à la fois un processus décisionnel et informationnel comme le montre la définition d'Amos David et Odile Thiery :

---

<sup>44</sup> « Processus de collecte, traitement et de diffusion de l'information qui a pour objet la réduction de la part d'incertitude dans la prise de toute décision stratégique » (Révelli 2000, p.18).

« On peut considérer l'intelligence économique comme l'utilisation de l'information pour le processus décisionnel stratégique. De ce point de vue le processus de l'IE couvre deux domaines scientifiques établis : systèmes d'information et processus décisionnel. Ainsi, l'IE concerne tous les secteurs socio-économiques et non limités aux entreprises ou aux industries » (David et Thiery 2003 p.16).

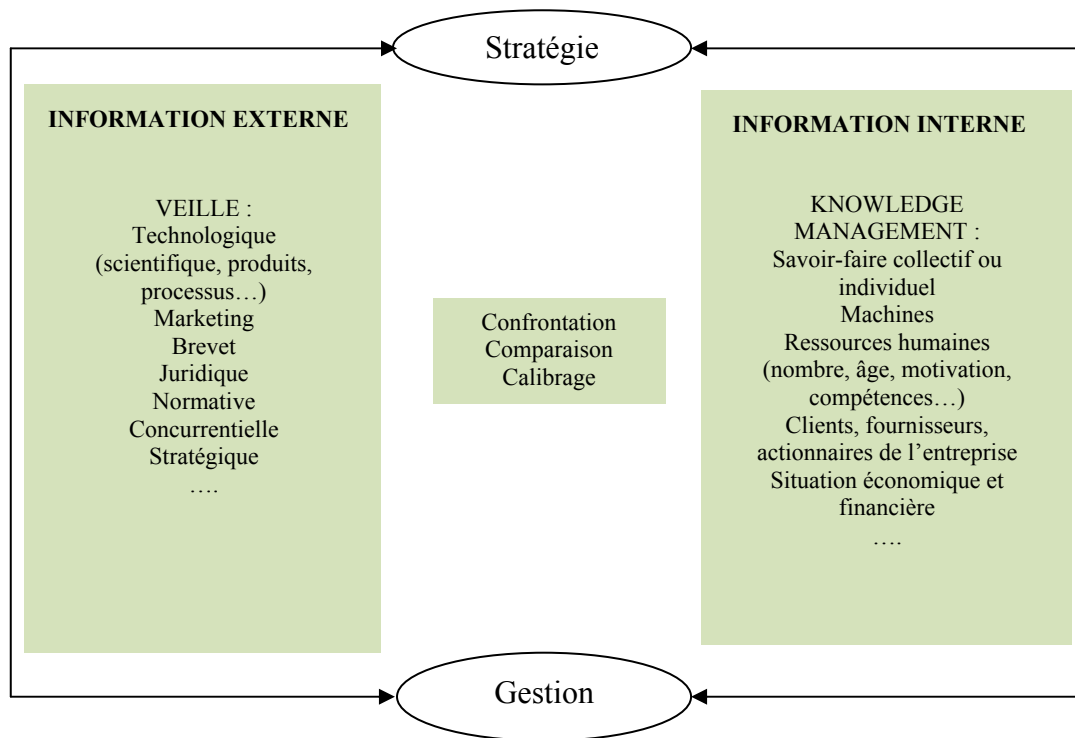


Figure 2.1 : Importance stratégique du management de l'information, adapté de (Menendez et coll. 2002, p.16)

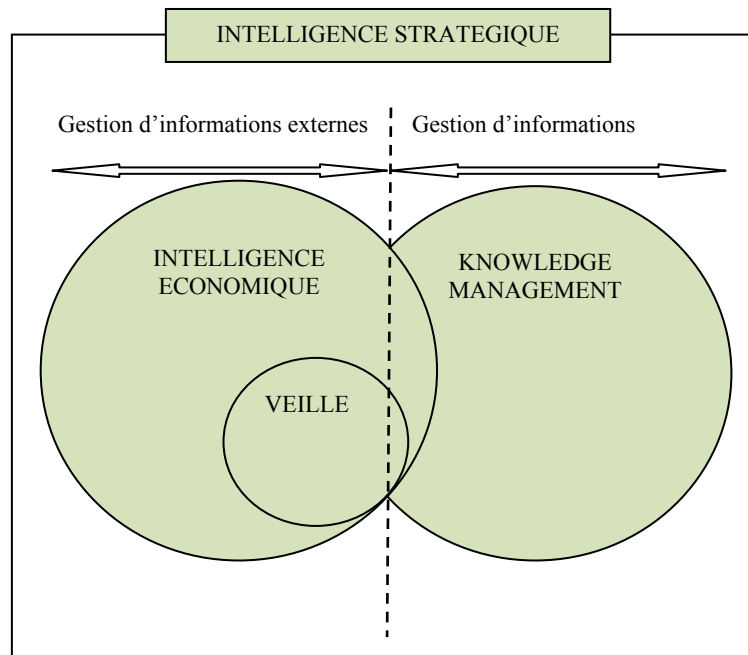


Figure 2.2 : Concepts associés à l'IE (Menendez et coll. 2002, p.19)

La définition de l'IE a fait l'objet de nombreux débats théoriques et pratiques<sup>45</sup>. De ce fait, plusieurs définitions ont été apportées à l'IE, chacune soulignant des caractéristiques différentes de l'IE. A travers des définitions différentes de l'IE, nous retrouvons les dimensions suivantes : la veille, la gestion de connaissances, la protection du patrimoine, la coordination, l'influence et la prise de décision, etc.

Deux de ces dimensions de l'IE qui nous intéressent se trouvent dans la définition de l'IE issue du rapport d'Henri Martre à l'initiative du Commissariat général du plan dans lequel l'IE est définie comme:

« L'ensemble des actions coordonnées de recherche, de traitement et de distribution en vue de son exploitation, de l'information utile aux acteurs économiques. Ces diverses actions sont menées légalement avec toutes les garanties de protection nécessaire à la préservation du patrimoine de l'entreprise dans les meilleures conditions de qualité, de délai et de coûts... » (Martre 1994, p.11).

Dans cette définition, en plus de la gestion de l'information, on y trouve la protection ou la préservation du patrimoine de l'entreprise qui explique le besoin de protéger les informations sensibles, valorisables et historiques de l'entreprise. Cette définition souligne également que

<sup>45</sup> Wikipédia consulté le 7 Mai, 2010.

les *actions informationnelles* sont menées légalement. Ceci montre que le processus de veille dans l'IE ne correspond pas à une forme d'espionnage.

Nous trouvons la définition de Bernard Besson et Jean-Claude Possin comme un résumé des différentes caractéristiques de l'intelligence économique. D'après ces deux auteurs, l'intelligence économique est « un système collectif d'acquisition, de production et de transformation de l'information en connaissances utiles. Ce système vise à l'amélioration du processus décisionnel, à l'exercice de l'influence, à la détection des opportunités, à la prévention des menaces et des risques, à la protection et à l'enrichissement du patrimoine, au développement de la performance, au maintien et au redressement de l'image, à la création de valeurs, à la sécurité et à la sureté. Offensif et défensif, ce système implique un projet, une stratégie, une éthique, une déontologie, une culture de l'anticipation, un partage de connaissances et la maîtrise des fonctions individuelles et collectives de mémoire, de réseaux, et d'analyse de l'information. Il est un modèle de management » (Besson et Possin 2006, p.36-37).

Pour rebondir sur cette diversité des approches conceptuelles de l'IE, nous reprenons la classification qui a été faite au niveau national en France par rapport à l'enseignement supérieur dans ce domaine. Cette classification est contenue dans le *Référentiel de formation en intelligence économique* qui a été publié en 2005 à la suite du 2<sup>ème</sup> Rapport officiel sur l'IE du député Bernard Carayon sorti en juin 2003, sous l'égide d'Alain Juillet (le haut responsable pour l'intelligence économique) et du Secrétariat Général à la Défense Nationale. Dans cette classification, cinq pôles de formation ont été identifiés chacun se composant de plusieurs axes:

- **Pole**  
**1 (Environnement de l'économie et compétitivité des nations)** ; il contient 4 axes à savoir :
  - mondialisati  
on de l'économie et compétitivité des nations ;
  - économie de  
l'information et de la connaissance ;
  - risques et  
menaces : la notion de sécurité économique ;

- les grands dispositifs nationaux d'intelligence économique.

- **Pole**

**2 (Intelligence économique et organisations)** ; il contient 3 axes à savoir :

- finalité et enjeux de l'intelligence économique dans les organisations ;
- le projet d'intelligence économique ;
- audit et pilotage du système d'intelligence économique.

- **Pole 3**

**(Management de l'information et des connaissances)** ; il contient 4 axes à savoir :

- identification et expression des besoins, animation du cycle de l'information ;
- gestion du recueil de l'information ;
- exploitation de l'information ;
- management des connaissances.

- **Pole 4**

**(Protection et défense du patrimoine informationnel et des connaissances)** ; il contient 4 axes à savoir :

- patrimoine informationnel ;
- propriété intellectuelle et industrielle ;
- politique de sécurité de l'information et de systèmes d'information ;
- la gestion des crises.

- **Pole**

**5 (Stratégies d'influence et de contre-influence)** ; il contient 3 axes à savoir :



- stratégies  
d'influence et de contre-influence ;
- utilisation  
offensive et défensive de l'information ;
- le lobbying.

### **2.2.1 Les acteurs de l'IE**

Il est important de souligner que le processus de l'intelligence économique implique plusieurs acteurs qui interviennent dans différentes phases du processus. Deux acteurs principaux sont généralement identifiés : le décideur et le veilleur. On peut également identifier d'autres acteurs tels que le coordinateur-animateur et l'analyste qui sont eux-mêmes des veilleurs, mais porteurs d'autres fonctions.

#### ***Le décideur***

Le décideur est l'individu disposant du pouvoir de décider. Il est celui qui est capable d'identifier et de poser le problème à résoudre en termes d'enjeux, de risques ou de menaces qui pèsent sur l'entreprise (David et Thiery 2001, p.3). Il est souvent confronté à une diversité de problème le conduisant à prendre différentes décisions (Afolabi 2007, p.68). Simon (1960, p.39-44) considère qu'un gestionnaire (manager) peut être assimilé à un décideur, car, pour lui, gérer c'est décider. Mintzberg (1984, p.65-111), en revanche, a critiqué ce propos en soulignant que l'activité et le rôle de manager sont à la fois interpersonnels, informationnels, et décisionnels. Dans le contexte de l'IE, le décideur est celui qui définit le problème décisionnel à résoudre. Selon (Kislin 2007, p.149), le décideur :

- « identifie les signaux faibles de l'environnement et les stratégies qui permettraient d'améliorer les performances de l'entreprise ;
- opère parmi ces stratégies des choix en fonction des caractéristiques du secteur d'activité, des concurrents et des partenaires ;
- alloue des ressources financières et humaines aux actions spécifiques engendrées par la prise de décision ;
- mesure et assume les conséquences liées à la prise de décision ».

### ***Le veilleur***

Le veilleur est celui qui est chargé de l'activité de veille. Lopes DaSilva le présente comme celui « qui est capable de mettre en place un dispositif de veille » (Lopes 2002, p.47). Il est chargé de collecter, d'analyser et diffuser l'information en vue de rendre plus intelligible et plus lisible l'environnement externe et interne de l'entreprise (Kislin 2003, p.136). Selon le référentiel métier de l'ADBS, le veilleur est celui qui « alimente les décideurs d'une entreprise en informations sélectionnées et traitées en vue de les alerter sur l'évolution de l'environnement (technique, concurrentiel, économique, réglementaire, etc.) de l'entreprise et de les aider dans leurs prises de décision. »<sup>46</sup>. Le veilleur peut ainsi être considéré comme un médiateur entre le décideur et le monde de l'information. Il est un spécialiste qui surveille en permanence les contenus de l'information, et suit les évolutions de flux informationnels et alimente le système d'information de l'entreprise (Kislin 2007, p.149).

### ***Le coordinateur-animateur***

Le coordinateur-animateur est celui qui coordonne le processus d'intelligence économique entre les divers acteurs impliqués (Knauf 2007, p.240). « Il fidélise le décideur en lui envoyant régulièrement des informations sur l'évolution de son environnement » (Knauf 2007, p.250). Il le conseille et l'accompagne. Il a une bonne connaissance du réseau local (Goria et Knauf 2006, p.4-5) et il participe à des actions de sensibilisation et de communication.

### **L'analyste**

L'analyste est un veilleur qui est aussi un expert dans l'analyse de l'information et dans l'extraction d'indicateurs. Il travaille avec d'autres veilleurs et le décideur afin de bien extraire et interpréter les indicateurs pour la prise de décision.

## **2.3 Processus de l'IE selon SITE**

L'équipe SITE dont l'intitulé exact est Modélisation et Développement de Systèmes d'Intelligence Economique a, dès le début de sa création en 2001, défini comme

---

<sup>46</sup> Référence tiré de la thèse de Philippe Kislin (2007, p.147). **Source d'origine** : ADBS (Association des professionnels de l'information et de la documentation), Management de l'information : panorama des métiers, 2002.

problématique l'étude et la modélisation ainsi que le développement de systèmes d'informations stratégiques dans le cadre de l'intelligence économique. L'équipe SITE est l'une des équipes de recherche du laboratoire lorrain en informatique et ses applications (LORIA) à Nancy.

### **2.3.1 Les axes de recherche de l'équipe SITE**

Dans le but de traiter sa problématique, SITE définit trois axes de recherche :

- Modélisation  
de l'utilisateur-acteur
- Modélisation  
de l'interaction entre l'utilisateur et le médiateur
- Conception  
et exploitation d'un entrepôt de données

Nous détaillerons ces axes de recherche ci-dessous.

#### ***2.3.1.1 Modélisation de l'utilisateur-acteur***

Cet axe de recherche est centré sur la modélisation de l'utilisateur et la prise en compte de ses particularités dans le processus de l'intelligence économique. Comme nous l'avons expliqué précédemment (cf. 2.2), la capacité de tirer de connaissances de l'information que l'on détient est l'une des spécificités de la démarche de l'IE. Cette capacité est liée au niveau de connaissances des acteurs humains qui interviennent dans le processus de l'IE. Notons également que les connaissances dont disposent les acteurs sont directement liées à leurs intentions, expériences, compétences, préférences, etc. Donc, le but de cet axe de recherche est de proposer des modèles et des méthodes permettant d'intégrer les connaissances sur les acteurs dans le système d'informations stratégiques et le système de recherche d'informations afin de permettre au système de produire des résultats aussi pertinents que possible en réponse au besoin en information de l'utilisateur.

#### ***2.3.1.2 Modélisation de l'interaction entre l'utilisateur et le médiateur***

L'IE comme expliqué dans le paragraphe 2.2, implique plusieurs acteurs qui interagissent dans les différentes phases du processus. Cet axe de recherche a donc comme objectif de

modéliser l'interaction entre des collaborateurs ou autrement dit des partenaires<sup>47</sup>. L'interaction dans ce contexte s'effectue généralement entre un utilisateur et un médiateur, entre le décideur et un veilleur, entre des veilleurs ou simplement entre deux utilisateurs d'un système d'information. Cet axe de recherche s'intéresse au concept de communication et de relation interpersonnel dans le cadre de la collecte et la recherche d'information dans le processus d'IE.

### ***2.3.1.3 Conception et exploitation d'un entrepôt de données***

Cet axe de recherche est centré sur la modélisation et la conception d'un système d'informations stratégiques de qualité, qui permettent de répondre aux besoins de différents acteurs de l'entreprise. L'objectif est de proposer une architecture, un modèle et une démarche pour concevoir un système d'information tout en intégrant les paramètres des utilisateurs et les évolutions de celles-ci. L'évolution des paramètres de l'utilisateur peut surgir de l'évolution de l'utilisateur lui-même ou de l'évolution de l'environnement de l'entreprise. Cet axe concerne les problématiques informatiques telles que la modélisation de données, la méthodologie de conception d'un système d'information adaptatif et les systèmes d'informations stratégiques.

Nous présentons dans la section suivante le processus de l'IE adopté par l'équipe SITE et les modèles issus des réflexions des membres de l'équipe pour la mise en œuvre du processus.

### **2.3.2 Processus de l'IE**

D'après l'équipe SITE, l'IE est un processus à la fois décisionnel et informationnel composé des étapes suivantes :

1. identification d'un problème décisionnel ;
2. traduction du problème décisionnel en problème de recherche d'informations ;
3. identification des sources pertinentes d'information ;
4. validation des sources d'information
5. collecte des informations pertinentes ;
6. analyse des informations collectées pour extraire des indicateurs pour la décision ;

---

<sup>47</sup> Nous préférons utiliser le mot « partenaires » au lieu de « collaborateurs » car l'interaction entre deux acteurs dans ce contexte peut ne pas dépendre une situation de collaboration.

7. interprétation des indicateurs ;
8. prise de décision.

Ces étapes peuvent être résumées en cinq phases du processus de l'IE à savoir :

- La phase décisionnelle : elle concerne l'identification du problème décisionnel et sa traduction en problème informationnel. Elle couvre les étapes 1 et 2 du processus. Dans cette phase, l'objectif est d'identifier l'enjeu du problème décisionnel. Si un problème décisionnel n'est pas bien défini et s'il n'est pas bien traduit en problème informationnel, toutes les activités impliquées dans les autres phases du processus de l'IE peuvent être inutiles. Les acteurs principaux dans cette phase sont les décideurs et les veilleurs.
- La phase de collecte d'informations : cette phase couvre les étapes 3, 4 et 5. Après la traduction d'un problème décisionnel en problème de recherche d'informations, il est nécessaire d'identifier les sources pertinentes d'information, de les valider et de collecter les informations pertinentes pour la résolution du problème. Les acteurs qui interviennent dans cette phase sont principalement les veilleurs. Les concepteurs de système d'information peuvent intervenir également dans cette phase.
- La phase d'analyse : elle couvre les étapes 6 et 7. Cette phase est importante, car elle permet de traiter les informations collectées afin d'extraire les indicateurs nécessaires pour la prise de décision. Les veilleurs sont des acteurs principaux dans cette phase.
- La phase de la prise de décision : elle couvre les étapes 1, 7 et 8. Une interprétation juste des indicateurs aboutira à la prise de bonnes décisions. Les décideurs sont les acteurs principaux dans cette phase.
- La phase de protection du patrimoine informationnel : cette phase n'est pas une phase isolée, car

elle couvre toutes les huit étapes du processus de l'IE. De l'identification du problème décisionnel à la prise de décision, toutes les informations collectées et exprimées doivent être protégées contre l'espionnage, le piratage des concurrents et toute diffusion nuisant à l'entreprise. Les acteurs intervenant dans cette phase sont les décideurs, les veilleurs, les infomédiaires, les analystes et les concepteurs des systèmes d'information.

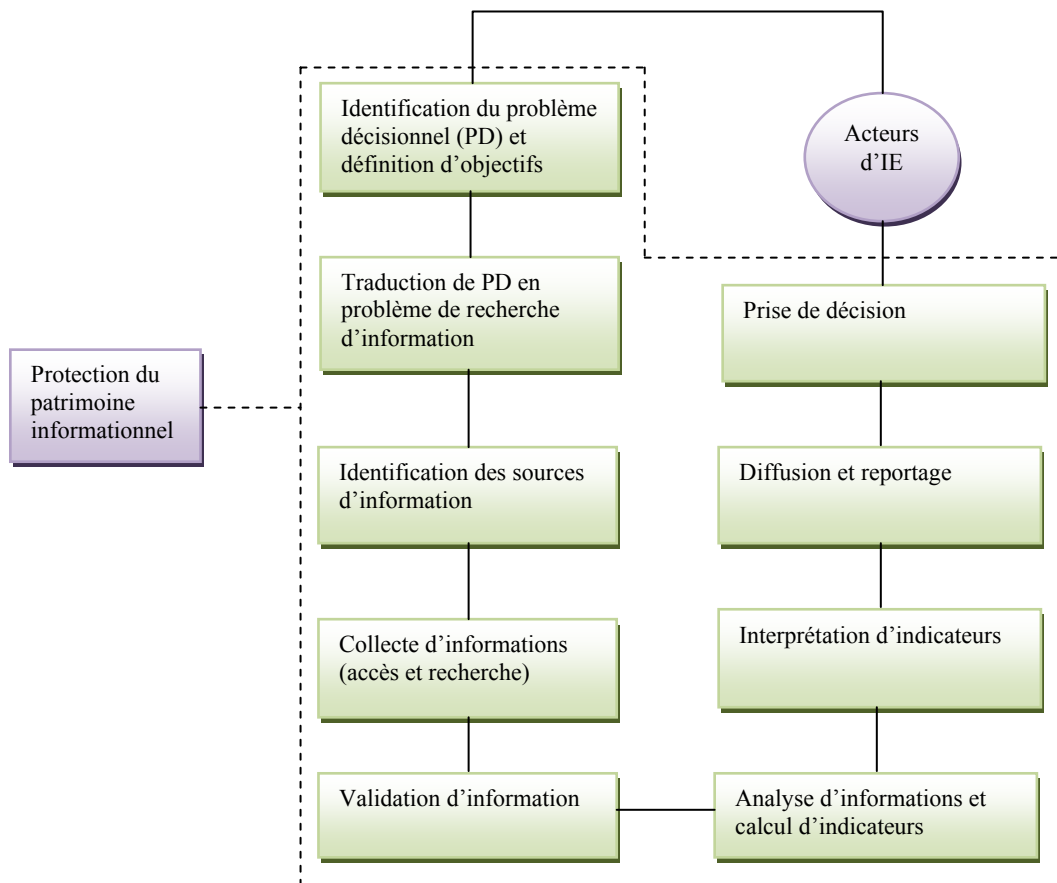


Figure 2.3 : le cadre d'intelligence économique

### 2.3.3 Le paradigme des abeilles

Selon Amos David, responsable de l'équipe SITE, les problématiques liées au processus d'intelligence économique peuvent être comparées aux propriétés ou facultés des sociétés des abeilles (repérer, indiquer, collecter, construire, organiser et protéger) (David 2006, p. 9-10). Nous présentons ci-dessous les propriétés et les liens avec l'intelligence économique et la collaboration entre les acteurs d'IE en conformité avec le réseau social des abeilles introduit par (David 2006).

**Repérer** : (localiser et valider) : *les abeilles ont une formidable faculté pour repérer les fleurs qui peuvent fournir du nectar. Il ne s'agit donc pas de localiser une plante, mais d'identifier les plantes utiles.* Cette faculté nous paraît un élément essentiel voir indispensable, car elle est déterminante dans un processus d'IE. En effet, non seulement l'acteur doit pouvoir identifier l'information, il doit également être en capacité de juger de la pertinence ou de l'utilité d'une information. L'activité du veilleur est basée principalement sur ce type de faculté.

**Indiquer** (langage commun de communication) : les abeilles sont capables de communiquer aux autres abeilles la direction et la distance d'une plante contenant du nectar. Là aussi nous considérons qu'il est possible d'établir une similitude avec les acteurs d'un processus d'IE qui doivent être capables de trouver les bonnes informations. A cet effet, les acteurs doivent disposer et mettre en œuvre des compétences de communication adaptées pour transmettre d'une part, l'information, et d'autre part, la localisation de l'information.

**Collecter** (utiliser l'existant) : *les abeilles ne se contentent pas de repérer, mais elles collectent le nectar pour la société. Pour cela elles disposent d'outils naturels pour le faire. Elles collectent le nectar pour une utilité commune de la société et elles contribuent inconsciemment à la pollinisation des plantes.* S'il nous fallait comparer l'action des abeilles au processus d'IE nous pourrions considérer que l'acteur doit également disposer d'outils spécifiques afin de collecter l'information. Cependant à l'instar des abeilles ces outils n'ont rien de naturel. C'est pour cette raison que l'homme conçoit et développe des outils adaptés à ses besoins. L'élaboration de ces outils de collecte d'informations apparaît donc comme tout à fait indispensable pour l'homme. Le processus d'IE à travers la collecte d'informations concerne bien évidemment un groupe d'individus ou une entreprise. Nous pourrions d'une certaine manière considérer que nous sommes un peu dans une forme de pollinisation puisque l'information collectée vient enrichir la croissance de l'entreprise par le processus d'IE.

**Construire** (fabriquer l'inexistant) : *une autre faculté admirable des abeilles concerne la faculté de construction. Les abeilles fabriquent les ruches avec une précision presque parfaite. On remarque donc qu'elles disposent de facultés à fabriquer un objet inexistant pour des besoins de leur société.* On peut considérer que les acteurs du processus d'IE doivent eux aussi tout comme les abeilles faire advenir un objet au départ inexistant. Ils y parviennent grâce à leur capacité à communiquer et à élaborer des processus de collaboration. Et c'est

ainsi qu'ils fabriquent les objets nécessaires pour répondre aux besoins et au pilotage de l'organisme socio-économique.

**Organiser** (partage des rôles, sens d'appartenance) : *la structure organisationnelle d'une colonie d'abeilles est presque parfaite. Chaque membre de la colonie connaît son rôle et les hiérarchies sont parfaitement respectées. On remarque également une forme de changement de rôle en fonction de l'âge, que nous pouvons associer à une forme d'attribution de rôle en fonction d'expérience.* Dans le processus d'IE si l'on tient compte de la métaphore de la ruche il serait donc souhaitable de modifier la distribution des rôles et des fonctions afin de répondre au mieux aux besoins de l'organisme. Cette mobilité des rôles et des fonctions pourrait s'appuyer sur l'expérience des différents membres de l'organisation.

**Protéger** (se défendre avec des armes adaptées - piqûre, chaleur par vibration) : *les abeilles mellifères européennes ne peuvent pas se défendre contre les frelons géants asiatiques. En effet leur dard est sans effet sur ces frelons. Par contre la variété japonaise d'abeilles mellifères, dont le dard est également inefficace, a cependant su développer une stratégie de défense efficace contre le prédateur : elles font mourir de chaleur les frelons en s'agglutinant dessus par groupe de plusieurs centaines d'abeilles pour un frelon. Cette faculté montre la nécessité d'avoir une attitude collective pour la défense du patrimoine de l'organisme, notamment dans un processus d'IE. Les stratégies de défense doivent être adaptées à des environnements et à ceux qui cherchent à nuire à la bonne marche de l'organisme.*

Tout comme les abeilles il nous faut trouver les stratégies pour défendre le patrimoine de l'organisme de façon à repousser les attaques des personnes qui voudraient se livrer au pillage de l'information.

La comparaison avec le monde des abeilles et notamment avec leurs formes d'organisations nous paraît tout à fait pertinente pour éclairer le processus d'IE. Ce dernier ne peut faire l'économie d'une parfaite connaissance de l'environnement. Mais cette connaissance n'est pas suffisante en soi. Il convient d'avoir un projet commun et partagé. C'est la condition sine qua non pour garantir la réussite du projet collaboratif. C'est également en développant un sentiment d'appartenance à un groupe et en fédérant les acteurs autour d'enjeux collectifs que le travail collaboratif prend tout son sens.



### **2.3.4 Les modèles de l'IE au sein de l'équipe SITE**

Plusieurs modèles ont été développés au cours des années au sein de l'équipe SITE dans le but de résoudre les problèmes liés à l'IE. Parmi ces modèles, nous expliquerons le modèle pour l'explicitation du problème décisionnel (MEPD) proposé par Najoua Bouaka pour faciliter l'identification et la représentation du problème décisionnel et le modèle de recherche d'informations du veilleur proposé par Philippe Kislin. Les autres modèles sont résumés dans le tableau 2.1.

#### ***2.3.4.1 Définition d'un problème décisionnel***

D'après (Bouaka 2004, p.98), un problème décisionnel dans un contexte de l'IE peut être défini comme étant « l'écart qu'il peut y avoir entre une situation donnée et une situation de référence jugée stable ». Elle ajoute que cet écart se manifeste par un « signal ». Dans son exemple : la chute des ventes et la démission de plusieurs employés dans une entreprise ne sont pas forcément des problèmes décisionnels, mais ils représentent des signaux pouvant aider à identifier le vrai problème qui se cache sous cette situation. Ceci veut dire qu'un signal n'est qu'un premier élément dans l'identification d'un problème décisionnel. Notons que le fait qu'un décideur présente un problème décisionnel montre qu'il est confronté à une difficulté et que probablement il ne parvient pas à identifier la ou les causes du problème en raison d'un manque de connaissances précises sur la situation en question. Ce manque de connaissance peut être traduit en besoin informationnel pouvant aider le décideur à résoudre son problème décisionnel.

Bouaka décrit trois facettes d'un problème décisionnel nécessaires pour son explicitation. Ces facettes, qui peuvent influencer le choix de sources d'information et leur utilisation pour la résolution du problème décisionnel, comportent :

- le décideur qui identifie le problème ;
- l'environnement qui détermine le contexte du problème.
- l'organisation qui subira l'effet du problème ;

**Modélisation du décideur** : Bouaka, dans son modèle du décideur, intègre ses caractéristiques individuelles telles que : son identité, sa formation initiale, son style

cognitif<sup>48</sup>, ses traits de personnalité et ses expériences. Ces caractéristiques influent sur la compréhension du problème par le décideur.

**Modélisation de l'environnement** : l'environnement d'une entreprise concerne le monde qui l'entoure. Comme expliqué précédemment (cf. 2.2), la prise de décision stratégique dépend de la maîtrise de l'information interne et externe. Modélisation de l'environnement selon Bouaka concerne la gestion des informations en provenance de l'environnement externe de l'entreprise afin de détecter tout événement susceptible d'affecter son fonctionnement. Le but est donc de repérer les facteurs qu'un décideur considère sensibles et pouvant aider à détecter des signaux faibles. Bouaka adopte le point de vue de Bourgeois (1980, p.25-39) qui propose deux catégories d'environnement :

- environnement immédiat qui est l'environnement proche de l'entreprise qui l'influence d'une façon directe. Cela concerne, par exemple : des clients, des fournisseurs et des concurrents de l'entreprise;
- environnement global qui est l'environnement universel, qu'il soit au niveau régional, national, international, regroupant l'environnement social, économique, politique, législatif, etc.

**La modélisation de l'organisation** : permet de rassembler les informations concernant les caractéristiques de l'entreprise. Ces caractéristiques constituent des facteurs déterminants qui influencent le comportement du décideur par rapport au problème identifié. Cette modélisation permet d'identifier les effets d'observation du décideur sur l'entreprise. Elle met en évidence des informations sur la vision que le décideur a de l'entreprise. Cela concerne la perception de ce dernier. Selon Bouaka, en modélisant l'entreprise, l'accent doit être mis sur toute entité qui a le potentiel d'être source d'enjeu, soit la source ou la cible d'une action de l'entreprise. Tout événement observé (constituant l'enjeu) peut être décomposé en trois éléments :

- l'objet de l'environnement qui est la cible sur laquelle le décideur souhaite agir (Kislin 2007, p.237)

---

<sup>48</sup> « Le style cognitif peut être défini comme la façon propre à chacun de percevoir et de comprendre l'information perçue face à une nouvelle connaissance. Ce style influe sur la manière dont le décideur traite l'information reçue de son environnement et la manière de la communiquer à une autre personne, entre autres, le veilleur » (Bouaka 2004, p.108).

- le signal émis par cet objet qui représente ce qui pousse le décideur à formuler un problème décisionnel
- l'hypothèse que nous pouvons déduire de la détection de ce signal. Elle comporte le(s) risque(s) perçu(s) et les conséquences (en termes de gains ou de pertes) envisagées si le décideur réagit ou ne réagit pas au signal.

Bouaka fait l'hypothèse que « si le décideur arrive à identifier les signaux en provenance de son environnement, on peut l'aider à projeter sa réflexion dans le temps et déterminer l'enjeu relatif a cette observation. De ce fait, la dimension de l'enjeu demeure une dimension fondamentale pour passer du niveau d'explicitation d'un problème au niveau du problème de recherche d'informations ».

Ces trois facettes d'un problème décisionnel constituent le modèle pour l'explicitation du problème décisionnel (MEPD) proposé par Bouaka. Ce modèle a été repris par Kislin (2007, p.225-234) dans son modèle de recherche d'informations du veilleur WISP (*Watcher Information Search Problem*).

#### ***2.3.4.2 Passage du problème décisionnel au problème informationnel***

Dans le processus de l'IE selon l'équipe SITE, la deuxième étape du processus concerne la traduction du problème décisionnel en problème informationnel. Le modèle de Bouaka et celui de Kislin s'intercalent pour traduire un problème décisionnel en problème informationnel. Kislin identifie six éléments principaux qui constituent le problème informationnel dans son modèle :

- demande : elle représente les différentes formulations et reformulations du besoin informationnel par le décideur;
- enjeu : représente l'enjeu perçu du problème décisionnel. Les trois éléments d'enjeu du modèle MEPD sont repris par Kislin ;
- indicateur : correspond à la traduction validée de la demande dans un langage 'documentaire' particulier. C'est le noyau du modèle WISP ;

- problème de recherche d'informations : représente l'objectif de recherche d'informations du veilleur en langage naturel ;
- résultat : représente les produits informationnels issus de la veille ;
- analyse : représente les analyses réalisées sur les résultats de recherche et sur le processus de recherche.

Ces éléments comportent plusieurs sous-éléments permettant de décrire au mieux le problème informationnel.

#### 2.3.4.3 D'autres modèles développés au sein de l'équipe SITE

Les autres modèles qui ont été développés au sein de l'équipe SITE sont :

- MIRABEL : *Model for Information Retrieval query Annotations Based on Expression Levels*
- RUBICUBE : Récupération, Utilisateur, Besoins, Identification, Classification, Usinage, Base métiers, Enrichissement
- MORPRIE : Modèle de Résolution d'un problème de recherche d'informations en Intelligence Économique
- AMIE : *Annotation model for Information Exchange*
- CADRIE : Caractérisation des compétences du Coordinateur-Animateur d'un Dispositif Régional d'Intelligence Économique
- COMIR : *Contextual model for Multimedia Information Retrieval*

Modèle	objectif	Etape d'implication dans le processus de l'IE selon SITE
MEPD : Bouaka Najoua (2004)	Ce modèle combine à la fois le contexte du problème, les représentations du décideur et les enjeux liés à la prise de décision. Son objectif est de faciliter l'identification et la représentation du problème décisionnel d'une part et la préparation du projet de recherche d'informations d'autre part. Il aide à la compréhension du problème au travers de l'identification de différentes caractéristiques de l'utilisateur et l'estimation du niveau du risque identifié.	<ul style="list-style-type: none"> <li>- identification et définition d'un problème décisionnel</li> </ul>
WISP : Kislil Phillip (2005)	Traduction du problème décisionnel en problème de recherche d'informations et spécification du problème informationnel par le veilleur.	<ul style="list-style-type: none"> <li>- traduction du problème décisionnel en problème de recherche d'informations</li> <li>- identification des</li> </ul>

		<ul style="list-style-type: none"> <li>sources d'information pertinentes</li> <li>- collecte des informations pertinentes</li> </ul>
MIRABEL : Goria Stéphane (2006)	C'est un modèle d'aide à l'expression d'un problème informationnel pour initier la recherche et l'analyse de son contexte en fonction d'un processus d'inférences au sujet de la compréhension du problème par le décideur (utilisateur) et le veilleur. Il permet de mettre en évidence des variations d'interprétation d'un problème ou d'un concept selon les personnes qui l'interprètent.	<ul style="list-style-type: none"> <li>- identification et définition d'un problème décisionnel</li> <li>- traduction du problème décisionnel en problème de recherche d'informations</li> </ul>
RUBICUBE : Frédérique Peguiron (2006)	Adapte la conception d'un système d'information en fonction des catégories de besoins des utilisateurs en amont du processus de conception d'un entrepôt de données dédiés à un environnement d'enseignement universitaire.	<ul style="list-style-type: none"> <li>- identification des sources d'information pertinentes</li> <li>- collecte des informations pertinentes.</li> </ul>
MORPRI2E : Afolabi Babajide (2007)	Représente les utilisateurs en fonction de leurs besoins spécifiques dépendants du contexte et modélise un système d'information qui s'adapte en fonction des changements des besoins de ses utilisateurs.	<ul style="list-style-type: none"> <li>- identification des sources d'information pertinentes</li> <li>- collecte des informations pertinentes</li> </ul>
AMIE : Roberts Charles (2007)	Destiné à la contextualisation d'informations à partir d'annotations dans un système d'intelligence économique aussi bien qu'à la validation des informations selon leur pertinence et leur fiabilité.	<ul style="list-style-type: none"> <li>- analyse des informations collectées pour extraire les indicateurs pour la prise de décision</li> </ul>
CADRIE : Audrey Knauf (2007)	Dédié à la spécification des rôles et des compétences du coordinateur-animateur d'un dispositif régional d'IE.	<ul style="list-style-type: none"> <li>- couvre toutes les phases du processus</li> </ul>
COMIR : Hanène Maghrebi (2010)	Ajoute des paramètres contextuels pour la représentation d'informations multimédia.	<ul style="list-style-type: none"> <li>- collecte des informations pertinentes</li> </ul>

Tableau 2.2 : Modèles développés au sein de l'équipe SITE (Oladejo 2008)

#### 2.4 La recherche collaborative d'information dans le processus de l'IE

Les différentes définitions de l'intelligence économique que nous avons expliquées dans ce chapitre montrent que la collecte ou la recherche d'informations est une activité centrale dans la démarche de l'IE. L'une de ces définitions indique que les actions de collecte et de recherche d'informations dans l'IE sont coordonnées (Martre 1994, p.11) et une autre définition indique que l'IE est un système collectif (Besson et Possin 2006, p.36-37). Nous

remarquons également que la notion de « travailler ensemble » est souvent évoquée dans les démarches de l'IE : David dans le paradigme des abeilles parle de travail collaboratif dans l'IE (David 2006, p. 9-10) ; Kislin parle de la coopération entre le décideur et le veilleur (Kislin 2003, p.99).

Tout au long du processus de l'IE, différents acteurs travaillent ensemble pour la réussite de la démarche. Un décideur peut collaborer avec un veilleur pour bien traduire un problème décisionnel en problème informationnel. Des veilleurs peuvent collaborer entre eux pour rechercher des informations pertinentes à la résolution d'un problème informationnel donné. Nous pouvons considérer que l'importance de la collaboration dans le processus de l'IE, surtout lors de la recherche d'information est inestimable

Notre but dans ce travail de thèse est de concevoir et de développer des méthodes et des technologies pour faciliter la recherche collaborative d'informations par les acteurs impliqués dans le processus d'intelligence économique. Le tableau 2.2 permet de visualiser le niveau d'implication de ce présent travail de recherche dans le processus de l'IE.

<b>Etape</b>	<b>Acteur</b>	<b>Implication de notre thèse</b>
Identification d'un problème décisionnel	Décideur	Non
Traduction du problème décisionnel en problèmes de recherche d'information	Décideur + veilleur	Oui
Identification des sources pertinentes d'informations	Veilleurs	Oui
Validation des sources d'information	Veilleurs	Oui
Collecte des informations pertinentes	Veilleurs	Oui
Analyse des informations collectées pour extraire des indicateurs pour la décision	Veilleurs	Oui
Interprétation des indicateurs	Décideur	Non
Prise de décision	Décideur	Non

Tableau 2.2 : Notre travail de recherche par rapport au processus de l'IE selon l'équipe SITE

## **2.5 Conclusion du deuxième chapitre**

Nous avons présenté le concept de l'IE dans ce chapitre. Le processus de l'IE ainsi que les acteurs impliqués dans le processus ont été présentés. Pour conclure ce chapitre et avant

d'énoncer les prochains développements de ce travail, il nous paraît important de préciser la nécessité de comprendre le concept de la RCI. Pour ce faire, il faut d'abord comprendre les deux concepts qui la composent. Nous avons développé la notion de collaboration dans le premier chapitre et nous allons dans le chapitre suivant décrire le concept de la recherche d'information tout en élaborant les évolutions qui ont eu lieu dans ce domaine.







# Chapitre 3

---

## De la recherche d'information à la recherche sociale d'information

### 3.1 Introduction

Le terme *information retrieval* a été proposé par Calvin Mooers en 1950 pour désigner le processus d'indexation automatique et de recherche d'information (Mooers 1959, p.81). Selon lui, « la recherche d'information est le nom d'un processus ou d'une méthode par lequel un utilisateur d'informations est capable de convertir son besoin informationnel en une liste de documents contenant les informations utiles à la satisfaction de son besoin. La recherche d'information comporte des aspects intellectuels de la description de l'information et ses spécifications pour la recherche, et l'ensemble des systèmes, des techniques ou des machines qui sont employés pour effectuer cette tâche de recherche »<sup>49</sup> (Mooers 1951, p.25). A la suite des travaux de Moers, F.W Lancaster en 1979 a défini la recherche d'information comme « le processus de rechercher un ensemble de documents avec le but d'identifier des documents qui sont liés à un thème particulier <sup>50</sup> ». Pour l'auteur, le terme « recherche d'information » n'est pas satisfaisant puisqu'il ne décrit pas l'activité impliquée. Pour lui, un système de recherche d'information (SRI) ne retrouve pas l'information car l'information est intangible<sup>51</sup>. Il a ainsi distingué la recherche documentaire (*document retrieval*) d'une recherche de renseignement (*fact retrieval*) (Lancaster 1979, p.13).

Une ambiguïté s'installe dans les définitions données à la « recherche d'information ». Cette ambiguïté existe à la fois en anglais et en français. Elle est issue du simple fait de mélanger le processus de recherche d'information avec le SRI. Le processus de recherche est plus complexe et robuste que le fonctionnement ou l'utilisation du SRI. Par exemple, dans la définition donnée par Baeza-Yates et Ribeiro-Neto (1999, p.1), la recherche d'information

---

<sup>49</sup> « *Information retrieval is the name for the process or methode whereby a prospective user of information is able to convert his need for information into an actual list of citations to documents in storage containing information usefull to him* » (Mooers 1951, p.25).

<sup>50</sup> « *Information retrieval is the process of searching a collection of documents with the goal of identifying documents that relate to a particular topic* » (Lancaster 1979, p.11)

<sup>51</sup> « *...information retrieval is not a particularly satisfactory term to describe the type of activity to which it is usually applied. An information retrieval system does not retrieve information. Indeed, information is intangible* » (Lancaster 1979, p.12)

concerne la représentation, le stockage, l'organisation et l'accès aux objets informationnels. Cette définition porte plutôt sur le système de recherche que sur le processus de recherche d'information. En revanche, la recherche d'information, telle qu'elle est définie par Fidel et coll. (2001, p.236), peut être interprétée dans un sens large qui consiste en des processus tels que l'identification du problème informationnel, l'analyse de besoin informationnel, la formulation de la requête, l'interaction avec le système, l'évaluation des résultats, et leur présentation. Cette dernière définition implique une dimension beaucoup plus large que celle du SRI.

La « recherche d'information » est un vaste domaine d'étude au carrefour de plusieurs disciplines telles que la science de l'information, l'informatique, la linguistique, et la psychologie cognitive. Deux axes principaux de recherche découlent de ce domaine : l'étude du comportement informationnel de l'utilisateur et l'étude de SRI.

Dans ce chapitre, nous présenterons ces deux axes d'études en RI, en relevant les principaux apports réalisés dans ce domaine au cours des 60 dernières années. En premier lieu, nous examinerons le comportement informationnel des utilisateurs sous quelques perspectives d'étude. Puis, nous présenterons quelques modèles de recherche d'information. Nous présenterons ensuite le concept général d'un système de recherche d'information, et les évolutions majeures qui lui sont associées.

### **3.2 Le comportement informationnel de l'utilisateur**

Les études sur le comportement informationnel de l'utilisateur ont débuté il y a plus de soixante ans. Avant 1980, ce domaine de recherche est appelé « étude des utilisateurs<sup>52</sup> » et a cherché à répondre à la question : comment les systèmes d'informations et la bibliothèque sont-ils utilisés et dans quel but (Wilson 2010, p.29) ? A partir des années 1980, la tendance a été d'établir, pour ce sujet, des cadres conceptuels à partir desquels sont nés plusieurs modèles et théories faisant notamment intervenir le besoin utilisateur et l'*information seeking*. Plusieurs approches ont été envisagées :

- l'approche cognitive (Ingwersen et Jarvelin 2005, p.23-53), étudie la façon dont un utilisateur traite de l'information. L'accent est mis sur la structure cognitive de l'utilisateur vis-à-

---

<sup>52</sup> *User studies* (Wilson 2010, p.29)

vis de son problème, de sa connaissance du domaine exploré, de ses expériences et du SRI. Cette approche, selon Ihadjadene et Chaudiron, « est perçue comme particulièrement féconde, tant sur le plan conceptuel que méthodologique, par une communauté de chercheurs en sciences de l'information ayant placé l'utilisateur au centre de leur préoccupations » (Ihadjadene et Chaudiron 2008, p.197). Amos David, en s'inspirant des habitudes évocatrices d'apprenant, a utilisé cette approche pour développer un modèle cognitif de recherche d'information - EQUA<sup>2</sup>e<sup>53</sup> (David 2003, p.17-18) ;

- l'approche phénoménologique (Savolainen 2008, p.15-75), permet d'étudier le comportement informationnel de l'utilisateur comme une pratique quotidienne comportant l'accès à l'information, l'utilisation de l'information et son partage ;
- l'approche de constructivisme social (Tuominen 2005, p.238), met l'accent sur l'interaction sociale dans le comportement informationnel de l'utilisateur;
- l'approche de théorie d'activité, permet d'étudier le comportement informationnel de l'utilisateur sous l'angle (l'aspect) de son « activité. » Dans cette théorie, une activité est composée d'actions et chaque action est composée d'opérations. Notons que toute activité se déroule dans un contexte socioculturel et historique. Un problème informationnel suscite un besoin correspondant et nécessitant une activité pour le résoudre. L'accès à l'information pour résoudre ce problème est un ensemble des actions qui nécessite d'effectuer des opérations, comme la formulation d'une requête et l'évaluation d'un document (Wilson 2006).

Wilson a proposé un modèle imbriqué incluant les trois concepts du comportement informationnel de l'utilisateur: *l'information retrieval (IR)*, *l'information seeking (IS)* et *l'information behaviour (IB)*. Comme le montre la figure 3.1, ce modèle imbriqué comporte trois couches. La couche extérieure correspond à l'IB, qui englobe l'IS et l'IR correspondant aux deuxième et troisième couches respectivement. Nous ne pouvons donc tenter de comprendre l'IR sans comprendre l'IS et l'IB. Dans sa publication, Wilson n'a pas utilisé le

---

<sup>53</sup> Le modèle EQUA<sup>2</sup>e (David 2002, p.33-36) signifiant en anglais : *explore* (explorer), *query* (interroger), *analyze* (analyser), *annotate* (annoter)

terme *information retrieval*, mais originellement celui d'*information search*. Nous avons choisi le terme *information retrieval* car il exprime mieux notre conception de la troisième couche du modèle imbriqué de Wilson. D'abord, il est important de souligner qu'il y a une nette distinction entre ces trois concepts selon les termes utilisés en anglais pour les décrire, ce qui n'est pas le cas dans la traduction française où tous ces concepts se traduisent par « recherche d'information. » Notre première préoccupation sera donc de traduire ces termes en français en les adaptant pour que l'on puisse les distinguer.

Nous ne sommes pas les premiers à souligner cette ambiguïté dans la traduction de ces trois termes. Certains auteurs français ont essayé de distinguer ces termes les uns des autres et ont proposé des traductions qui représentent mieux ces concepts. Parmi eux, nous trouvons J. Maniez, M. Ihadjadene et S. Chaudiron. Selon Maniez, le terme *information retrieval* pourrait être remplacé par le terme système de « retrouvage ». En fait, selon lui, les SRI peuvent être considérés comme des sous-ensembles de systèmes de recherche d'objets (exemples d'objets : renseignement, outil, service, produit de consommation etc.). Maniez montre que « la particularité des SRI tient au caractère imprécis de la notion de « thème » ou de *topic* en anglais qui est employée dans le traitement documentaire pour ensuite définir la pertinence de l'adéquation entre un document et la requête » (Ihadjadene et Chaudiron 2008, p.185 ; Maniez 2002, p.132-140).

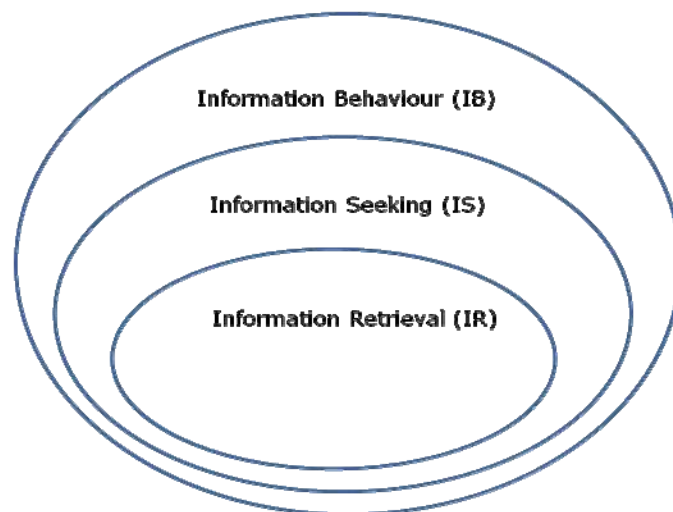


Figure 3.1 : Modèle imbriqué du comportement informationnel, adapté de (Wilson 1999)

L'*information seeking* est un concept qui concerne les méthodes employées par l'utilisateur pour découvrir les sources d'information ou pour y avoir accès (Wilson 1999). Elle désigne le

« contexte dans lequel se trouve un utilisateur au moment de choisir les sources d'information qu'il juge pertinentes pour répondre à son besoin informationnel » (Ihadjadene et Chaudiron 2008, p.185). Un utilisateur peut choisir d'utiliser un moteur de recherche ou bien une bibliothèque. Il peut aussi choisir de rechercher les informations auprès d'un ami, d'un professeur ou d'un journaliste. Quelle que soit la méthode choisie, ces activités font partie de l'*information seeking*. Ayant choisi une source d'information, par exemple un moteur de recherche, l'utilisateur commence un processus de formulation et de reformulation de requêtes. Ces requêtes, qui représentent son besoin informationnel, sont formulées tout en tenant compte des spécificités fonctionnelles du moteur de recherche ou du SRI qu'il utilise. Ce contexte, dans lequel l'utilisateur interroge un SRI en formulant des équations de recherche, est désigné sous le terme *information retrieval*.

Dans le cadre de ce travail de thèse, pour éviter les ambiguïtés, nous adoptons le terme d'« accès à l'information » utilisé par (Ihadjadene et Chaudiron 2008, p.185) pour désigner le contexte global de recherche d'information qui est appelé en anglais *information seeking*. En ce qui concerne le contexte spécifique lié à la formulation de requête et le « retrouvage » d'information utilisant un SRI, nous regroupons l'*information retrieval* et l'*information search* sous le terme traduit en français par « recherche d'information ». Nous adoptons le terme « comportement informationnel » pour désigner l'*information behaviour*.

Le comportement informationnel, qui est notre préoccupation majeure dans cette section, intègre toutes les pratiques informationnelles des utilisateurs recherchant de l'information, généralement dans des environnements principalement numériques. Il comporte également la recherche active et passive d'information (Wilson 1999) ainsi que l'évitement d'information (Case 2002, p.6-7). Il recouvre la totalité du comportement humain en relation avec des sources ou canaux d'information. Il inclut toutes les activités dans lesquelles s'engage une personne : identification des besoins informationnels, recherche d'informations (quelle que soit la manière dont la recherche est faite) et transmission des informations (Wilson 1999). Comme nous l'avons indiqué ci-dessus, il englobe l'accès à l'information et la recherche d'information.

### 3.2.1 L'accès à l'information

L'accès à l'information peut être considéré comme un processus d'élaboration du sens (*sense-making*) au travers duquel un individu développe son point de vue personnel par rapport à son besoin informationnel (Dervin 1986, p.20-22). L'individu assimile des informations provenant de plusieurs sources et les intègre dans sa structure cognitive. Les informations provenant de sources formelles et structurées, telles que des systèmes d'information, sont combinées avec d'autres informations de sources informelles, issues de l'expérience quotidienne de l'individu. Khulthau a proposé un modèle de recherche d'information à six étapes pour décrire le processus de l'accès à l'information. Chaque étape est caractérisée par des éléments cognitifs, affectifs et physiques (Khulthau 1991, p.366-368). En 2004, elle a ajouté la septième étape à ce modèle (Khulthau 2004, p.82). Ces étapes sont :

- initiation, la première phase du modèle de Khulthau décrit l'état de conscience d'un individu face à un manque de connaissance. Il essaie alors d'identifier son besoin d'information. Ce stade est généralement caractérisé par de l'incertitude et de l'appréhension. L'utilisateur essaie de comprendre son problème en le rapprochant de ses expériences antérieures et de ses connaissances personnelles ;
- sélection, l'utilisateur identifie et sélectionne les sujets généraux à rechercher et les approches à utiliser. Après la sélection, l'incertitude de l'utilisateur commence à diminuer, possiblement accompagné par un sentiment d'optimisme. Il est donc prêt à commencer la recherche ;
- exploration, cette phase correspond à la recherche des informations sur les sujets généraux sélectionnés. Pendant son déroulement, la compréhension de ces sujets augmente chez l'utilisateur, et les sentiments de confusion, d'incertitude et de doute peuvent rejaillir. Il intègre les informations contenues dans les documents retrouvés et les rapprochent à ce qu'il connaissait au préalable. A ce stade, le problème de corrélation entre les diverses informations se pose et il peut entraîner une certaine confusion, voir un sentiment de frustration chez l'utilisateur ;
- formulation, cette phase consiste à formuler un sujet précis de recherche. Au niveau cognitif, l'utilisateur identifie et sélectionne, dans les documents retrouvés des idées

appropriées qui lui permettent de former sa propre perspective du sujet de recherche. Durant cette étape, la confiance augmente et un sentiment de clarté s'installe ;

- **Collection**, c'est le stade où l'interaction entre l'utilisateur et le SRI devient effective et efficace. Les informations pertinentes sur le sujet sont collectées ;
- **Présentation**, l'utilisateur est maintenant doté des nouvelles connaissances qu'il peut appliquer pour résoudre son problème ou les communiquer aux autres ;
- **Évaluation**, c'est l'étape finale de la recherche. Elle permet une satisfaction finale de la part de l'utilisateur. L'utilisateur prend conscience de l'augmentation de son niveau de connaissance. L'identification d'informations complémentaires est également effectuée par l'utilisateur (Vivian et Dinet 2008, p.90).

Étapes :	Affectifs	Cognitifs	Physiques
<b>Initiation</b>	Incertitude	Analyse de la demande	Echange avec autrui
<b>Sélection</b>	Optimisme	Choix de mots-clés	Consultation de répertoires, discussions
<b>Exploration</b>	Doute, incertitude, confusion	Identification des sources	Lecture et prise de note
<b>Formulation</b>	Optimisme	Production d'idées, formulation de critères	Lecture de notes
<b>Collection</b>	Intérêt accru	Prélèvement d'informations	Prises de notes précises
<b>Présentation</b>	Optimisme	Catégorisation des informations	Organisation des notes prises
<b>Évaluation</b>	Satisfaction finale	Conscience de l'augmentation du niveau de connaissance, Identification d'informations complémentaires	Relecture finale et retour aux sources



Tableau 3.1 : Modèle de Khulthau sur le processus de recherche d'information<sup>54</sup> (Khulthau 1999, p.367 ; Khulthau 2004, p.82)

### 3.2.2 Le besoin informationnel

Dans la section 1.2.2.2, nous avons décrit l'information comme une différence entre deux états de connaissances (Brookes 1975, p.48 ; Mizzaro 1996, p.233-236). Le besoin informationnel peut donc être considéré comme un besoin pour combler le *gap* qui fait passer l'état de connaissance de l'utilisateur d'un niveau à un autre. Ceci montre qu'un besoin informationnel est un besoin cognitif<sup>55</sup>. Ce concept est donc beaucoup étudié parmi les chercheurs utilisant une approche cognitive pour analyser le comportement informationnel de l'utilisateur. Notons entre autres les travaux de Mackay, Taylor, O'Connor, Belkin et coll, Ingwersen et Mizzaro.

Mackay l'a expliqué par une perception incomplète du monde de la part de l'utilisateur. Ce dernier cherche à réduire cette inadéquation en interagissant avec son environnement dans un domaine d'intérêt particulier (Mackay 1960, p.789). O'Connor montre que le concept de besoin informationnel est d'une nature ambiguë (O'Connor 1968, p.200). Belkin et coll. ont parlé de l'état insatisfaisant de connaissance<sup>56</sup> (Belkin et coll. 1982a, p.61-71 ; Belkin et coll. 1982b, p. 145-164). Ingwersen a parlé de l'état de connaissance incomplet et de l'état d'incertitude de la connaissance (Ingwersen 1992, p.123-156). Ces deux derniers rejoignent à l'état insatisfaisant de connaissance de Belkin (1978, p. 81).

Ingwersen (1986, p.223) montre que le besoin informationnel peut se manifester sous trois formes : besoin vérificatif, besoin conscient thématique<sup>57</sup> et besoin « embrouillé »<sup>58</sup>. Taylor développe un modèle de besoin par étapes : besoin « viscéral », besoin conscient, besoin formalisé et besoin « compromis » (Taylor 1968, p.126-127). Suivant le même principe, Mizzaro (1998, p.308-310) identifie les quatre éléments suivants pour décrire la nature et l'expression du besoin informationnel d'un utilisateur :

---

<sup>54</sup> Cette table est une traduction de l'anglais reprise de (Vivian et Dinet 2008, p.90) avec quelques modifications ajoutées.

<sup>55</sup> Définition cognitif : PSYCHOL. APPL. Qui concerne les moyens et mécanismes d'acquisition des connaissances. Système cognitif, aptitude, dissonance, structure cognitive.

<sup>56</sup> « *Anomalous State of Knowledge* » généralement appelé *ASK*

<sup>57</sup> *Conscious topical need*

<sup>58</sup> *Muddled need*

- le besoin réel d'information (RIN)<sup>59</sup> est défini de la sorte : Un utilisateur confronté à un problème, cherche à le résoudre et cette volonté se manifeste par un besoin informationnel. Le RIN est associé à l'ensemble complet des informations qui permettraient à l'utilisateur, s'il les détenait, de résoudre son problème de manière appropriée. C'est donc l'abstraction du vrai besoin, comme l'indique Taylor qui, dans sa classification l'appelle besoin « viscéral ». Selon lui, ce vrai besoin n'est pas nécessairement exprimé de manière concrète dans l'esprit de l'utilisateur;
- le besoin perçu (PIN)<sup>60</sup> correspond au besoin « conscient » de Taylor. C'est une représentation implicite, dans l'esprit de l'utilisateur, du problème auquel il fait face. C'est une représentation mentale du RIN qui peut ne pas correspondre au RIN si l'utilisateur l'a mal perçu, d'où l'importance d'un intermédiaire (expert) pour l'aider à mieux comprendre son besoin ;
- la demande (EIN)<sup>61</sup> correspond au besoin informationnel perçu;
- la requête est une formalisation de la demande (EIN) dans le langage du SRI à utiliser.

Notons que le « besoin formalisé » défini par Taylor correspond à « la demande » de Mizzaro car il représente également l'expression du besoin informationnel de l'utilisateur sans tenir compte de spécificité du SRI. Le « besoin compromis » correspond à « la requête ».

Ces quatre (entités) composantes (RIN, PIN, EIN et requête) et les trois opérations (perception, expression et formalisation) qui les lient sont représentées graphiquement dans la figure 3.2. Ces quatre entités peuvent être ordonnées ainsi : requête  $\leq$  EIN  $\leq$  PIN  $\leq$  RIN selon la complétude de leur description et leur capacité à exprimer le besoin réel (Simonnot 2008, p.170).

Les SRI prétendent répondre aux besoins informationnels des utilisateurs en ne s'intéressant qu'à la requête. Notre opinion est que l'on ne peut pas satisfaire entièrement le besoin

<sup>59</sup> *Real Information Need (RIN)*

<sup>60</sup> *Perceived Information Need (PIN)*

<sup>61</sup> *Expressed Information Need (EIN)*

informationnel de l'utilisateur sans vraiment comprendre ces quatre éléments de son besoin. La requête (4<sup>ème</sup> élément du modèle de Mizzaro) correspondant au besoin réellement soumis au SRI, n'est pas suffisante pour représenter l'intégralité du besoin informationnel de l'utilisateur.

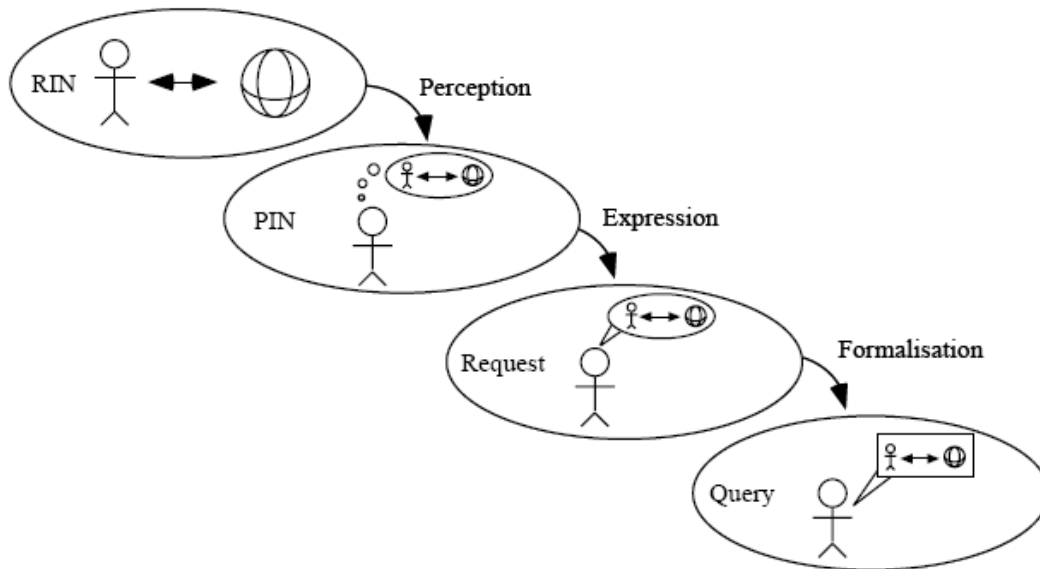


Figure 3.2 : RIN, PIN, demande et requête (Mizzaro 1998, p.308)

### 3.3 Système de recherche d'information

#### 3.3.1 Introduction

Nous traitons dans cette section le concept de SRI qui est le deuxième axe d'étude en recherche d'information (prise dans son sens global). Nous nous intéresserons à tout ce qui se rapporte à l'automatisation de la récupération d'information.

D'un point de vue historique, notons que l'augmentation du volume d'information scientifique (produite notamment depuis la fin de la 2<sup>ème</sup> guerre mondiale) ainsi que le développement de l'ordinateur, sont les principaux facteurs à l'origine de l'automatisation de la classification et des processus d'indexation de l'information. Dans ce contexte, en 1948, un événement majeur a eu lieu à Londres : La *Royal Society Scientific Information Conference*. Cette conférence, rassemblant des bibliothèques, des communautés scientifiques, avait pour

but d'améliorer les méthodes existantes de collection, d'indexation et de distribution de la littérature scientifique<sup>62</sup>.

Dans les années 1950, la recherche d'information devient un domaine de recherche ayant pour but d'améliorer l'accès à l'information et d'automatiser le processus de recherche d'information. Trois thématiques majeures ont été identifiées par Rijsbergen dans ce domaine, à savoir : l'analyse de contenu, la structure d'information et l'évaluation (Rijsbergen 1975).

L'analyse de contenu se rapporte à la description du contenu d'un document dans une forme permettant de le traiter informatiquement. Luhn (1957, p.309-317) a proposé d'utiliser les mots d'un document comme des termes d'indexation dans un système d'indexation automatique. Selon lui, la fréquence d'occurrence d'un mot dans un article fournit une mesure utile de sa signification ou de son importance. Il ajoute que l'emplacement relatif des mots, ayant valeur de signification dans une phrase, fournit aussi une mesure utile pour la détermination de la signification de la phrase. Ainsi le facteur de signification d'une phrase est calculé à partir de la combinaison des deux mesures<sup>63</sup> (Luhn 1958, p. 160). Cette idée de Luhn, est devenue un fondement sur lequel repose de nombreux systèmes d'analyse du texte. Un autre travail significatif sur l'indexation automatique est le projet SMART (Salton 1968 ; Salton 1971) qui a abouti au premier SRI expérimental.

La deuxième thématique concerne l'exploitation des relations entre les documents afin d'améliorer l'efficacité des stratégies de recherche. Elle couvre l'organisation logique du document et les méthodes de classification automatique. Good (1958 ; 1965, p.43-54) et Fairthorne (1961, p.1-10) furent parmi les premiers à proposer l'idée que la classification automatique puisse améliorer la recherche documentaire.

---

<sup>62</sup> The report of the Royal Society Empire Scientific Conference of 1946 asked the society "...to convene a conference of the libraries, societies and institutions responsible for publishing, abstracting and information services, in order to examine the possibility of improvement in existing methods of collection, indexing and distribution of scientific literature, and for the extension of existing abstracting services (The Royal Society Scientific Conference 21 June – 2 July 1948. Report and papers submitted. London: The Society).

<sup>63</sup> Selon Luhn, « the frequency of word occurrence in an article furnishes a useful measurement of word significance. It is further proposed that the relative position within a sentence of words having given values of significance furnish a useful measurement for determining the significance of sentences. The significance factor of a sentence will therefore be based on the combination of these two measurements » (Luhn 1958, p. 160).

La troisième thématique concerne l'évaluation des SRI. Elle inclut la mesure d'efficacité de la récupération de l'information. La plupart des travaux en R&D<sup>64</sup>, dans le domaine de la RI, ont pour but d'améliorer son efficacité. Les projets d'évaluation de systèmes d'indexation menés à Cranfields (Royaume-Uni) sous la direction de C. Cleverdon sont les premiers dans ce domaine (Cleverdon 1966 ; Cleverdon 1967).

Au cours des années 70 et 80, plusieurs travaux de recherche effectués en RI ont donné naissance à plusieurs modèles d'indexation et méthodes d'accès à l'information. Parmi les modèles proposés, deux d'entre eux sont parmi les plus utilisés : le modèle vectoriel<sup>65</sup> (Salton et coll. 1975, p. 613-620) et le modèle probabiliste<sup>66</sup> (Robertson et Sparck Jones 1976, p. 140-145) (cf. 3.3.2.)

En 1992, le gouvernement des Etats-Unis à travers l'institut national des standards et technologies (NIST)<sup>67</sup> a commencé des campagnes d'évaluation TREC<sup>68</sup> dans le but de promouvoir la recherche en RI. Les TREC fournissent aux groupes de recherche la capacité d'évaluer et de comparer leurs systèmes en utilisant un corpus de documents, de thèmes et de métriques communs. Cette méthodologie est devenue au fil des années et malgré les critiques, la méthodologie de référence. Elle s'est imposée sur tous les continents dans des programmes financés par des organismes officiels : citons notamment le programme japonais NTCIR<sup>69</sup> et le projet européen CLEF<sup>70</sup> (Ihadjadene et Chaudiron 2008, p.194). Avec l'introduction du Web au début des années 1990 a surgi le besoin de développer des approches et des techniques plus sophistiquées pour la recherche d'information (Foley 2008, p.11).

### 3.3.2 Le mode de fonctionnement d'un SRI

Traditionnellement, le fonctionnement du SRI comporte trois éléments principaux : la représentation de document, la formulation de requête et la fonction d'appariement. La représentation de document consiste à représenter son contenu par le processus d'indexation, soit automatiquement, soit manuellement par un indexeur. La formulation de requête est une

---

<sup>64</sup> Recherche et développement

<sup>65</sup> Vector Space Model

<sup>66</sup> Probabilistic Model

<sup>67</sup> *National Institute of Standard and Technology*

<sup>68</sup> *Text REtrieval Conference* (la conférence sur la recherche d'information textuelle)

<sup>69</sup> <http://research.nii.ac.jp/ntcir/>

<sup>70</sup> <http://www.clef-campaign.org>

représentation du besoin informationnel de l'utilisateur. Le lien entre les documents et les requêtes s'effectue par une fonction d'appariement. Cette fonction compare les deux types de représentation (documents et requêtes) et retrouve les documents qui répondent au besoin informationnel que l'utilisateur a exprimé dans sa requête. Ce modèle simple de SRI est représenté par la figure 3.3 ci-dessous.

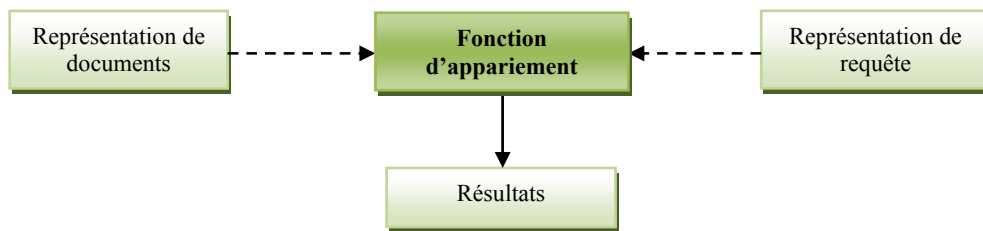


Figure 3.3 : Un modèle simple de la RI, adapté de (Ingwersen 1992, p49)

Rijsbergen (1979) décrit, au moyen d'une « boîte noire », un SRI typique. Son modèle, présenté en figure 3.4, est composé de trois parties : un *input* (entrée), un processeur et un *output* (sortie). L'*input* comporte les documents et les requêtes. Le processeur applique la fonction d'appariement tout en effectuant le processus de représentation et d'indexation des documents et des requêtes. Les résultats de recherche correspondent à l'*output*. Son modèle comporte un facteur de bouclage de pertinence (relevance feedback); cela implique que les résultats de recherche fournis par le SRI peuvent influencer l'utilisateur et le conduire à reformuler sa requête afin de mieux exprimer son besoin informationnel. Ceci correspond au point de vue cognitif de la RI parce que le bouclage de pertinence se traduit par le fait que la connaissance utilisateur du SRI augmente lorsqu'il reformule sa requête pour mieux représenter son besoin. Ceci permet également à l'utilisateur de mieux comprendre son besoin. Un schéma plus détaillé qui regroupe le processus de RI et une architecture générale de SRI est présenté dans la figure 3.5.

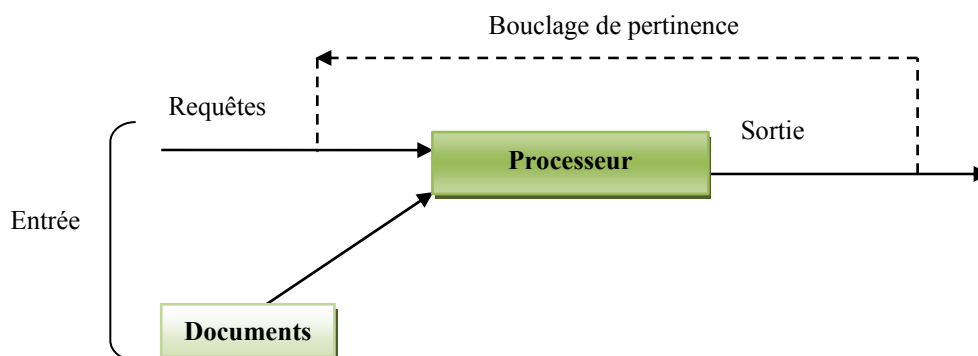


Figure 3.4 : Un SRI typique, adapté de (Rijsbergen 1979)

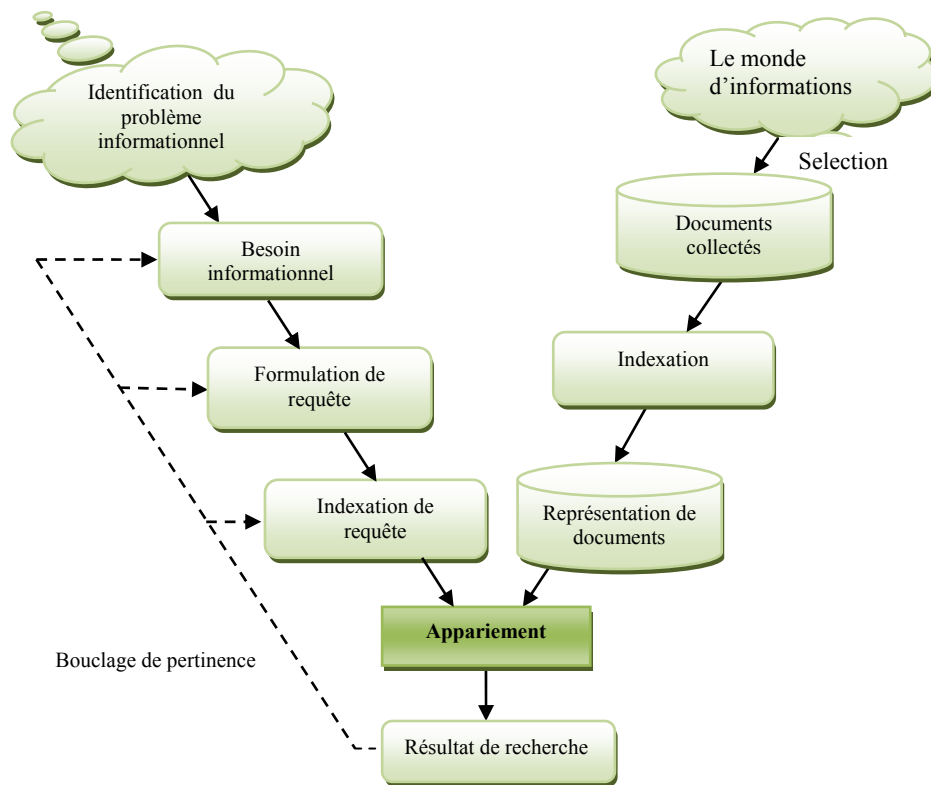


Figure 3.5 Un schéma global regroupant le processus de RI et une architecture générale de SRI

Aujourd'hui, la conception des SRI a beaucoup évolué. Les études sur le comportement informationnel de l'utilisateur et les besoins informationnels ont beaucoup influencé le développement et la conception des SRI. Sans trop entrer dans les détails, nous montrerons dans cette section et ses sous-sections la spécificité des SRI du point de vue de l'automatisation.

Indépendamment de sa conception, le SRI (un système de recherche bibliographique, un moteur de recherche en ligne, un moteur de recherche du bureau<sup>71</sup>) comporte toujours les trois éléments : requête, document et fonction d'appariement - indispensables à son fonctionnement. Ces trois éléments seront présentés par la suite.

<sup>71</sup> Desktop search engine

### **3.3.2.1 La collection, l'indexation et la classification des documents**

Dans cette section, nous présenterons les trois processus associés à l'élément « document » dans les SRI.

#### *3.3.2.1.1 La collection*

La première phase du développement d'un SRI est la collection de documents à intégrer dans le système. Elle est réalisée avant le développement du système mais le processus de collection peut continuer après la mise en route du SRI. Il existe plusieurs modes de collecte de documents selon le type de SRI développé. En ce qui concerne les moteurs de recherche en ligne, par exemple, la collection de documents se fait par le processus de *crawling*. C'est un processus qui permet à un robot d'indexation d'explorer le Web d'une manière automatique afin de collecter les documents (pages web, images, vidéos, documents Word, PDF ou PostScript, etc.).

#### *3.3.2.1.2 L'indexation*

Quel que soit le mode de collection des documents, dans la plupart des SRI, ces documents ne sont pas interrogés directement. Ils sont d'abord analysés et ensuite représentés sous forme de base documentaire. Cette analyse porte aussi bien sur le contenant que sur le contenu des documents (Afolabi 2007, p.138). Cette démarche d'analyse et de représentation des documents est effectuée afin d'améliorer l'efficacité du processus d'appariement lors de la recherche d'information. En procédant ainsi, la taille des données à interroger par la fonction d'appariement en temps réel lors d'une session de recherche d'information est réduite.

Le point de départ du processus d'analyse textuelle peut être un document entier, un résumé, un titre ou une liste des mots-clés. De ce point de départ, le processus d'analyse doit produire une représentation du document dans une forme capable d'être traitée informatiquement (Rijsbergen 1979).

L'indexation selon Kislin (2007, p.166) est « le résultat d'un traitement qui a pour but la création d'une représentation la plus fidèle et la plus exhaustive possible du contenu du document afin d'en faciliter les recherches et d'en garantir la pertinence future ». L'indexation manuelle implique un ou plusieurs indexeurs, chargés de faire le choix des descripteurs pour les valeurs des propriétés choisies pour la description. En ce qui concerne l'indexation automatique, le système est chargé d'extraire les concepts représentatifs du texte



d'un document textuel en se fondant sur les techniques linguistiques. Deux méthodes communes de prétraitement pour l'indexation automatique sont le *stopping* et le *stemming* (lemmatisation).

Le *stopping* est un processus qui permet d'enlever certains mots dits « vides »<sup>72</sup> ou anti-dictionnaire (par exemple les articles, les prépositions etc.) du texte d'un document. Une liste des mots vides généraux peut être retrouvée sur le Web et elle peut être, selon le domaine, étendue ou réduite. Le *stemming*, en revanche, sert à réduire des termes à leur racine commune. Par exemple, les mots collaborer, collaborant, collabore et collaboré peuvent être considérés comme des mots différents si la méthode de *stemming* n'est pas appliquée. Ces quatre mots, qui ont pour racine commune le verbe « collaborer », peuvent rendre un document non-pertinent si le document comporte une de ces conjugaisons et la requête en comporte une autre. Avec le *stemming*, ce problème est évité en réduisant ces quatre mots à leur racine commune.

Il est important de souligner qu'il existe aussi des travaux portant sur l'indexation automatique des documents multimédia. Ces travaux sont fondés sur des études de reconnaissance de forme. Enfin, une indexation humaine peut être combinée avec une indexation automatique dans le cadre d'une indexation assistée afin de pallier les manques d'indexation automatique.

### 3.3.2.1.3 La classification

Le troisième processus associé au traitement de documents dans les SRI est la classification. Elle est un processus de regroupement logique des documents. Ce regroupement peut être fait par sujet et ou par un autre critère sélectionné à la suite d'une analyse de contenu. La classification suit un ensemble de conventions, de méthodes et de règles de procédure structurées en système (CREPUQ<sup>73</sup> 1994, p.6). Elle peut être considérée comme une hiérarchisation des connaissances dans un domaine particulier.

Selon Rijsbergen (1979), les méthodes de classification, dans la RI, sont appliquées principalement dans le regroupement des mots-clés et dans celui des documents. Classer les

---

<sup>72</sup> *Stopwords*

<sup>73</sup> « *La gestion des archives informatiques* ». Conférence des recteurs et des principaux des universités du Québec (CREPUQ).

documents ou les mots-clés, correspond à les situer les uns par rapport aux autres. La classification a comme but d'améliorer l'efficacité du processus de recherche d'information. Le regroupement est fondé sur l'hypothèse que les documents ayant un contenu similaire ont la même pertinence vis-à-vis de la même requête. Selon R.M. Hayes, la classification est définie comme le regroupement des *items* (ex. documents, représentations des documents) qui sont par la suite considérés comme une unité, perdant ainsi leur identité individuelle. Ceci signifie qu'un document regroupé dans une classe est toujours identifié à cette classe. Tous les documents dans une classe sont traités comme identiques jusqu'à ce qu'ils soient examinés individuellement (Hayes 1963, p. 287)

Les techniques de classification peuvent être manuelles ou automatiques. Selon Chen et coll. (2002, p.2), la classification manuelle s'occupe de données variées et est très efficace en terme de précision. Ils citent comme exemples les sites de Yahoo, Amazon, et eBay dont les moteurs de recherche fonctionnent selon la classification manuelle. Ils remarquent que l'inconvénient de cette technique est liée à son inefficacité face aux objets très larges avec une structure de catégorisation complexe (Chen et coll. 2002, p.2).

La classification automatique est soit à la base de règles ou à la base de systèmes d'apprentissage. Deux types de classification découlent des systèmes d'apprentissage : la classification supervisée et la classification non-supervisée. La première correspond à une classification de document par un mécanisme externe (tel qu'un *feed-back* humain). Ce mécanisme fournit des informations sur la classification exacte des documents. Donnons quelques exemples de classificateurs utilisés pour la classification supervisée : réseaux neurones (Li et coll. 1991, p.313-318 ; Farkas 1994, p.710-713), *support vector machine* (Joachims 1998, p.138-140), *Kohonen type self-organizing maps* (Hyötyniemi 1996, p.64-67) etc. Certains de ces classificateurs peuvent être utilisés avec l'apprentissage non supervisé, c'est à dire, des documents non étiquetés, mais la précision d'un classificateur peut être améliorée en utilisant un petit ensemble de documents étiquetés. En revanche, la classification non-supervisée est faite sans aucune référence à une information externe. Elle est réalisée en calculant la mesure de rapprochement entre deux documents.

### **3.3.2.2 La représentation du besoin informationnel utilisateur sous forme de requêtes**

Le deuxième élément fonctionnel d'un SRI est la requête utilisateur qui est une expression de son besoin informationnel. La requête est présentée par l'utilisateur dans un langage compréhensible par le SRI. Cette étape constitue la première difficulté pour l'utilisateur. Pour que celui-ci exprime correctement son besoin informationnel en requête, il faut qu'il possède une bonne connaissance du fonctionnement du SRI. En fait, il faut que son modèle mental du SRI corresponde au modèle fonctionnel du système lui-même. Il est important de souligner à nouveau que le besoin informationnel utilisateur est plus profond qu'une simple requête formulée, ce qui explique pourquoi ce besoin informationnel est devenu un domaine d'étude pour des chercheurs en science de l'information.

Pour un SRI, la requête représente le besoin informationnel, qu'elle soit bien formulée ou non. Le système prend cette requête comme *une entrée*, et la traite. Dans certains SRI, la requête passe par le processus d'indexation afin d'en extraire une représentation qui est par la suite utilisée pour la mise en correspondance avec la représentation des documents. Après cette étape, le système fournit à l'utilisateur un ensemble de résultats s'il en existe. L'utilisateur peut évaluer les documents issus du résultat de recherche afin de juger leur pertinence. Cette évaluation peut être exprimée implicitement ou explicitement. Explicitement, l'utilisateur peut ajouter des annotations aux résultats de recherche. En revanche, une évaluation implicite peut se traduire par une reformulation de la requête utilisateur. La reformulation a lieu lorsque l'utilisateur a intégré l'information retrouvée dans sa structure cognitive.

### **3.3.2.3 La fonction d'appariement requête – document**

Le troisième élément du fonctionnement d'un SRI est la fonction d'appariement. C'est une fonction de mise en correspondance entre la représentation du contenu des documents et la requête de l'utilisateur. Elle représente le cœur du SRI et contribue à sa qualité.

Dans le cas simple de recherche purement manuelle d'un document, l'utilisateur est le seul juge de la pertinence du document retrouvé dans un ensemble de documents. On parle de pertinence utilisateur. En revanche, dans le cas d'un SRI, la fonction d'appariement intervient dans la détermination de la pertinence parce que l'ensemble de documents est indexé et qu'une requête est soumise par l'utilisateur. On parle alors de pertinence système.

La qualité d'un SRI dépend de l'écart existant entre la pertinence utilisateur et la pertinence système (Kislin 2007, p.168). La pertinence système dépend non seulement de l'algorithme de mise en correspondance entre les documents et les requêtes mais elle dépend aussi fortement des formalismes utilisés pour la description des documents et pour l'expression des requêtes. Si l'algorithme d'appariement est très efficace mais la représentation des documents n'est pas bien faite, le système aura du mal à retrouver des documents pertinents associés à la requête de l'utilisateur. Il en est de même quand le traitement des termes de la requête est inefficace.

En résumé, il existe deux fonctions importantes permettant de caractériser et d'identifier un modèle de SRI. La première correspond à la fonction de représentation et d'indexation des documents et des requêtes, et la deuxième, correspond à la fonction d'appariement.

### **3.3.3 Les modèles de système de recherche d'information**

Rappelons que l'objectif majeur de tout SRI est d'identifier des documents pertinents associés à une requête particulière dans l'ensemble des documents contenus dans la base documentaire. Comme évoqué dans la section précédente, les SRI peuvent se classer en différentes catégories selon la méthode qu'ils utilisent pour apparier les requêtes aux documents. Il existe plusieurs modèles de SRI qui ont été développés, mais nous n'évoquerons dans cette section que les trois modèles classiques qui s'appliquent dans certains modules de notre prototype présenté dans le sixième chapitre. Ces trois modèles sont : le modèle booléen (*Boolean model*), le modèle vectoriel (*Vector space model*) et le modèle probabiliste (*Probabilistic model*).

#### **3.3.3.1 Le modèle booléen**

Le modèle booléen utilise la théorie des ensembles et l'algèbre de Boole. Il est probablement le plus répandu dans le domaine de la recherche d'information. Les documents ainsi que les requêtes des utilisateurs sont considérés comme des ensembles de termes. Ce modèle utilise des fichiers inverses pour l'indexation des documents. Prenons l'exemple d'un SRI dont la base est constituée des documents  $d_1, d_2, d_3, \dots, d_n$ . Pour chaque document se trouvant dans la base, les termes  $t_1, t_2, t_3, \dots, t_m$  sont utilisés pour le décrire.

On a donc :

$d_j(t_1, t_2, t_3, \dots, t_m)$  où terme  $t_i$  décrit le document  $d_j$

Pour la liste inverse, qui permet d'associer à chaque terme l'ensemble des documents qu'il décrit, on a :

$$t_i (d_1, d_2, d_3, \dots d_l)$$

Une formulation de requête est faite par une combinaison des termes et des opérateurs booléens tels que la conjonction (ET), la disjonction (OU) et la négation (SAUF). Les termes sont séparés les uns des autres par ces opérateurs. Pour traiter une requête booléenne, la requête est d'abord décomposée, et, pour chaque terme, la base documentaire est interrogée pour faire ressortir un ensemble de documents contenant ce terme. Les opérateurs booléens sont par la suite appliqués sur les ensembles retrouvés pour chaque terme afin de créer un ensemble final de documents qui corresponde à l'équation de recherche formulée (Foley 2008, p.16).

La figure 3.6 montre un exemple de requête booléenne. La requête est exprimée par (new ET york) SAUF city. Cette requête retrouvera les documents qui contiennent à la fois le terme « new » et le terme « york » mais qui ne contiennent pas le terme « city ».

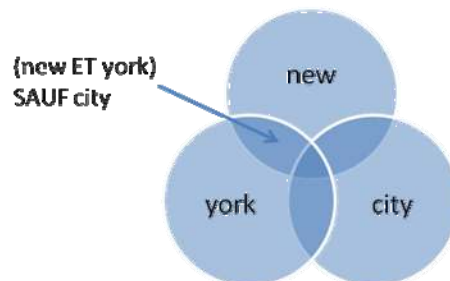


Figure 3.6 : Un exemple de requête booléenne

Une requête booléenne permet à l'utilisateur d'exprimer précisément son besoin informationnel. C'est un point fort du modèle booléen. La précision de ce formalisme fait que ce modèle utilisant les requêtes booléennes a été largement utilisé par les « intermédiaires formés » (bibliothécaires) à l'époque des débuts de la recherche d'information (Foley 2008, p.17). Le modèle booléen s'adapte très bien à la recherche en texte intégral. Il utilise l'indexation des fichiers inverses rapidement actualisés, tout en permettant l'intégration en continu de nouveaux documents (Kislin 2007, p.169 ; Belkin et Croft 1987, p.113-114).

Bien que ce modèle soit très répandu, il est très critiqué car il est fondé sur un critère de décision binaire. Par conséquent, si les documents ne correspondent pas exactement à l'équation de recherche, ils sont jugés comme non pertinents et écartés. On peut ainsi se retrouver dans la situation où le nombre de documents retrouvés est très petit. A l'inverse, on peut également avoir un nombre de documents retrouvés très grand.

Une autre faiblesse de ce modèle est l'absence de fonction de pondération pour différencier les termes de la requête. Ainsi, les résultats ne sont pas ordonnés en fonction de leur degré de pertinence.

Pour le rendre plus « flexible », le modèle booléen a été étendu, (Salton 1983, p.1023) permettant ainsi l'utilisation de l'opérateur de proximité (« *near* ») des caractères jokers, des troncatures, des expressions régulières (*regular expressions*) dans la formulation de la requête. Le modèle étendu permet aussi d'intégrer une aide au choix des descripteurs par l'intermédiaire de listes contrôlées, d'ontologies (Loiseau et coll. 2004, p.29-33) ou de thesaurus (Kislin 2007, p.170).

### 3.3.3.2 Le modèle vectoriel

Le besoin de pondérer les termes dans les documents et le classement des documents ont conduit au développement des modèles de recherche d'information « *best-match* ». Dans cette catégorie, le modèle vectoriel (*Vector Space Model*) proposé par Salton (1975, p.613) emploie une approche statistique pour caractériser quantitativement les termes dans les documents et mesurer la similarité d'un document vis-à-vis d'une requête. Le modèle représente les documents et les requêtes par des vecteurs dans un espace à  $n$  dimensions où à chaque terme descripteur correspond une dimension de l'espace et l'ensemble  $T$  des termes d'indexation constitue la base de l'espace (David 1999, p.25). Le modèle mesure la « similarité » entre deux vecteurs, qu'il s'agisse de deux vecteurs de requêtes, de deux vecteurs de documents, ou encore d'un vecteur de document et d'un vecteur de requête.

Dans ce modèle, un document  $D_j$  peut être représenté par :

$$D_j = w_{j1}t_1, w_{j2}t_2, \dots, w_{jn}t_n$$

où  $w_{ji}$  représente le poids de pertinence affecté au terme  $t_i$  du document  $D_j$

et une requête  $Q$  peut être représentée par :

$$Q = w_1t_1, w_2t_2, \dots, w_nt_n$$

où  $w_i$  représente le degré d'importance du terme  $t_i$  dans la requête

Comme les documents et les requêtes partagent le même type de représentation, l'ordre d'un document  $D_j$  vis-à-vis d'une requête  $Q$  est déterminé par le calcul de la similarité entre le document et la requête, soit  $\text{sim}(D_j, Q)$ . La mesure de similarité souvent utilisée est le cosinus de l'angle formé par le vecteur de document et le vecteur de requête :

$$\text{sim}(D_j, Q) = \cos(D_j, Q) = \frac{d_j \cdot q}{\|d_j\| \times \|q\|} = \frac{\sum_{i=0}^{I-1} w_q(i) \cdot w_d(j, i)}{\sqrt{\left(\sum_{i=0}^{I-1} w_q(i)^2\right) \left(\sum_{i=0}^{I-1} w_d(j, i)^2\right)}} \quad (3.1)$$

où

$\cos(D_j, Q)$  représente la mesure de similarité entre le document  $D_j$  et la requête  $Q$  ;

$d_j$  représente le vecteur du document  $D_j$  ;

$q$  représente le vecteur de la requête  $Q$  ;

$w_q(i)$  représente le degré d'importance du terme  $i$  dans la requête ;

$w_d(j, i)$  représente le poids de pertinence du terme  $i$  dans le document  $j$ .

Ce modèle offre de nombreux avantages : l'augmentation des performances système grâce à la pondération des termes ; les documents répondant approximativement à une requête peuvent être fournis comme résultats de recherche; les documents constituant le résultat de recherche peuvent être triés selon leur degré de similarité avec la requête. Comme les documents et les requêtes sont représentés en tant que vecteurs dans un même espace vectoriel, ce modèle permet d'implémenter les techniques de bouclage de pertinence et d'expansion de requête.

Mais le modèle vectoriel n'est pas sans défaut. Ses limitations se manifestent dans :

- La pondération des documents longs : la pertinence d'un document doit être indépendante de sa taille. Néanmoins, sans un moyen de normalisation de la taille des documents,

les documents longs paraîtront avoir une fréquence de terme plus élevée que les autres, à cause de leur taille. Une solution a été proposée dans (Salton et Buckley 1988, p.516-521 ; Singhal et coll. 1996, p. 21-29).

- La synonymie et la polysémie : l'une des limitations majeures de ce modèle concerne son incapacité de gérer la synonymie et la polysémie. Deux mots différents peuvent signifier la même chose dans le cas de la synonymie. En revanche, un mot peut avoir plusieurs sens quand il s'agit de la polysémie. Des documents non-pertinents peuvent avoir des valeurs de similarité élevées parce qu'ils contiennent des termes qui se trouvent dans la requête mais qui n'ont pas les mêmes significations. En revanche, des documents pertinents à une requête peuvent avoir des valeurs de similarité basse parce que les termes utilisés dans la requête ne sont pas trouvés dans les documents alors que leurs synonymes y sont contenus. Une solution à ce problème a été proposée dans l'indexation sémantique latente<sup>74</sup> (Bestgen 2004, p.1 ; Deerwester 1990, p.391-407).

### **3.3.3.3 Le modèle probabiliste**

Le fonctionnement des SRI peut être interprété du point de vue probabiliste. On rencontre souvent, dans la littérature sur la recherche d'information, l'expression : « les documents qui peuvent répondre au besoin informationnel de l'utilisateur ». Cette phrase indique qu'il existe une notion de probabilité dans le jugement de la pertinence d'un document par rapport au besoin informationnel de l'utilisateur. Cette notion intervient à deux niveaux :

- le niveau de « pertinence système » par lequel le système juge certains documents comme pertinents à la requête de l'utilisateur
- le niveau de « pertinence utilisateur » par lequel un utilisateur juge certains documents comme pertinents à son besoin informationnel

Quel que soit le niveau considéré, une première question se pose : Quelle est la probabilité qu'un document D (ou sa représentation) réponde de manière pertinente à une requête Q ?

---

<sup>74</sup> L'indexation sémantique latente est une technique utilisée pour l'analyse automatique du langage. Elle « vise à construire un espace sémantique de très grande dimension à partir de l'analyse statistique de l'ensemble des cooccurrences dans un corpus de textes » Bestgen (2004, p.1).



Une seconde question se pose : Quelle est la probabilité que les termes utilisés par l'utilisateur pour formuler sa requête correspondent aux termes utilisés pour représenter les documents ?

La première question peut être considérée comme un fondement sur lequel repose le modèle probabiliste. L'approche probabiliste fut présentée premièrement par Maron et Kuhns (1960). Elle a été par la suite élaborée, testée et appliquée par d'autres comme Robertson et Spärck Jones (1976, p.140-145), Van Rijsbegen (1979), Croft et Tuttle (1979) et Fuhr (1992, p. 243-255). Le modèle probabiliste calcule la probabilité qu'un document ( $D$ ) soit pertinent ( $R$ ) pour une requête ( $Q$ ) donnée. Cette valeur, désignée par  $P(R|D)$ , sert à classer les documents selon leur ordre de pertinence vis-à-vis de la requête, ce qui constitue un résultat de recherche en soi. Il s'agit du « principe de classement probabiliste<sup>75</sup> » décrit par Robertson en 1977. Une explication a été donnée par (Kislin 2007, p.172) comme suit : « Etant donnée une requête de l'utilisateur, il existe un ensemble des documents qui contient exactement les documents « *pertinents* » que nous pourrions qualifier d'ensemble idéal de réponses. Si nous connaissons la description de cet ensemble idéal, nous n'aurons alors aucune difficulté à retrouver les documents qui les composent. Répondre à une requête revient donc à spécifier les propriétés de cet ensemble. Cependant ces propriétés ne sont pas connues au moment de la requête, il faut d'abord *deviner* ce qu'elles pourraient être. Cette première tentative permet de générer une description probabiliste initiale de l'ensemble, qui est ensuite utilisée pour retrouver un ensemble primitif de documents. Il faut ensuite une interaction avec l'utilisateur pour améliorer la description probabiliste de l'ensemble idéal (ou plutôt de l'échantillon représentant cet ensemble idéal ».

Robertson et Spärck Jones (1976, p.141) ont appliqué le théorème de Baye (Baye's theorem) et *log-odds* (Robertson 1977, p.282) sur le principe de classement ( $P(R|D)$ ) pour arriver à une fonction de classement optimale :

$$\text{sim}(D,Q) = \log \frac{P(D|R)}{P(D|\bar{R})} \quad (3.2)$$

où

---

<sup>75</sup> Probability Ranking Principle (Robertson 1977, p.281) : « *if a reference retrieval system's response to each request is a ranking of the documents in the collections in order of decreasing probability of relevance to the user who submitted the request, where the probabilities are estimated as accurately as possible on the basis of whatever data have been made available to the system for this purpose, then the overall effectiveness of the system to its users will be the best that is obtainable on the basis of those data* »

$P(D|R)$  est la probabilité qu'un document (D) soit pertinent (R)

$P(D|\bar{R})$  est la probabilité qu'un document est non-pertinent ( $\bar{R}$ )

Selon Foley (2008, p.22), c'est à ce point que les différents modèles probabilistes, fondés sur leur hypothèse sous-jacente, divergent. Selon lui, le *Binary Independence Model* proposé par Robertson et Spärk Jones (1976) est la forme la plus simple du modèle probabiliste. Il calcule la probabilité de pertinence d'un document en simplifiant  $P(D|R)$  en termes de ces attributs (i.e. termes). Les termes de document et de requête sont représentés par un poids binaire avec une hypothèse d'indépendance entre les termes afin d'arriver à :

$$\text{sim}(D,Q) = \sum_{t_i \in Q,D} \log \frac{p_i(1-q_i)}{q_i(1-p_i)} \quad (3.3)$$

où

$p$  est la probabilité qu'un document soit pertinent étant donné qu'il contient un terme  $t_i$ ,

$P(t_i|R)$

$q$  est la probabilité qu'un document est non-pertinent étant donné qu'il contient un terme  $t_i$ ,

$P(t_i|\bar{R})$

Nous pouvons représenter  $p$  et  $q$  par les proportions suivantes:

$$p = r_i/R \quad (3.4)$$

$$q = \frac{n_i - r_i}{N - R} \quad (3.5)$$

où

$N$  est le nombre de documents dans l'ensemble

$n_i$  est le nombre de documents qui contiennent le terme  $i$

$R$  est le nombre de documents pertinents connu dans l'ensemble

$r_i$  est le nombre de documents pertinents connu qui contiennent le terme  $i$

En substituant les valeurs de  $p$  et de  $q$  dans les équations 3.4 et 3.5 dans l'équation 3.3, nous avons une formule de pondération de pertinence (Robertson 1976) :

$$r_w = \log \frac{\left(\frac{r_i}{R}\right)\left(1 - \frac{n_i - r_i}{N - R}\right)}{\left(\frac{n_i - r_i}{N - R}\right)\left(1 - \frac{r_i}{R}\right)} \quad (3.6)$$

Puisque l'ensemble de documents pertinents  $R$  n'est pas connu au départ de la recherche, il nous faut une estimation pour  $p$  et  $q$ . La méthode la plus utilisée est de remettre  $r_i$  et  $R$  à zéro et puis ajouter des constantes à chacune de ses valeurs dans l'équation 3.6 (Robertson et Spärk Jones 1976).

Reprenons l'explication donnée par Foley (2008, p.21-25) : le *Binary Independence Model* a été implémenté par l'équipe de *City University* au TREC-1 (Robertson et coll. 1992, p.21-30). Le résultat était néanmoins mauvais car le modèle n'a pas intégré la fréquence de termes dans le document ainsi que la normalisation de la longueur de document. Du coup, deux modèles de classement sont introduits au TREC-2. Les modèles BM15 et BM11 furent introduits afin d'expérimenter avec différentes fonctions de pondération en intégrant la fréquence de terme et la normalisation de la longueur de document (Robertson et coll. 1993, p.21-34). Les deux fonctions de pondération ont été combinées pour obtenir la fonction de pondération BM25 qui a été expérimentée au TREC-3 (Robertson et coll. 1992, p.109-126). Ce dernier est devenu le modèle probabiliste le plus utilisé pour la recherche d'information. Il est représenté par l'équation 3.7

$$w(i,j) = \log \frac{\left(\frac{r_i + 0.5}{R + 0.5}\right)\left(1 - \frac{n_i - r_i + 0.5}{N - R + 0.5}\right)}{\left(\frac{n_i - r_i + 0.5}{N - R + 0.5}\right)\left(1 - \frac{r_i + 0.5}{R + 0.5}\right)} \times \frac{tf(i,j) \times (k+1)}{k+1 \times ((1-b) + (b \times ndl_j)) + tf(i,j)} \quad (3.7)$$

où

$N$  est le nombre de documents dans l'ensemble

$n_i$  est le nombre de documents qui contiennent le terme  $i$

$R$  est le nombre de documents pertinents connu dans l'ensemble

$r_i$  est le nombre de documents pertinents connu qui contiennent le terme  $i$

$tf_{i,j}$  est la mesure de fréquence de terme  $i$  dans le document  $j$

$ndl_j$  est la longueur normalisée (dl) du document  $j$

$$= \frac{dl}{\text{dl moyen de tout documents}}$$

$kI$  est le constant qui détermine l'influence de  $tf_{i,j}$

$b$  est le constant qui détermine l'influence de la normalisation de la longueur de document

Les trois modèles que nous venons de décrire ont été développés dans le but d'aider l'utilisateur à trouver les documents **pertinents** à sa requête. Il nous paraît donc, important de traiter la notion de pertinence dans la recherche d'information car elle occupe une place très importante dans les études sur les SRI. Cette notion fera l'objet de la section suivante.

### 3.3.4 La notion de pertinence en recherche d'information

Comme nous l'avons dit dans la section 3.3.2.3, la pertinence peut être vue selon la perspective utilisateur ou la perspective système. Calcaly et coll. (2004, p.181) définissent la pertinence comme « l'adéquation entre le résultat d'une recherche d'informations et l'énoncé de la question posée ». Selon Dinet et Rouet (1998, p.153), elle est « un jugement de valeur attribué selon le niveau de lecture du sujet, ses connaissances concernant le thème, ses objectifs, le contexte ». Brigitte Simonot (2008, p.162-163) a tenté d'expliquer la notion de pertinence selon une distinction entre deux mots en anglais - *relevance* et *pertinence* - qui relèvent de cette notion. Selon elle, le terme anglais *relevance*, par son étymologie, est apparenté au verbe « relever » au sens de « être du ressort de » mais aussi « mettre en relief ». Elle ajoute : « la *relevance* serait ce qui, pour une question donnée, permet de distinguer un objet informationnel des autres ». Le terme *pertinence* en revanche vient du latin *pertinere* qui est composé du verbe *tenere* qui signifie « se souvenir de quelque chose » mais aussi « comprendre, conserver dans son esprit » et du préfixe *per* qui signifie « à travers, au moyen de ». Ses explications sur ces deux termes sont résumées ainsi : « La *relevance* concerne l'adéquation d'un document ou d'objet informationnel à un besoin d'information ou à une demande d'information en général » et la *pertinence*, en revanche, concerne « l'adéquation d'un document ou d'un objet informationnel à la demande d'un individu donné : pour qu'un document *relevant* soit pertinent pour cet individu, il doit être compréhensible par lui et lié aux connaissances dont il dispose sur le sujet » (Simonot 2008, p.163).

Nous pouvons rapprocher l'explication de *relevance* selon Brigitte Simonot à ce que nous considérons comme la « pertinence système » et son explication de *pertinence* à ce que nous considérons comme la « pertinence utilisateur ». Plusieurs études ont été faites sur la notion

de pertinence en recherche d'information qui en effet a abouti à des modèles de la pertinence.

Parmi ces modèles, nous trouvons :

- le modèle d'interaction stratifié de Saracevic qui sort d'une analyse des manifestations de la pertinence pour en distinguer les différents attributs (Saracevic 1996, p.201-218 ; Saracevic 1997, p.315-320). Ce modèle a été revisité par Cosijn et Ingwersen (2000, p.537) ;
- le modèle de Mizarro intégrant les différentes relations de pertinence (Mizarro 1998, p.307-315) ;
- le modèle cognitif de Wang et Soergel (1998, p.116-117) qui traite le processus de prise de décision des utilisateurs lorsqu'ils choisissent ou rejettent un document proposé par un système de recherche d'information.

Nous sommes intéressés par le modèle de Saracevic dans lequel il présente une série de pertinences. Saracevic a abordé le sujet de pertinence en expliquant l'interaction homme-machine dans la recherche d'information. Les éléments majeurs dans son modèle sont l'utilisateur et le système qui sont caractérisés par de nombreux attributs. Il propose que l'interaction entre l'utilisateur et le SRI puisse être modélisé comme une suite de processus se produisant à plusieurs niveaux ou strates interconnectés (Saracevic 1996, p.212). Il définit alors cinq catégories de pertinence :

- la pertinence système ou algorithmique : il s'agit de l'évaluation de la fonction d'appariement d'un SRI. L'algorithme d'appariement est chargé de comparer une requête de l'utilisateur avec les représentations des documents dans sa base afin de fournir à l'utilisateur des documents que le système juge pertinents à la requête. Cette adéquation entre les documents et les requêtes constitue la pertinence système ou algorithmique
- la pertinence-sujet ou thématique (*topicality*) : il s'agit d'une relation entre le sujet des documents et celui de la requête. On suppose que les documents et les requêtes comportent des sujets. Cette pertinence est donc définie en fonction d'une somme de correspondance ou d'une absence de correspondance entre le sujet de la requête et le sujet couvert par des objets informationnels retrouvés.
- la pertinence cognitive : c'est la relation entre l'état de connaissance et le besoin informationnel cognitif de l'utilisateur et les objets informationnels retrouvés. Elle est

inférée à partir de la correspondance entre les objets informationnels et les connaissances de l'utilisateur, de la nouveauté, et de la qualité de l'information.

- la pertinence situationnelle ou utilité : elle se réfère à la relation entre la tâche ou le problème à résoudre et les objets informationnels retrouvés. Elle est inférée à partir de leur utilité à la prise de décision, de l'adéquation de l'information pour résoudre le problème donné, et de la réduction d'incertitude.
- la pertinence affective ou motivationnelle : elle se réfère à la relation entre les intentions, les buts et les motivations de l'utilisateur et les objets informationnels retrouvés. Elle est inférée de la satisfaction, du succès, de la réussite et des goûts de l'utilisateur.

Nous pouvons également rapprocher ces catégories de pertinence aux deux perspectives mentionnées ci-dessus. La pertinence système ou algorithmique correspond à ce que nous avons précédemment appelé la « pertinence système ». La pertinence-sujet ou thématique en revanche se trouve entre la pertinence système et la pertinence utilisateur. Les trois dernières catégories sont strictement liées avec l'utilisateur donc nous les considérons comme des sous-éléments de la pertinence utilisateur.

### **3.3.5 Les limitations de SRI classique<sup>76</sup>**

Une problématique toujours présente dans le domaine de la recherche d'information est la pertinence d'une information retrouvée par rapport au besoin informationnel de l'utilisateur. Le problème qui demeure est de rapprocher la pertinence système de la pertinence utilisateur. La pertinence système, que nous considérons comme une pertinence objective, est différente de la pertinence utilisateur, que nous considérons comme une pertinence subjective. En effet, l'utilisateur ne juge pas la pertinence d'un document en regardant uniquement la similarité existant entre la représentation du document et sa requête. Pour juger, l'utilisateur fait intervenir d'autres facteurs comme le montre le modèle de Saracevic (cf.3.3.4). Un SRI, en revanche, fournira toujours le même résultat à une requête donnée si la base documentaire ne change pas. Si on prend le cas de deux utilisateurs différents qui formulent deux requêtes identiques, le fait qu'elles soient identiques n'implique pas que les utilisateurs aient le même

---

<sup>76</sup> Nous entendons par SRI classique ou traditionnel, tout SRI qui n'intègre pas les connaissances sur les utilisateurs dans son mode de fonctionnement et qui ne s'adapte pas à ses utilisateurs.

besoin informationnel. De ce fait, leurs jugements (appréciations) de la pertinence des documents retrouvés par le SRI seront différents.

Le but d'un SRI étant de fournir à l'utilisateur des documents répondant à son besoin informationnel, il est nécessaire qu'une réflexion sur ce problème de pertinence soit engagée et influence la conception et le développement des méthodes et des outils de recherche d'information. Le problème de pertinence comporte plusieurs aspects, qui sont :

- a) le décalage entre le besoin informationnel de l'utilisateur et sa représentation en termes de requête : les études sur le besoin informationnel de l'utilisateur ont montré qu'il est parfois difficile pour un utilisateur d'exprimer son besoin dans un langage compréhensible par le SRI. En effet, il est naturel pour un utilisateur d'exprimer son besoin informationnel dans un langage naturel pour lequel la plupart des SRI ne dispose pas de moyens d'interprétation. Même quand les SRI sont présentés comme capables d'interpréter ce langage naturel, il existe un fossé sémantique. Prenons un exemple, donné par (David 1999, p.23), de deux besoins informationnels :

- « je cherche les derniers livres de Mitterrand »
- « je cherche un bon livre sur la modélisation de l'utilisateur »

Comment peut-on exprimer « les derniers livres » et « un bon livre » dans une équation de recherche (requête) ? Même si un attribut « appréciation » est utilisé pour représenter les livres, l'affectation de la valeur « bon » dépend de l'indexeur et non de l'auteur ou de l'utilisateur. Aujourd'hui, on trouve des moteurs de recherche d'information qui suggèrent des requêtes aux utilisateurs à partir du premier mot entré par l'utilisateur. Même si cela peut s'avérer utile, on remarque malheureusement qu'une erreur de représentation du besoin dans le premier mot ne peut mener qu'à des suggestions des requêtes erronées.

- b) la différence entre les termes utilisés pour l'indexation des documents dans le SRI et les termes utilisés par l'utilisateur pour exprimer son besoin informationnel : comme nous l'avons expliqué sur la pertinence-sujet (cf. 3.3.4) ou thématique, s'il existe une différence entre le sujet couvert par les documents (exprimé dans leurs représentations,) et le sujet de la requête de l'utilisateur, ceci aboutira à une absence de correspondance. Plusieurs auteurs ont évoqué ce problème, sachant que de nos jours,

cela devient toujours plus difficile d'aboutir à un consensus sur la représentation des concepts.

- c) l'incapacité de « capter » la sémantique de la requête utilisateur est aussi un facteur qui contribue au problème de pertinence dans les SRI.
  
- d) le décalage entre le contexte de production d'informations et le contexte d'utilisation par les utilisateurs : comme nous l'avons montré dans le premier chapitre (cf. 1.2.2.3), le sens d'une information est étroitement lié au contexte dans lequel elle est produite. En revanche, la pertinence d'une information est strictement liée au contexte dans lequel elle sera utilisée. Ce décalage peut contribuer à accentuer les problèmes liés à la pertinence, car la plupart du temps, ce ne sont pas les producteurs d'informations qui les utilisent.
  
- e) le décalage entre le modèle fonctionnel du SRI et le modèle mental que l'utilisateur en possède : il s'agit ici d'une corrélation entre les connaissances dont dispose l'utilisateur sur le fonctionnement du SRI et le mode de fonctionnement réel du SRI. Si un utilisateur possède un modèle mental inexact du fonctionnement du SRI, ceci se traduira dans son interaction avec le SRI et contribuera au problème de pertinence. Par exemple, la plupart des utilisateurs des moteurs de recherche en ligne manque de connaissance sur la façon dont ces derniers fonctionnent. Par exemple, ils ne savent pas utiliser la fonctionnalité de « recherche avancée ». Aujourd'hui, les moteurs de recherche en ligne, comme Google, disposent de beaucoup de fonctionnalités inexploitées par les utilisateurs, à cause du décalage dans leur modèle mental du système.
  
- f) l'écart entre le comportement informationnel de l'utilisateur et la représentation de l'utilisateur dans le SRI : nous expliquerons ce point dans le quatrième chapitre car nous n'avons pas encore expliqué la représentation de l'utilisateur dans les SRI.

### **3.4 L'intégration de la représentation de l'utilisateur dans les SRI**

Dans les SRI traditionnels, les particularités des utilisateurs ne sont pas prises en compte car tous les utilisateurs sont considérés de manière identique. Mais la réalité est différente car



chaque utilisateur a ses particularités cognitives et affectives. Puisque le jugement de pertinence par l'utilisateur constitue un critère déterminant pour l'efficacité d'un SRI, il est nécessaire d'intégrer la représentation de l'utilisateur dans le SRI. Depuis la fin des années 1980, on constate que la représentation des connaissances sur l'utilisateur appelée « modèle de l'utilisateur » est intégrée dans la conception et le développement des SRI. Cette intégration des particularités de l'utilisateur montre que les chercheurs se sont servis des idées développées dans les deux axes de recherche en RI (le processus de RI et le SRI) afin de satisfaire au mieux les utilisateurs des SRI.

Les spécificités des utilisateurs sont nécessaires dans toutes les phases du processus de recherche d'information car le système a besoin de connaître le niveau cognitif de l'utilisateur et ses préférences pour interpréter son besoin, ce qui implique un intérêt pour la « compréhension » du besoin et non le deviner. Pour modifier la requête initiale de l'utilisateur, le système a besoin de connaître son évaluation des solutions déjà proposées. L'implantation du modèle de l'utilisateur dans un SRI permet, de ce fait, de personnaliser ses résultats de recherche tout en s'adaptant au mieux à ses connaissances, à ses préférences, à ses goûts et à son besoin.

En général, il existe 3 catégories de modélisation de l'utilisateur :

- la modélisation implicite, qui est fondée sur les interactions de l'utilisateur avec le SRI (Rich 1983, p.201). Les préférences de l'utilisateur sont implicitement inférées à travers ses activités. Le fait de cliquer sur un document, ou de faire défiler l'écran en consultant un document ou d'imprimer un document ou le temps passé sur un document peuvent être utilisés implicitement pour inférer le besoin utilisateur ;
- la modélisation explicite de l'utilisateur, où on demande explicitement à l'utilisateur de spécifier ses préférences et d'évaluer les résultats de recherche fournis par le SRI. Cette approche s'intéresse à modéliser les connaissances (savoir, savoir-faire) de l'utilisateur et son processus cognitif ;
- la modélisation hybride de l'utilisateur, qui est une combinaison des approches implicites et explicites.

Giorgio et coll. (1990, p.168) ont identifié quatre dimensions de la modélisation de l'utilisateur :

- *implicite*  
*versus explicite* : ils expliquent cette dimension d'une manière différente de celle que nous avons explicitée ci-dessus. Ils considèrent un modèle de l'utilisateur comme étant implicite si : 1) il a été développé lors de la conception du système ; 2) il est utilisé pour faire des choix techniques sur la structure et l'organisation du système, mais sans avoir été représenté explicitement dans le système final. Un modèle implicite selon eux, ne peut pas être modifié ou mis à jour sans faire recours à la modification intense de la structure et de l'organisation du système dans laquelle il a été implanté. En revanche, ils considèrent un modèle de l'utilisateur comme explicite s'il est directement représenté dans le système en tant que connaissances « séparées » mais utilisables en interne pour le fonctionnement du système et accessibles en externe pour d'éventuelles modifications et mises à jour ;
- *exprimé*  
*(given) ou inféré (inferred)* : un modèle explicite peut être classifié comme « exprimé » s'il a été défini, codé lors de la conception du système puis stocké pour être utilisé lors du fonctionnement du système. Ce type de modèle peut être modifié ou mis à jour seulement par l'intervention du concepteur du système et lorsque le système est mis hors opération. En revanche, si un modèle explicite de l'utilisateur est automatiquement intégré lors du fonctionnement du système sans l'intervention du concepteur, un tel modèle est considéré comme inféré ;
- *statique*  
*versus dynamique* : un modèle est dit « statique » s'il est conçu pour rester fixe et ne pas changer dans le temps. A l'opposé, un modèle conçu pour être raffiné, étendu et mis à jour progressivement lors du fonctionnement du système – et ceci afin d'intégrer de nouvelles connaissances sur l'utilisateur – est considéré comme « dynamique » ;
- *canonique*  
*versus individuel* : un modèle canonique vise à « capter » les caractéristiques, les traits communs à un ensemble d'utilisateurs, tout en ignorant les particularités individuelles. En revanche, un modèle prenant en compte la représentation de tous les traits de

chaque individu dans une population donnée, c'est-à-dire à la fois ses traits spécifiques et les traits communs avec d'autres utilisateurs, est dit « individuel ».

Dans la réalité, pour rendre compte de la complexité liée à la modélisation de l'utilisateur, ces dimensions peuvent être combinées afin de donner une bonne représentation. Avant l'automatisation des SRI, le traitement humain par les bibliothécaires ou les intermédiaires, aidant l'utilisateur à bien définir son besoin informationnel en intégrant ses particularités et son contexte, permettait de réduire fortement cette complexité. Par exemple, si un professeur d'école primaire recherche un bon livre sur la science, la bibliothécaire ne va pas l'orienter sur la même sélection de livres qu'elle ferait pour un professeur d'université faisant la même demande.

L'implantation du modèle de l'utilisateur s'applique à la fois au SRI et au système de filtrage d'information (SFI). Expliquons la différence entre SRI et SFI : le SRI requiert une activité de formulation systématique du besoin de l'utilisateur tandis que le SFI pérennise ce besoin d'information et permet l'acheminement au cours du temps des documents correspondant à cette expression de besoin. Ainsi le SRI permet la découverte ponctuelle de documents alors que le SFI permet la réception permanente de documents (Berrut et Denos 2003, p.242-244). Alors que le SRI a pour but de fournir à l'utilisateur les documents en se fondant sur son contexte actuel (c'est-à-dire une session unique de recherche), le SFI, en revanche, achemine des documents qui se présentent vers un utilisateur ou un groupe d'utilisateurs, en se fondant sur son profil<sup>77</sup> à long terme.

### **3.4.1 Système de recherche d'information personnalisé**

L'implantation d'un modèle individuel de l'utilisateur dans un SRI permet de personnaliser les réponses de recherche. Bueno et coll. (2007, p.2) ont proposé une approche de

---

<sup>77</sup> Il est nécessaire de distinguer le « profil de l'utilisateur » du « modèle de l'utilisateur ». En général le profil de l'utilisateur s'applique dans le SFI tandis que le modèle de l'utilisateur s'applique dans le SRI. Selon le dictionnaire de l'information par Cacaly et coll. (2004, p.186), le profil est une « caractérisation des sujets d'intérêt d'une personne ou d'un groupe de personnes. L'enregistrement de cette caractérisation permet d'organiser une diffusion sélective et ciblée de l'information ou le développement de prestations de type push. Cela peut consister en une équation de recherche mémorisée avec l'adresse du destinataire et utilisée périodiquement pour signaler les nouveautés ». Le modèle de l'utilisateur, en revanche, peut consister de deux parties : une partie statique et une partie dynamique. Alors que la partie statique correspond aux informations personnelles de l'utilisateur et toutes les autres informations sur lui qui ne changent pas dans le temps, la partie dynamique, en revanche, correspond au contexte de l'utilisateur à chaque session de recherche. Le contexte dans ce cas comporte les requêtes, l'objectif et les préférences de l'utilisateur.

personnalisation fondée sur l'objectif de recherche de l'utilisateur. Ils proposent deux représentations de l'utilisateur : une représentation plus ou moins statique, correspondant à son profil, et une représentation dynamique, renouvelée lorsque l'utilisateur exprime un nouveau besoin informationnel. Leur système permet à l'utilisateur d'exprimer son objectif de recherche en langage naturel afin d'éviter le problème liée à l'expression de besoin informationnel en langage du SRI. Dans le cadre de leur travail, une session est constituée d'un objectif de recherche avec toutes les interactions associées. Les interactions comportent les requêtes formulées, les documents consultés et les évaluations émises pour répondre à un besoin informationnel donné. Ces auteurs utilisent deux algorithmes de personnalisation : NBM (Naives Bayes Metiore) et WNB (Weighted Naives Bayes Metiore) dans leur prototype pour « prédire » les évaluations de l'utilisateur pour chaque document, selon son objectif de recherche actuel. Ceci permet de fournir à l'utilisateur les résultats de recherche personnalisés fondés sur son objectif, ses requêtes et ses évaluations des documents déjà consultés.

L'approche de Vallet et coll. (2006, p.2) pour la personnalisation des réponses des SRI montre une évolution dans la conception des SRI. Vallet et coll. expriment l'idée que les préférences de l'utilisateur risquent de ne pas être pertinentes dans toutes les situations. Ils sont de l'opinion que la multiplicité, l'hétérogénéité et la variation (fluctuation) voire la contradiction caractérisant les préférences humaines devraient être interprétées dans leur contexte, en prenant en compte le but de l'utilisateur et de sa tâche actuelle. Ils ont ainsi proposé une méthode pour le développement d'une représentation dynamique du contexte sémantique de la tâche de recherche au cours de son exécution. Cette représentation est ensuite utilisée pour activer les sous-ensembles des préférences de l'utilisateur, de manière à éliminer les préférences hors contexte. Leur approche est fondée sur une représentation orientée ontologique du domaine de recherche fournissant ainsi des descriptions enrichies des éléments sémantiques dans les actions de recherche et dans les préférences de l'utilisateur (voir la figure 3.7). Ceci permet de rapprocher les préférences et le contexte. L'analyse des relations sémantiques explicites entre les préférences et le contexte peut apporter une amélioration au niveau de la personnalisation des réponses par rapport aux techniques statistiques utilisés pour calculer la similarité entre les termes.

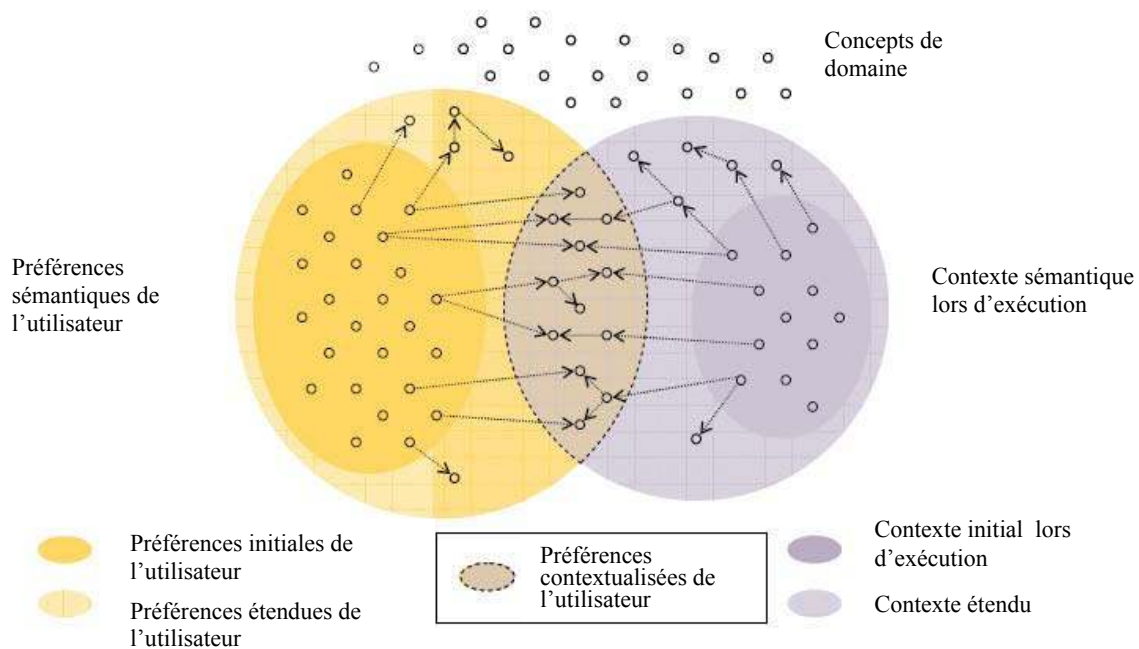


Figure 3.7 : Activation contextuelle des préférences sémantiques de l'utilisateur (Vallet et coll. 2006, p.4)

Une approche plus récente dans la personnalisation est présentée dans (Maghrebi 2010, p.223-238). Maghrebi propose d'intégrer le contexte d'utilisation d'un document dans la représentation du document. Elle propose également une modélisation dynamique du besoin de l'utilisateur par le contexte d'utilisation. Elle combine une modélisation statique avec une modélisation dynamique de l'utilisateur. Cette approche permet une personnalisation des réponses, fondée sur le contexte dans lequel les réponses de recherche seront utilisées.

### 3.5 La recherche sociale d'information

L'étude du comportement informationnel utilisateur montre que la recherche d'information est une activité à la fois cognitive et sociale. L'utilisateur a toujours tendance à consulter d'autres utilisateurs dans sa quête informationnelle. La recherche sociale d'information (RSI) est donc née du fait que les utilisateurs peuvent s'influencer réciproquement, de manière implicite ou explicite, dans les démarches qu'ils entreprennent pour résoudre leurs problèmes informationnels. La RSI est à l'intersection des concepts de la recherche d'information et des réseaux sociaux. Elle comporte toutes les techniques permettant à un utilisateur d'avoir accès

aux informations susceptibles de répondre à son besoin informationnel, tout en exploitant les connaissances d'autres utilisateurs. Parmi ces techniques se trouvent l'indexation sociale, le *bookmarking* social, le filtrage collaboratif, l'analyse des réseaux sociaux et le partage de requête.

Selon Kirsch (2005, p.34) un système de recherche sociale d'information (SRSI) comporte trois éléments : document, requête, individu. Ce qui domine les systèmes de RI traditionnels est la modélisation des relations entre les documents et les requêtes. Les SRSI, de leur côté, intègrent en plus la modélisation des relations entre les individus (voir les figures 3.8 et 3.9), ce qui fait leur force. Dans les SRSI, l'individu peut jouer le rôle de producteur de l'information ou celui d'utilisateur de l'information, et nous prendrons garde à bien faire cette distinction.

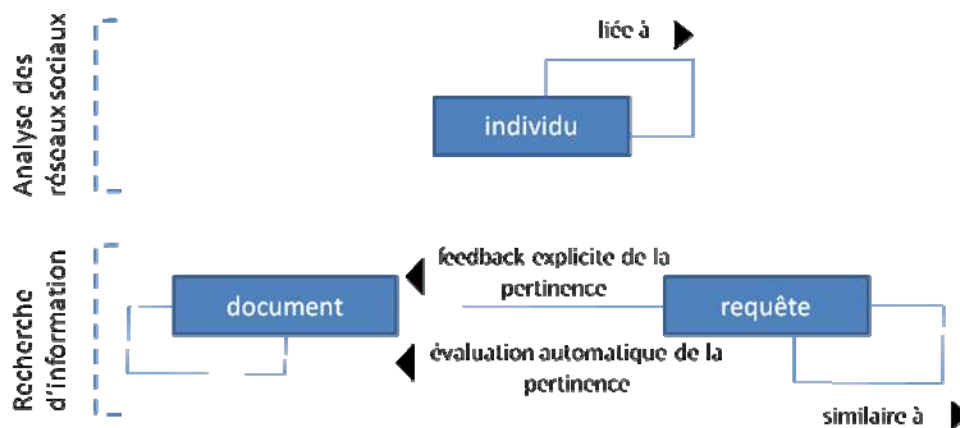


Figure 3.8 : RI traditionnelle et analyse des réseaux sociaux (Kirsch 2005, p.35)

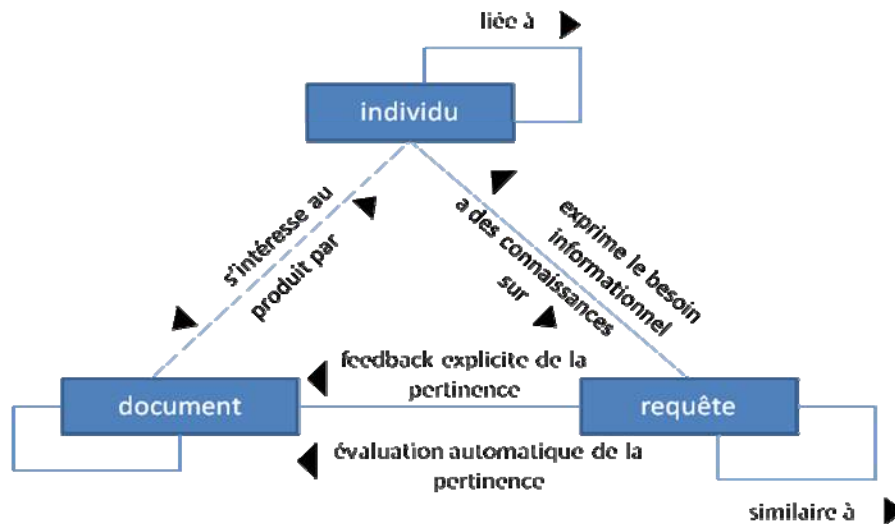


Figure 3.9 : *Modèle de domaine* pour la RSI (Kirsch 2005, p.35)

La RSI est fondée sur le fait que le producteur d'une information est inséparable de son produit (Kirsch 2005, p.34), et que nous « utilisons des gens pour trouver des contenus et des contenus pour trouver des gens<sup>78</sup> ». Une information est toujours produite par un producteur, elle est l'expression de la pensée d'un individu. Cela signifie qu'un jugement porté sur une information pourrait être utilisé pour établir un jugement sur le producteur de telle information. De la même façon, un jugement ou une évaluation sur le producteur d'une information peut être influencé par le jugement de l'utilisateur sur telle information. Par exemple, dans le domaine des publications scientifiques : quand on cherche un article sur un sujet, s'il existe un auteur célèbre et reconnu comme étant un expert dans le domaine correspondant, il y a une tendance à considérer ses articles comme des documents pertinents.

Il est important d'intégrer l'analyse des réseaux sociaux dans la recherche d'information. Prenons un autre exemple : comment jugeons-nous l'autorité d'un auteur sur un sujet par rapport à celle d'autres auteurs travaillant sur le même sujet au sein d'une communauté de recherche. Imaginons que le résultat de recherche d'une requête formulée sur une base de publications fasse ressortir sept articles écrits par des auteurs différents. Mais, en analysant un peu plus, on découvrirait que l'un des auteurs a déjà collaboré avec cinq autres parmi les six autres. On est alors influencé par la réputation de cet auteur dans la communauté, et on est

<sup>78</sup> « *We use people to find content and use content to find people* ». <http://semanticstudios.com/publications/semantics/000006.php>, site visité le 17 Mars, 2009.

tenté de le considérer comme une autorité sur le sujet et, en conséquence, à probablement consulter son article en premier<sup>79</sup>.

Selon Kirsch (2005, p.36-37), le web sémantique<sup>80</sup> illustre l'importance de la recherche sociale d'information car il permet la modélisation explicite des trois éléments de la RSI. Cette modélisation utilise des standards, et nous en choisissons trois pour poursuivre notre explication : le standard FOAF, le standard Dublin Core et le standard OpenSearch. Le standard FOAF permet de modéliser les individus en utilisant un type <foaf : person> et de modéliser les relations entre eux en utilisant une propriété <foaf : knows>. Les intérêts ou les préférences d'un individu sont modélisés par la propriété <foaf : interest>. Le standard Dublin Core, de sa part, permet de modéliser le lien entre un document et son auteur en utilisant <dc : creator> et de modéliser les relations entre les documents par la propriété <dc : relation>. La propriété <dc : subject> contient le sujet et les mots-clés d'un document. Le standard OpenSearch, en revanche, permet de publier les informations sur les résultats de recherche. La combinaison de ces trois standards permet une description à la fois complète et lisible à la machine (*machine-readable*) des trois éléments de la RSI (Kirsch 2005, p.37).

### 3.5.1 Le filtrage collaboratif

Le filtrage collaboratif est une autre technique qui s'inscrit dans la recherche sociale d'information. Il regroupe les méthodes pour le développement du système de recommandation<sup>81</sup> en se fondant sur les préférences et les apports de la communauté d'utilisateurs du système. Le filtrage collaboratif permet de contourner certaines difficultés liées au système de filtrage par le contenu et au système de recherche d'information. L'idée est non seulement de personnaliser la réponse mais d'utiliser ce que d'autres ont déjà trouvé et évalué. « Le principe est de filtrer le flot de documents entrant en fonction de l'opinion que d'autres utilisateurs de la communauté ont déjà portée sur les documents » (Berrut et Denos 2003, p.247). Cette hypothèse suppose que ce qui est déjà jugé pertinent par des utilisateurs

---

<sup>79</sup> Une idée similaire est exprimée par Kirsch (2005, p.36).

<sup>80</sup> Le web sémantique est un concept qui « désigne un ensemble de technologies visant à rendre le contenu des ressources du World Wide Web accessible et utilisable par les programmes et agents logiciels, grâce à un système de métadonnées formelles, utilisant notamment la famille de langages développés par le W3C » (Wikipedia, consulté le 23 Juillet 2010)

<sup>81</sup> « Les systèmes de recommandation sont une forme spécifique de filtrage de l'information (SI) visant à présenter les éléments d'information (films, musique, livres, news, images, pages Web, etc) qui sont susceptibles d'intéresser l'utilisateur. Généralement, un système de recommandation permet de comparer le profil d'un utilisateur à certaines caractéristiques de référence, et cherche à prédire l'« avis » que donnerait un utilisateur » (Wikipedia, consulté le 23 Juillet, 2010).



pourrait être pertinent pour un autre utilisateur ayant un problème informationnel similaire. Le reproche que l'on peut formuler sur cette hypothèse est le suivant : à tout moment, un utilisateur peut porter un jugement erroné sur un document et ainsi influencer de manière inadéquate le jugement d'autres utilisateurs qui consulteraient ce document.

Dans les systèmes de filtrage collaboratif, les méthodes statistiques sont utilisées pour faire des prévisions basées sur des intérêts des utilisateurs. Ces prévisions sont à leur tour exploitées pour faire des propositions à un utilisateur individuel, en se fondant sur la corrélation entre son propre profil personnel et les profils d'autres utilisateurs qui présentent des intérêts et des goûts semblables (Berrut et Denos 2003, p.247). Pour construire leur profil, les utilisateurs fournissent des évaluations sur les objets informationnels. Les évaluations peuvent prendre la forme de notes (scalaires) ou une forme binaire telle que d'accord / pas d'accord ou une forme unaire telle qu'un enregistrement ou une trace qui montre qu'un utilisateur a choisi un *item* ou a réalisé un achat (dans le domaine du commerce) (Schafer et coll. 2007, p.293). L'évaluation peut prendre également une forme textuelle. Les évaluations des utilisateurs sont comparées pour mesurer leurs similitudes. Des prévisions sont calculées sous forme de moyennes pondérées, à partir des évaluations d'autres utilisateurs ayant des goûts semblables ou complètement opposés (Berrut et Denos 2003, p.247).

Selon (Herlocker et coll. 1999, p.1), le filtrage collaboratif possède les avantages suivants sur le filtrage par le contenu :

- le filtrage d'objets informationnels dont le contenu est difficile à analyser (par exemple : musique, vidéo...);
- le filtrage fondé sur l'évaluation de certains aspects des objets informationnels (ex : leur qualité...) et fondé sur le goût de l'utilisateur ;
- la génération de recommandations sérendipiteuses<sup>82</sup> : le filtrage collaboratif peut permettre de recommander aux utilisateurs des items intéressants mais qui ne correspondent pas forcément aux attentes exprimées.

---

<sup>82</sup> Le mot « sérendipiteux » est la forme adjectivale du mot « sérendipité ». La sérendipité est le fait de découvrir quelque chose par accident et sagacité alors que l'on est à la recherche de quelque chose d'autre (*accident and sagacity wh ile in pursuit o fso mething else* ) (H. Walpole 1754). Source : [http://www.intelligence-creative.com/354\\_serendipite\\_definition.html](http://www.intelligence-creative.com/354_serendipite_definition.html), site visité le 28 Août 2009.

Les difficultés issues de la synonymie et de la polysémie rencontrées par les SRI sont surmontées par des « lecteurs humains ». Ainsi les objets informationnels sémantiquement similaires mais ne partageant pas les mêmes termes pourraient être recommandés grâce au filtrage collaboratif en se fondant sur les évaluations humaines de tels objets.

(Maltz et Ehrlich 1995, p.203) a mis en évidence une autre motivation en faveur du filtrage collaboratif en comparant la richesse de l'environnement des objets réels à la pauvreté dans laquelle les utilisateurs des systèmes opèrent. Lorsqu'un utilisateur consulte un document électronique, il ne peut, en général, d'aucune manière savoir s'il est le premier utilisateur à consulter le document ou si d'autres l'ont déjà consulté avant lui. L'un des apports du filtrage collaboratif est d'associer aux documents électroniques les activités qui ont été fait sur eux. Hill et coll. (1992, p.6-7), de leur côté, ont remarqué que les objets de la vie quotidienne accumulent à l'usage l'usure qui est un indicateur de leur utilisation : les pages des livres se corrent, les reliures se plient, et les marges sont tachées par des empreintes digitales. Les objets plus usés peuvent être vus comme les plus utilisés, ainsi leur usure sert d'*index* pour trouver les informations pertinentes dans l'objet. Nous pourrions alors dire que la possibilité de donner aux utilisateurs l'accès à l'histoire de l'usage d'un objet peut leur permettre de profiter des conseils subtils généralement employés pour prendre des décisions de lecture ou de non lecture dans le monde réel (Berrut et Denos 2003, p.249).

Pour qu'un système de filtrage collaboratif soit efficace, deux conditions sont nécessaires : un grand nombre d'utilisateurs et une évaluation, au moins, de chaque document du système. Des problèmes apparaissent alors pour les nouveaux documents. En effet, ils ne peuvent être diffusés que lorsqu'un minimum d'informations les concernant est collecté à partir de l'évaluation d'au moins l'un des utilisateurs. La limitation majeure donc, du système de filtrage collaboratif est liée au démarrage à froid. Les nouveaux utilisateurs commencent toujours avec un profil vide. Même s'il existe un profil de démarrage, le système a besoin d'une période d'apprentissage avant que le profil ne reflète correctement les préférences de l'utilisateur (Berrut et Denos 2003, p.254). En d'autres termes, le système ne peut pas filtrer des objets efficacement pour le compte d'un utilisateur pendant qu'il « apprend » sur lui. D'un autre côté, les personnes ayant des goûts peu fréquents risquent de ne pas recevoir de propositions. Ces carences peuvent être résolues en employant une approche hybride dans laquelle un système de filtrage collaboratif cohabite avec un système de filtrage par contenu.

Le premier système de filtrage collaboratif est le système Tapestry (Goldberg et coll. 1992, p.61-70) développé au centre de recherche Xerox Palo Alto (PARC). Il permet aux utilisateurs d'annoter les objets informationnels en texte libre ou de donner des appréciations dans le style « J'ai bien aimé » ou « Je déteste ». Ainsi les utilisateurs peuvent se recommander des documents les uns aux autres. D'autres systèmes de filtrage collaboratif parmi les premières générations sont Grouplens (Konstan et coll. 1997, p.77-87), PHOAKS (People Helping One Another Know Stuff) (Terveen et coll. 1997, p.59-62), Siteseer (Rucker et Polanco 1997, p.73-75) et Fab (Balabanovic et Shoham 1997, p.66-72) (ce dernier combine le filtrage par contenu et le filtrage collaboratif.)

Parmi les systèmes commerciaux en ligne qui emploient la technique du filtrage collaboratif, nous trouvons :

- Amazon<sup>83</sup> qui est notamment connu pour la recommandation des livres ;
- Amie Street<sup>84</sup> qui permet aux utilisateurs de recommander des musiques. D'un autre côté, il leur permet de découvrir des musiques qui peuvent les intéresser en leurs fournissant des recommandations selon leurs préférences. Il permet également un système collectif de calcul de prix d'une musique ;
- eBay<sup>85</sup> qui est un site d'e-commerce de ventes aux enchères permettant aux utilisateurs de vendre et d'acheter des biens. Il recommande aux utilisateurs des *items* selon leur profil ;
- Google News<sup>86</sup> qui est un site de journaux en ligne, rassemble des articles provenant de plus de 4,500 sources et regroupe les informations similaires pour les afficher en fonctions des intérêts de chaque utilisateur (Das et coll. 2007, p.2) ;
- Baynote<sup>87</sup> qui est une plateforme d'intelligence collective<sup>88</sup> permettant aux entreprises de déployer des recommandations « orientées visiteurs », de faire de la recherche sociale d'information et de fournir d'autres services à travers des canaux multiples tout en offrant un seul point d'entrée, d'essai et de contrôle. Baynote réclame faire 4 milliards de recommandations chaque mois ;

---

<sup>83</sup> <http://www.amazon.com>

<sup>84</sup> <http://amiestreet.com>

<sup>85</sup> <http://www.ebay.com>

<sup>86</sup> <http://news.google.com/>

<sup>87</sup> <http://www.baynote.com/>

<sup>88</sup> L'intelligence collective renvoie à une intelligence partagée ou à l'intelligence du groupe ou à l'intelligence des connexions et des relations qui résultent d'une coopération ou d'une compétition de plusieurs individus (Zara 2004, p.8)

- ChoiceStream<sup>89</sup> qui est un service du genre SAS<sup>90</sup> fournit des recommandations sur des produits ainsi que des contenus informationnels aux consommateurs (utilisateurs). ChoiceStream utilise une approche probabiliste pour la personnalisation tout en combinant un ensemble de techniques statistiques telles que le filtrage collaboratif, les tableaux de corrélation multiple, l'analyse par cohorte, la corrélation des attributs, le filtrage sélectif et le *multiple term scoring* (ChoiceStream 2009, p.16-19).

### 3.5.2 L'indexation sociale

Une autre technique associée à la recherche sociale d'information est l'indexation sociale. Elle est un système de classification dite collaborative qui permet aux non-spécialistes<sup>91</sup> d'ajouter des métadonnées à un document. Elle est également désignée par la folksonomie, le *social bookmarking* et l'étiquetage collaboratif (Golder et Huberman 2006, p.198). L'expression officielle française pour la folksonomie est « l'indexation personnelle<sup>92</sup> ». A l'inverse des systèmes hiérarchiques de classification où la contrainte de terminologie est toujours présente, l'indexation sociale permet aux utilisateurs ou contributeurs d'utiliser les termes qu'ils souhaitent pour décrire un document. Ces termes, souvent appelés mots-clés ou étiquettes, sont librement choisis. L'indexation sociale est caractérisée par des vocabulaires non normalisés, non contrôlés, non structurés et non spécifiques à un domaine.

Le processus d'indexation sociale prend de l'ampleur aujourd'hui et on le trouve partout dans les systèmes permettant aux utilisateurs de partager des contenus informationnels. Elle permet de créer des favoris et de leur ajouter des étiquettes servant de « couche d'indexation » ou de classification pour les liens afin de les partager avec d'autres utilisateurs.

Des exemples très connus de sites qui supportent l'indexation sociale incluent :

---

<sup>89</sup> <http://www.choicestream.com/>

<sup>90</sup> SAS (Software As Service), prononcé comme *SaaS* et appelé « logiciel en tant que service » en français, est un concept consistant à proposer un abonnement à un logiciel plutôt que l'achat d'une licence. Avec le développement des technologies de l'information et de la communication, de plus en plus d'offres *SaaS* se font au travers du web. Il n'y a alors plus besoin d'installer une application de bureau ou client-serveur. Ce concept, apparu au début des années 2000, prend la suite de celui du fournisseur de service d'application (« *application service provider* » - ASP) [Wikipédia : consulté le 5 mars, 2010].

<sup>91</sup> Les non-spécialistes correspondent à tous ceux qui ne sont pas des indexeurs professionnels

<sup>92</sup> <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000021530619&dateTexte=&categorieLien=id>

- delicio (<http://del.icio.us>), reddit (<http://reddit.com>), furl (<http://furl.net>), et Digg (<http://digg.net>) pour le bookmarking ;
- flickr (<http://flickr.com>) pour les photos ;
- CiteULike (<http://citeulike.org>), BibSonomy (<http://bibsonomy.org>), Connotea (<http://conotea.org>) et LibraryThing (<http://librarything.com>) qui permettent aux utilisateurs de gérer et partager les métadonnées bibliographiques sur le web (*Social reference managing* ou *collaborative cataloguing*) (Voss 2007, p.234).

L'indexation sociale ne doit pas être confondue avec la taxonomie ou l'ontologie. Patrick Lambe a bien fait de distinguer ces trois concepts à l'aide de sa carte de *findability*<sup>93</sup> (voir la figure 3.10). Sur la carte, remarquons que la taxonomie et l'ontologie ont des frontières bien définies, à la différence du monde des métadonnées, d'où la présence de l'océan du chaos. Une autre interprétation de cette carte consiste à dire que les mondes de la taxonomie et de l'ontologie sont entourés par l'océan de métadonnées dans lequel se trouvent l'île de la folksonomie et l'océan du chaos. Ceci veut dire que l'indexation sociale (folksonomie) peut être utilisée pour pallier les lacunes du système de classification et du système d'ontologie mais que cela peut engendrer du chaos car il n'y a pas de contrôle dans les termes utilisés pour 'tagger' les documents. D'un autre côté, l'indexation sociale peut fournir plus de termes d'indexation que l'indexation humaine contrôlée faite par des indexeurs professionnels ou l'indexation automatique faite par des machines.

Les avantages associés à l'indexation sociale sont nombreux :

- elle est faite par des humains ;
- elle est instantanée, cumulative, simple, rapide et gratuite ;
- elle permet l'application de la technique de filtrage collaboratif ;
- tous les genres de documents sont couverts (texte, audio, vidéo, image etc...).

En revanche, il existe beaucoup d'inconvénients associés à l'indexation sociale :

- elle engendre trop d'hétérogénéité ;

---

<sup>93</sup> [http://www.greenchameleon.com/gc/blog\\_detail/the\\_kingdom\\_of\\_taxonomy/](http://www.greenchameleon.com/gc/blog_detail/the_kingdom_of_taxonomy/)

- elle engendre une confusion entre la pertinence et la popularité. Le fait qu'un objet soit plus taggé que les autres dans un système n'implique pas qu'il soit plus pertinent que les autres ;
- elle peut aussi introduire des problèmes liés à la polysémie, à la synonymie, à la lemmatisation, à la redondance de langue et à l'orthographe ;
- il existe un risque que les termes employés ne décrivent pas correctement l'objet ;
- elle engendre trop de bruit.



Figure 3.10 : Une carte de différenciation entre la taxonomie, l'ontologie et la folksonomie<sup>94</sup>

### 3.6 Conclusion du troisième chapitre

Nous avons élaboré dans ce chapitre le concept de la RI et les évolutions qui ont eu lieu dans ce domaine. Nous avons également évoqué le besoin d'intégrer la dimension sociale dans la RI et certaines techniques existantes supportant cette dimension ont été développées. Toutes ces évolutions avaient pour but d'améliorer le processus de recherche d'information de l'utilisateur mais malheureusement nous constatons que l'on ne peut pas tout confier à la machine surtout tout ce qui touche l'aspect cognitif et social de la RI. Les utilisateurs ont

<sup>94</sup> [http://www.greenchameleon.com/gc/blog\\_detail/the\\_kingdom\\_of\\_taxonomy/](http://www.greenchameleon.com/gc/blog_detail/the_kingdom_of_taxonomy/)

besoin de partager leurs connaissances autour d'un problème informationnel d'une manière explicite, synchrone et directe entre eux. En conséquence, la tendance aujourd'hui est vers la recherche collaborative d'information en utilisant la technologie pour faciliter la collaboration explicite entre un groupe d'individus lors de la résolution d'un problème informationnel. Nous aborderons ainsi la recherche collaborative d'information dans le chapitre suivant.







# Chapitre 4

---

## Du comportement informationnel collaboratif à la recherche collaborative d'information

### 4.1 Introduction

Jusqu'à présent, nous avons développé les trois principaux concepts de ce sujet de thèse, c'est-à-dire la *collaboration*, la *recherche d'information* et l'*intelligence économique*. Comme expliqué au chapitre 2, le processus d'IE implique, entre autres, l'identification du besoin informationnel, l'accès à l'information et la recherche d'information<sup>95</sup>. Ces trois étapes peuvent être réalisées dans un processus de collaboration. Par conséquent, il est important d'intégrer la recherche collaborative d'information dans les étapes de processus de l'IE. Dans ce chapitre, nous détaillons la recherche collaborative d'information (RCI) en tant que combinaison des concepts de la collaboration et de la recherche d'information. Soulignons dans un premier temps un fait important : nous ne pouvons pas expliquer la RCI sans décrire la pratique informationnelle collaborative. Ceci nous amène donc à commencer ce chapitre par l'étude du comportement informationnel collaboratif. Dans un deuxième temps, nous expliquerons ce que nous entendons par RCI, les travaux de recherche existant dans ce domaine, ainsi que notre cadre conceptuel pour la gestion de la RCI.

### 4.2 Comportement informationnel collaboratif

Dans le chapitre précédent, nous avons étudié le comportement informationnel de l'utilisateur. La plupart des modèles existants se focalise sur le comportement informationnel individuel, surtout les modèles issus de l'approche cognitive et de l'approche de la théorie de l'activité

---

<sup>95</sup> Rapellons que l'accès à l'information est un concept qui concerne les méthodes employées par l'utilisateur pour découvrir les sources d'information ou pour y avoir accès (Wilson 1999). Elle désigne le « contexte dans lequel se trouve un utilisateur au moment de choisir les sources d'information qu'il juge pertinentes pour répondre à son besoin informationnel » (Ihadjadene et Chaudiron 2008, p.185). Ayant choisi une source d'information, par exemple un moteur de recherche, l'utilisateur commence un processus de formulation et de reformulation de requêtes. Ces requêtes, qui représentent son besoin informationnel, sont formulées tout en tenant compte des spécificités fonctionnelles du moteur de recherche ou du SRI qu'il utilise. Ce contexte, dans lequel l'utilisateur interroge un SRI en formulant des équations de recherche, est désigné sous le terme « la recherche d'information ».

(Wilson 1981 ; Khulthau 1991, p.361-371). Même si certains de ces modèles font référence aux contextes socio-culturels de l'utilisateur, ils traitent toujours la recherche d'information comme un processus individuel. En revanche, il a été remarqué que les utilisateurs recherchent de l'information d'une manière collaborative (Romano et coll. 1999, p.1 ; Reddy et Jansen 2008, p.256). Romano et coll. appellent cet écart entre la pratique informationnelle vraie de l'utilisateur et la modélisation faite de celui-ci « le paradoxe de la recherche d'information » :

*« D'une part, les recherches et développements dans la RI sont focalisés quasi exclusivement sur des individus. D'autre part, l'observation des gens effectuant une recherche d'informations montre que, quels que soient l'endroit et le temps de recherche, soit dans une bibliothèque, une organisation, ou dans un groupe de touristes, les ressources d'informations sont souvent utilisées d'une manière collaborative pour rechercher des informations et pour la prise de décision »* (Romano et coll. 1999, p.2).

Il est important de noter que la recherche d'information est un processus à la fois cognitif et social. Sur le point de vue cognitif, le but final de la RI est la production de connaissance. Karamuftuoglu a ainsi dit que « *il a été prouvé que la problématique fondamentale associée au processus de recherche d'information est la production et l'utilisation de connaissance et que la production de connaissance est une tâche collaborative* »<sup>96</sup> (Karamuftuoglu 1998, p.1070). La RI a été considérée comme une activité individuelle du fait que, dans ce domaine, l'accent a toujours été mis sur l'interaction entre un seul individu et un système. Aujourd'hui, les progrès dans les technologies de la communication, de l'information et du web ont démontré qu'il est possible de faciliter la collaboration entre individus par l'intermédiaire de la machine. Pour profiter de ces développements dans le domaine de la RI, il est nécessaire de revisiter la modélisation du processus de recherche d'information.

Depuis la fin des années 1990, des chercheurs étudient comment les utilisateurs résolvent leur problème informationnel de manière collaborative. Les principales questions soulevées sont (Hansen et Järvelin 2005, p.1105 ; Reddy et Jansen 2008, p.260) :

-

Comment les

activités de collaboration se manifestent-elles?

---

<sup>96</sup> *It is argued that the fundamental intellectual problems of information retrieval are the production and consumption of knowledge. Knowledge production is fundamentally a collaborative labour, which is deeply embedded in the practices of a community of participants constituting a domain* (Karamuftuoglu 1998, p.1070).

- A quelle fréquence ces activités de collaboration surviennent-elles pour les problèmes informationnels?
- A quel moment surviennent-elles lors du processus de recherche d'information ?
- Quelles sont les caractéristiques de telles activités ?

La RCI a été définie comme étant les activités entreprises par un groupe d'utilisateurs afin d'identifier et de résoudre un problème partagé (Pollock et coll. 2003, p.239). Comme nous l'avons fait pour la recherche d'information (cf. 3.2.), nous pouvons également étendre le modèle imbriqué de pratique informationnelle de Wilson (1999) pour rendre compte des trois niveaux comportementaux de la recherche collaborative d'information. Ces niveaux, comme le montre la figure 4.1, sont le comportement informationnel collaboratif<sup>97</sup>, l'accès collaboratif à l'information<sup>98</sup> et la recherche collaborative d'information<sup>99</sup>.

La couche extérieure de la figure 4.1 correspond au *comportement informationnel collaboratif* qui concerne toutes les pratiques ou comportements informationnels des utilisateurs recherchant de manière collaborative des informations, généralement dans des environnements numériques. La deuxième couche, correspondant à l'accès collaboratif à l'information, est utilisée pour désigner le contexte dans lequel se trouve un groupe d'utilisateurs au moment de choisir les sources d'information qu'il juge pertinentes pour répondre à ses besoins informationnels (il s'agit ici d'un besoin partagé). Le noyau, associé à la RCI, désigne le contexte d'interrogation d'un SRI, c'est-à-dire la formulation collaborative de requête et l'évaluation de la pertinence d'information retrouvée.

---

<sup>97</sup> Collaborative Information Behaviour (CIB)

<sup>98</sup> Collaborative Information Seeking (CIS)

<sup>99</sup> Collaborative Information Retrieval (CIR)

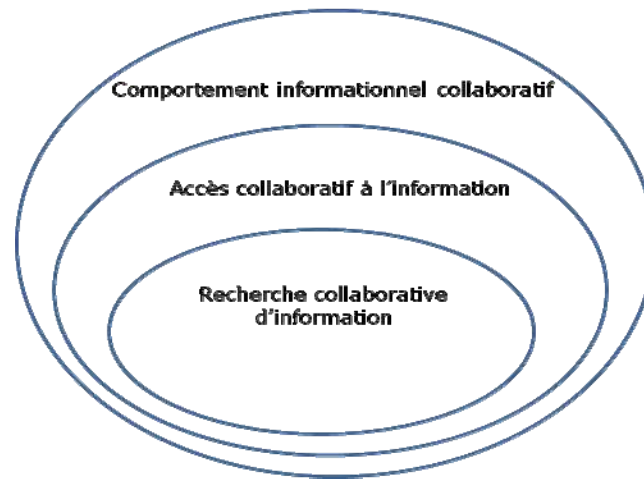


Figure 4.1 : Modèle imbriqué de la pratique informationnelle collaborative, adapté de (Wilson 1999)

Selon le concept de collaboration développé au chapitre 1, il n'est pas utile, lorsqu'on recherche de l'information d'une manière collaborative, de considérer chaque niveau comportemental (*behaviour, seeking* et *retrieval*) comme distinct l'un de l'autre. Nous préférons donc les considérer comme formant un continuum. Ainsi nous retiendrons le terme « recherche collaborative d'information » pour décrire toutes les activités collaboratives couvrant ces trois niveaux comportementaux. Dans le but de développer le concept de RCI, il nous paraît important de présenter, dans les sections suivantes, trois modèles qui ont été développés pour analyser le comportement informationnel collaboratif.

#### 4.2.1 Modèle pour la compréhension du CIB selon Reddy et Jassen

Reddy et Jassen (2008, p.261-268) ont créé un modèle, issu de deux études empiriques, expliquant le passage du comportement informationnel individuel au comportement informationnel collaboratif (voir figure 4.2). Ce modèle montre la différence entre ces deux comportements selon les trois niveaux comportementaux : recherche d'information, accès à l'information et pratique informationnelle. Pour ces auteurs, une situation de recherche d'information peut être décrite par :

- le comportement : un continuum allant de la recherche d'information à l'accès et à l'utilisation d'information ;

- le contexte :  
un continuum du contexte individuel (IIB<sup>100</sup>) au contexte collaboratif (CIB<sup>101</sup>).

Ils ont également identifié trois caractéristiques de l'environnement qui influent sur ces deux axes :

- la complexité du problème à résoudre : dans un contexte individuel, le problème informationnel est relativement simple. En revanche, un contexte collaboratif peut signifier un problème informationnel complexe ;
- le nombre d'agents impliqués dans l'activité : un agent pouvant être humain ou technique (ex. un SRI). Selon le contexte, le nombre d'agents impliqués peut varier ;
- le type d'interaction entre ses agents : selon le contexte, l'interaction peut être directe et simple ou elle peut être complexe, indirecte, conversationnelle et extrêmement riche.

Reddy et Jasen ont montré que le passage de la recherche individuelle à la recherche collaborative dépend de facteurs déclencheurs spécifiques (*triggers*) tels que l'insuffisance de connaissances chez l'individu, la complexité du problème, des ressources d'information trop dispersées. Ces facteurs amènent un individu à demander de l'aide à d'autres individus afin de résoudre son problème. Dinet (2007, p.8) a souligné que ce modèle peut être un outil pertinent pour décrire les situations rencontrées au niveau des comportements, mais qu'il ne renseigne pas sur les processus cognitifs sous-jacents. En conséquence, il a abordé les différences cognitives entre la recherche individuelle d'information et la RCI. Le modèle métacognitif qu'il a élaboré avec Robin Vivian sera présenté dans la section suivante.

---

<sup>100</sup> *Individual Information Behavior*

<sup>101</sup> *Collaborative Information Behavior*

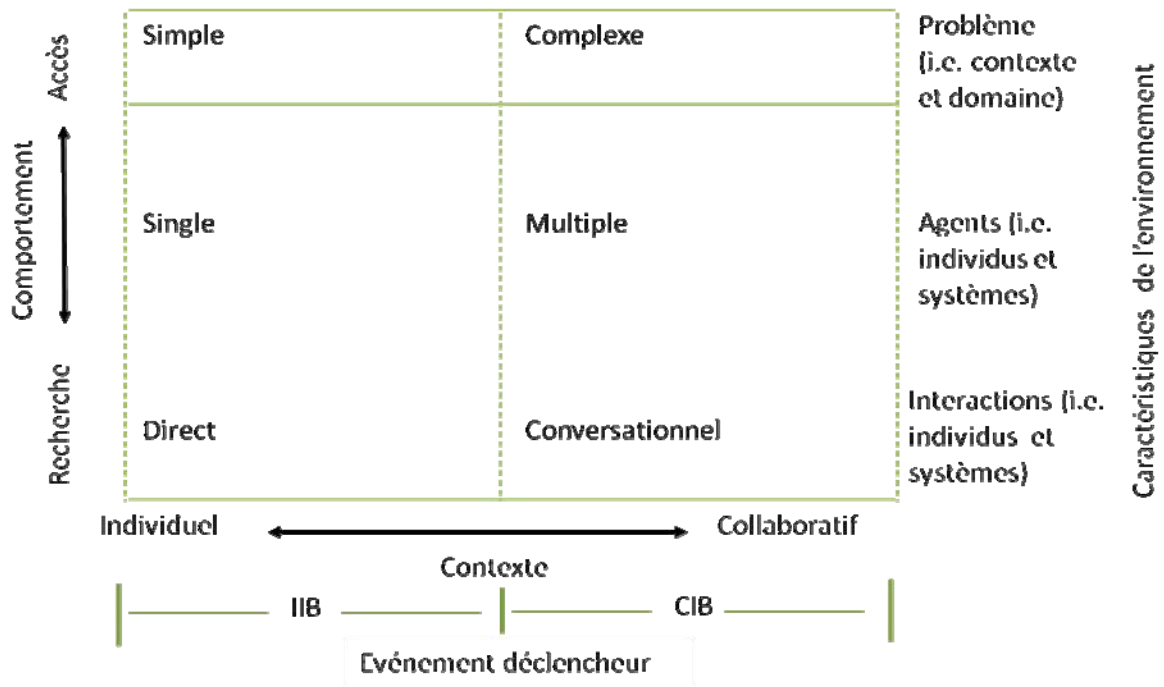


Figure 4.2 : Comportements impliqués lors des recherches individuelle et collaborative d'information, d'après Reddy et Jansen (2008, p.266).

#### 4.2.2 Modèle métacognitif de Vivian et Dinet

Vivian et Dinet, sur la base des modèles actuels de recherche d'information, ont émis l'hypothèse que la RCI demande un plus grand nombre de connaissances et de compétences que la recherche individuelle d'information, notamment en matière de gestion métacognitive de l'activité (Vivian et Dinet 2007, p.33-34 ; Dinet 2007, p.6-7). Ils sont d'avis que certaines compétences non sollicitées lors d'une activité de recherche d'information réalisée de manière individuelle se trouvent l'être lorsque cette même activité est réalisée de manière collaborative.

Pour eux, lors d'une recherche individuelle d'information, les connaissances déclaratives liées au contenu, à l'objectif et au thème de recherche et les connaissances procédurales liées à la méthode, la stratégie et à la procédure de recherche sont largement sollicitées. Ils ajoutent que certaines compétences métacognitives sont également mises en œuvre pour que l'utilisateur puisse gérer et contrôler ses propres comportements. Les facteurs affectifs interviennent également lors d'une recherche individuelle. Toutes les connaissances et les compétences sollicitées dans le cadre de la recherche individuelle le sont aussi dans celui de la recherche

collaborative, mais, en plus, de nouvelles compétences métacognitives liées à la gestion des relations interpersonnelles s'ajoutent. Ces dernières sont généralement rapprochées au savoir-être<sup>102</sup>. De nouveaux aspects affectifs tels que l'amitié et l'affinité interviennent dans la recherche collaborative d'information. Certains de ces facteurs ont été traités au chapitre 1 lors de l'étude des facteurs sociocognitifs associés à la collaboration.

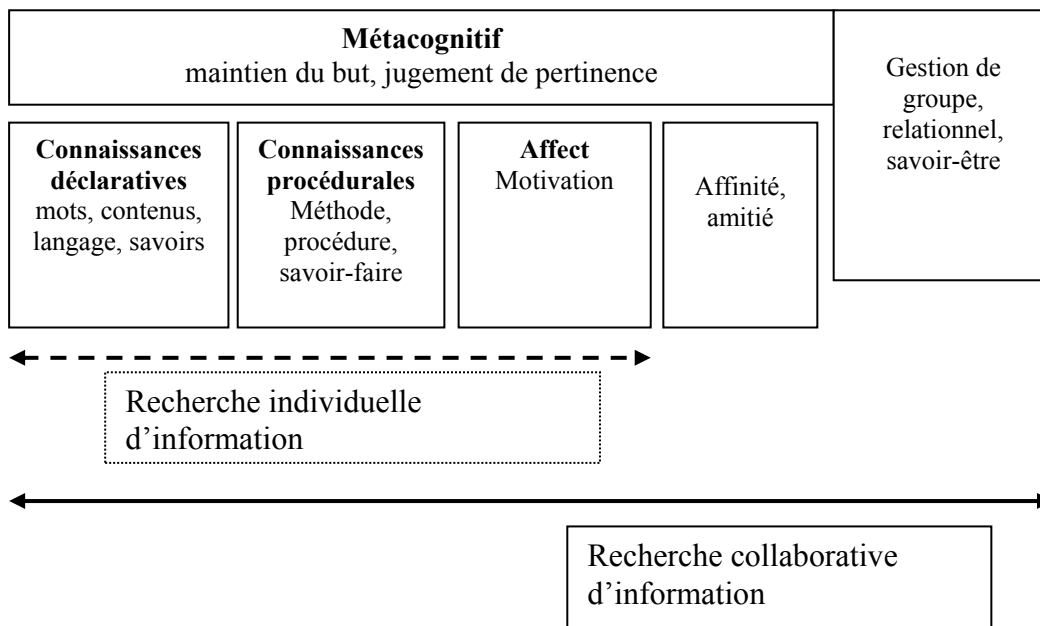


Figure 4.3 : Connaissances et métaconnaissances impliquées lors d'une recherche collaborative d'information (Dinet 2007, p.7)

Dans un autre travail, Dinet, Simonot et Vivian ont vérifié expérimentalement que les liens affectifs influent sur la recherche d'information réalisée de manière collaborative (Dinet et coll. 2008, p.142). Leurs résultats ont confirmé ceux obtenus dans les travaux antérieurs tels que : (Crook 1998, p.242-245 ; Hyun 1991, p. 69-91 ; Vass 2002, p.104-109).

L'un des travaux s'inscrivant dans l'analyse du comportement informationnel collaboratif est celui de Fidel et Petjerson (2004). Ils ont utilisé le cadre de l'analyse du travail cognitive dans leur étude. Ceci sera présenté dans la section suivante.

<sup>102</sup> Le savoir-être correspond à la capacité à s'adapter à l'environnement humaine en ajustant ses comportements. Dans le cadre de la relation interpersonnelle, il s'agit de la capacité à être à l'écoute d'un partenaire et à produire des actions et des réactions adaptées à la relation.



### 4.2.3 Cadre d'analyse de travail cognitif pour la recherche collaborative d'information

Le cadre de l'analyse du travail cognitif (*Cognitive Work Analysis*) a été développé par Jens Rasmussen (Vicente 1999, p.xiv). Ce cadre conceptuel, permettant l'analyse des travaux cognitifs, est fondé sur l'hypothèse que les actions humaines sont guidées par des objectifs. Il comporte sept dimensions (voir figure 4.4). Fidel et coll. (2004, p.942) l'ont utilisé dans leur analyse de l'interaction homme-information (autrement dit le comportement informationnel). Fidel et Pejtersen (2004) l'ont aussi mis en œuvre pour étudier le comportement collaboratif informationnel dans trois équipes de deux compagnies internationales. Leur opinion est la suivante : la recherche d'information devient collaborative lorsque les acteurs impliqués sont des collègues. Ils sont d'avis qu'un bibliothécaire aidant un utilisateur à rechercher de l'information ne correspond pas à une recherche collaborative. En revanche, pour eux, deux bibliothécaires recherchant ensemble de l'information, participent à une activité de recherche collaborative d'information.

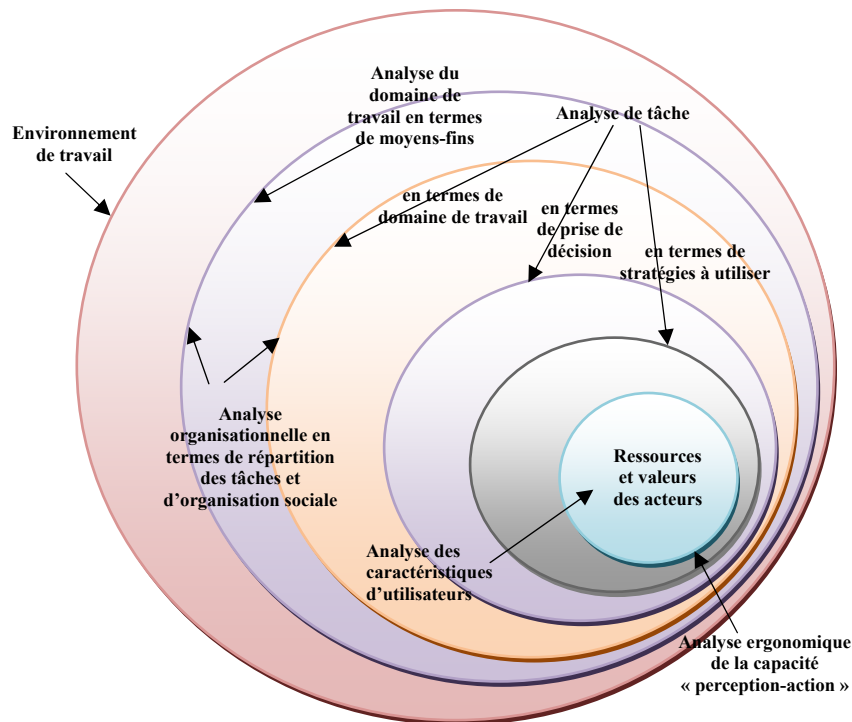


Figure 4.4: Cadre d'analyse de travail cognitif pour la RCI (Fidel et coll. 2004, p.943)

Nous ne sommes pas d'accord avec cet avis car la collaboration peut ne pas être limitée aux collègues. L'apport de la collaboration est dans la complémentarité des connaissances des participants. Néanmoins nous trouvons intéressante leur approche de l'analyse du comportement collaboratif informationnel dans une situation réelle, en utilisant le cadre

d'analyse de travail cognitif. Une telle analyse peut aider les concepteurs de système de RCI à mieux tenir compte des aspects humains dans leur conception. Des exemples de questions posées lors de l'analyse de chaque dimension sont présentés dans le tableau 4.1.

<b>Dimension</b>	<b>Exemples des questions à poser lors d'analyse</b>
<b>Environnement</b>	Quels sont les éléments hors de l'organisation qui jouent sur l'environnement de travail ?
<b>Domaine de travail</b>	Quels sont les buts, les contraintes, les priorités et les fonctions du domaine de travail ? Quels sont les processus physiques impliqués ? Quels sont les outils employés ?
<b>Analyse organisationnelle</b>	Comment le travail est-il divisé parmi les équipes ? Quels sont les critères utilisés ? Quelle est la nature de l'organisation : hiérarchique, démocratique, chaotique ? Quelles sont les valeurs de l'organisation ? Quels sont les outils utilisés ?
<b>Analyse de tâche en termes de domaine de travail</b>	Quelle est la tâche à réaliser ? Quels sont les buts de cette tâche qui suscitent des problèmes informationnels ? Quelles sont les contraintes et les fonctions impliquées ?
<b>Analyse de tâche en termes de la prise de décision</b>	Quelles sont les décisions à prendre ? Quelles sont les informations nécessaires ? Quelles sont les sources d'information utiles ?
<b>Analyse de tâche en termes de stratégies à utiliser</b>	Quelles sont les stratégies possibles (ex. exploration) ? Quelle stratégie préfère l'acteur ? De quel type d'information a-t-il besoin ? Quelles sont ses sources d'information préférées ?
<b>Ressources et valeurs des acteurs</b>	Quelle est sa formation ? Quel est son domaine d'expertise ? Quelles sont ses expériences dans le domaine du sujet et dans le domaine de travail, ses priorités/préférences personnelles, ses valeurs ?

Tableau 4.1 : Dimensions de l'analyse de travail cognitif et exemples des questions à poser lors d'analyse de chaque dimension, adapté de (Fidel et coll. 2004, p.943)

### **4.3 La recherche collaborative d'information (RCI)**

Il existe plusieurs définitions de la RCI. Nous en citerons quelques unes afin de donner un contexte pour la notre. Selon Foster, la recherche collaborative d'information peut être définie comme l'étude des systèmes et des pratiques permettant aux individus de collaborer lors de l'accès à l'information et lors de la recherche d'information (Foster 2006, p.330). Pour Fidel

et coll. (2001, p.236), la RCI comporte toute activité permettant de résoudre collectivement un problème informationnel. La définition de Fidel et coll. nous semble trop limitative, comme nous l'avons souligné précédemment, car pour eux la RCI dépeint seulement une situation de collaboration entre des collègues ou au sein d'une équipe. Pour Hansen et Järvelin (2005, p.1102), la RCI est toute activité d'accès à l'information, dans un but de résoudre un problème informationnel, impliquant (explicitement ou implicitement) plusieurs individus interagissant directement ou interagissant à travers des artefacts (ex. documents) dans des contextes plus ou moins définis et des environnements plus ou moins ouverts<sup>103</sup>.

En nous fondant sur notre étude du concept de collaboration et sur celui de recherche d'information (et en tenant compte des facteurs déclencheurs qui font passer un individu d'une situation de recherche d'information individuelle à une situation de recherche d'information collaborative), nous définissons la RCI de la manière suivante :

« la RCI consiste en des **méthodes** et des **systèmes** de gestion des **activités collectives** des utilisateurs dans un processus de recherche d'informations centré sur un problème informationnel partagé, ce, dans le but de faciliter une **collaboration directe** entre les utilisateurs ainsi que le **partage de connaissance** entre eux » (Odumuyiwa et David 2009, p.496).

Dans cette définition, les mots en gras représentent les éléments importants de la RCI. Notons d'abord que pour effectuer de la recherche collaborative, il faut des méthodes et des systèmes. Les méthodes peuvent être des modèles qui permettent de bien effectuer une activité de collaboration. Les systèmes peuvent être des environnements numériques ou des systèmes de recherche d'information etc... La RCI n'est pas limitée à un seul système. Les activités doivent être collectives et mutualisées. Une situation dans laquelle un participant devient le seul producteur de connaissance<sup>104</sup> et où les autres ne deviennent que des consommateurs, ne reflète pas une bonne situation de collaboration.

---

<sup>103</sup> *Collaborative Information Retrieval (CIR) : CIR is an information access activity related to a specific problem solving activity that, implicitly or explicitly, involves human beings interacting with other human(s) directly and/or through texts (e.g., documents, notes, figures) as information sources in a work task related information seeking and retrieval process either in a specific workplace setting or in a more open community or environment [HAN 05].*

<sup>104</sup> Par la production de connaissances dans ce contexte, nous entendons la formulation de requêtes, l'annotation et l'évaluation des objets informationnels par les collaborateurs.

Par collaboration directe, nous entendons une collaboration explicite et synchrone car la collaboration implicite n'est pas vraiment une activité de collaboration mais plutôt une coordination comme expliqué au chapitre 1. Nous avons montré jusqu'à présent que l'un des objectifs les plus importants de la recherche d'information (soit d'une manière individuelle, soit d'une manière collaborative) est l'acquisition et la production de connaissance. Une activité de recherche collaborative d'information doit donc être orientée vers la production, l'acquisition et le partage des connaissances.

La RCI peut être également considérée comme un domaine de recherche émergent combinant deux domaines de recherche : le collectif (*groupware*) et la recherche d'information. Le collectif est « l'ensemble des technologies et des méthodes de travail associées qui, par l'intermédiaire de la communication électronique, permettent le partage de l'information, sur un support numérique, à un groupe engagé dans un travail collaboratif » (Courbon et Tajan 1997, p.1). Les collectifs sont généralement classifiés selon un espace à deux dimensions : lieu et temps, comme le montre la figure 4.5. Cette classification a été reprise par certains auteurs tels que Taher (2004, p.12) et Foley (2008, p.48) pour classer les systèmes de recherche collaborative d'information (SRCI). Les collectifs ont été utilisés dans le passé pour faciliter la communication asynchrone tel que le courrier électronique. De plus en plus d'utilisations synchrones et en temps réel caractérisent les collectifs d'aujourd'hui.

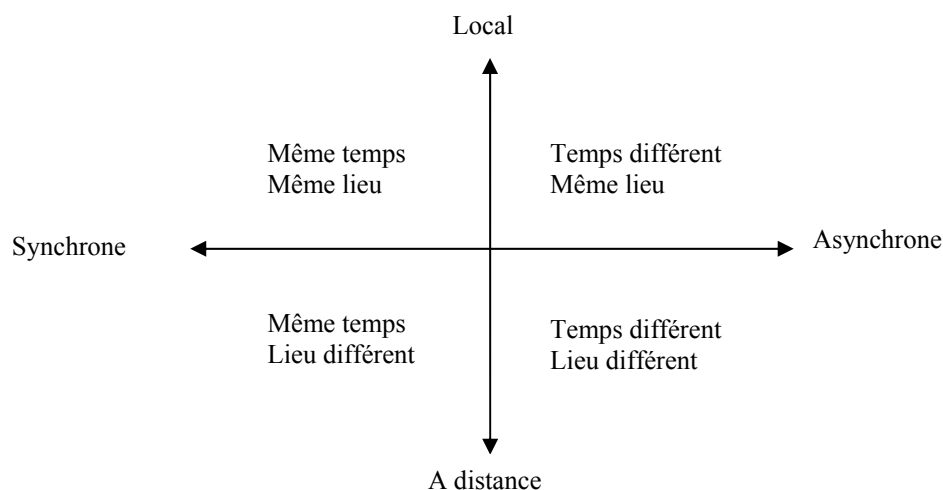


Figure 4.5 : Classification temps-lieu de collectifs

Une taxonomie que nous avons présentée dans la section 1.5.2 permet de mieux classer les SRCI. Elle comporte quatre dimensions : intention, profondeur, synchronisme et localisation. Cette classification est plus riche que celle expliquée dans la figure 4.5. Ayant abordé le concept de RCI, nous présenterons dans la suite de ce chapitre quelques travaux sur les SRCI.

#### 4.4 Systèmes de recherche collaborative d'information existants

Depuis les années 1990, plusieurs applications ont été développées dans le but de faciliter la RCI. Nous présenterons dans cette section quelques systèmes de recherche collaborative d'information que nous avons étudiés au cours de ce travail de thèse. Ils seront présentés sous deux perspectives : système de navigation collaboratif et système de recherche collaborative. Nous n'évoquerons pas dans cette section de systèmes se disant collaboratifs et dans lesquels l'intention de collaboration est implicite. Ce type de système a été expliqué dans le chapitre précédent dans la partie consacrée à la recherche sociale d'information (cf. 3.5).

##### 4.4.1 Systèmes permettant la navigation collaborative

Le GroupWeb (Greenberg et Roseman 1996, p. 271-272) est un navigateur du Web qui permet à un groupe d'individus de partager visuellement des pages Web et de naviguer ensemble en temps réel. Il fut développé dans le but d'intégrer les fonctionnalités des collecticiels dans la navigation web. Lors d'une session collaborative, un utilisateur (*le maître*) peut contrôler l'écran des autres utilisateurs. Le navigateur de tel *utilisateur-maître* devient le « *navigateur-maître* » tandis que ceux des autres participants deviennent les « *navigateurs-esclaves* ». Les navigateurs-esclaves peuvent voir les pages consultées par l'utilisateur-maître d'une manière WYSIWIS<sup>105</sup>. GroupWeb permet un défilement des pages web d'une manière synchrone et d'une manière indépendante. Lorsqu'il s'agit d'un défilement synchrone, à chaque fois que l'utilisateur maître défile une page, la page est défilée de la même façon dans les navigateurs-esclaves. En revanche, le système permet également aux utilisateurs-esclaves de défiler leurs pages d'une manière indépendante. Il permet à tous les utilisateurs d'ajouter des annotations à des pages consultées. Ces annotations sont partagées entre les membres du groupe. Le système dispose également de télépointeurs

---

<sup>105</sup> WYSIWIS est l'acronyme de la locution anglaise « What You See Is What I See », traduit littéralement en français « Ce que vous voyez est ce que je vois ».

pour faciliter la communication par gestes et pour attirer l'attention du groupe sur un élément précis dans la page consultée.

Un autre système qui s'appelle W4 (*World Wide Web for Workgroups* ) browser a été développé à la suite de GroupWeb par Gianoutsos et Grundy (1996, p.14-21). Il permet aux participants, en plus des fonctionnalités de GroupWeb, d'échanger des messages instantanés (chat) lors d'une session collaborative. Ils peuvent partager leurs favoris et également faire du *brainstorming* en partageant un écran blanc.

Cabri et coll. (1999, p. 138–145) ont développé un autre système de co-navigation en utilisant un serveur mandataire (*proxy*). Ils interposent un serveur mandataire entre le client (le navigateur web) et le serveur web pour relayer des requêtes entre les deux. Ceci permet d'éviter toute sorte d'installation d'autres logiciels sur les machines des participants. La gestion du groupe se fait sur le serveur mandataire. La création du groupe est faite par un utilisateur possédant un droit plus élevé que les autres. Chaque membre du groupe peut participer à une session collaborative à partir de son navigateur web mais ceci n'est possible que si l'utilisateur active l'option de serveur de mandataire sur son navigateur.



Figure 4.6 : Copie d'écran du système de cabri et coll. (1999, p.143)

La première demande d'une page web est interceptée par le serveur mandataire qui envoie ensuite une page d'authentification afin d'autoriser l'utilisateur à participer à la session collaborative. Pour les demandes suivantes de pages, le serveur mandataire vérifie si la page demandée a été consultée par un autre utilisateur. Si c'est le cas, le serveur mandataire répond

en envoyant la version de la page stockée dans le cache (car chaque nouvelle page consultée par les membres du groupe est stockée dans le cache). Dans le cas contraire, le serveur mandataire fait passer la demande au serveur web adéquat. Dans ces deux cas, la demande passe toujours par le serveur mandataire qui analyse et modifie la page demandée avant de l'envoyer à l'utilisateur. La page HTML envoyée est composée de deux cadres comme le montre la figure 4.6. Le cadre en haut est un *applet*<sup>106</sup> qui contient des informations sur les activités des membres du groupe. Il permet alors de voir qui sont les autres utilisateurs participant à la collaboration et les URLs visitées par les membres du groupe. Il fournit également une messagerie instantanée pour le groupe. Le cadre en bas contient une version modifiée de la page demandée. Ce système permet une navigation synchronisée entre les participants. Attere et coll. (2007, p.74-87) ont repris cette approche de serveur mandataire pour créer leur système de co-navigation qui s'appelle UsaProxy 2.

CoWeb (Jacobs et coll. 1996, p.1385-1395) fournit des fonctionnalités qui permettent à un groupe d'utilisateurs de collaborer en manipulant simultanément les éléments d'une page web. Par exemple, deux utilisateurs peuvent remplir ensemble un formulaire d'une page, mettre en évidence une partie d'une page (ex. une image) en la soulignant afin d'engager une discussion à son sujet. Le principe de fonctionnement de CoWeb est de remplacer tous les éléments d'une page HTML par des applets Java sans changer le contenu de la page. Les contenus de la page sont intégrés dans les applets qui fournissent, en plus des contenus, des fonctionnalités de collaboration. CoWeb supporte quatre modes de collaboration :

- WYSIWIS permet la synchronisation du navigateur ;
- « Même document » permet aux utilisateurs de travailler sur le même document en même temps sans une synchronisation stricte de la navigation. Cela signifie que lorsqu'un utilisateur défile sa page, cette action n'a aucun effet sur la page d'un autre utilisateur. En revanche, si un utilisateur remplit un champ dans un formulaire, les textes tapés sont immédiatement affichés sur la page de chaque participant ;
- « Versions » permet aux utilisateurs de travailler sur des versions différentes de la même page. Ils peuvent éventuellement fusionner toutes les modifications réalisées ;
- « Document privé » permet à un utilisateur de travailler d'une manière autonome sans partager sa page avec d'autres membres du groupe.

---

<sup>106</sup> Un applet est un programme écrit en langage Java, l'applet peut être intégré dans une page HTML. Lorsqu'un utilisateur consulte une page web comportant un applet, les codes de l'applet sont téléchargés sur son ordinateur puis ils sont exécutés par la machine virtuelle de Java si cette dernière est installée sur son poste.

De son côté, CSCW3 (Gross 1999, p.1-5) utilise le concept de « chambre virtuelle » pour regrouper les utilisateurs qui consultent la même page web en même temps. Il permet aux utilisateurs se trouvant dans la même chambre de communiquer et de synchroniser leur navigation.

Han et coll. (2000, p. 221-230) utilise une approche différente de celle citée ci-dessus. Ils combinent la navigation collaborative avec l'informatique ubiquitaire<sup>107</sup> dans leur système qui s'appelle « Websplitter ». Ce dernier permet de répartir les objets d'une page web entre les membres du groupe. Donc, dans une session collaborative, chaque utilisateur peut voir les différents aspects d'une page partagée. Selon ces auteurs, ce système ne marche qu'avec les pages web développées en XML.

Maekawa et coll. (2006, p.22-33) ont développé un système similaire à celui de Han et coll. Le système permet aux utilisateurs co-localisés de co-naviguer sur le web en utilisant leurs téléphones portables et PDAs. Le système décompose une page web en plusieurs composants qui sont ensuite répartis sur les dispositifs des collaborateurs en fonction de la capacité des dispositifs dont ils disposent.

Laurillau et Nigay (2000, p. 121-124) ont identifié quatre types de navigation possibles pouvant être implémentés dans un système de co-navigation :

- Visite guidée : un guide prend le contrôle de la navigation et les autres le suivent ;
- Navigation opportuniste : le groupe est ouvert, c'est-à-dire que n'importe qui peut devenir membre à n'importe quel moment. Chacun navigue indépendamment mais dispose du résultat du groupe ;
- Navigation coordonnée : tous les membres du groupe ont le même droit. Ils travaillent séparément sur le même sujet mais sans forcément partager le même but ;

---

<sup>107</sup> « L'informatique ubiquitaire est le modèle qui suit l'ordinateur de bureau au niveau de l'interaction homme-machine dans lequel le traitement de l'information a été complètement intégré dans tous les objets des activités journalières. Par opposition au paradigme de bureau, dans lequel un seul utilisateur engage consciemment un dispositif unique dans un but spécialisé, quelqu'un « qui utilise » l'informatique ubiquitaire engage beaucoup plus de dispositifs et de systèmes informatiques simultanément, au cours de ses activités ordinaires, et n'est pas nécessairement averti qu'il en est ainsi » (Wikipédia, consulté le 9 Août 2010).



- Navigation coopérative : le leader du groupe oriente la recherche et définit l'objectif à atteindre. Il peut également répartir les sous-espaces à explorer parmi les membres du groupe.

Laurillau et Nigay ont implémenté ces quatre types de navigation dans le système CoVitesse. Ce système permet à un utilisateur de créer un groupe, puis de choisir un type de navigation à suivre par le groupe. Le choix réalisé permet de signaler aux utilisateurs le mode de fonctionnement du groupe. Chaque utilisateur dispose, à la fin de la session collaborative, des résultats accumulés par le groupe.

Lieberman et coll. (1999, p. 65-68) ont développé *Let's Browse*, un système de co-navigation qui permet à plusieurs utilisateurs co-localisés de naviguer sur le web ensemble. Ce système crée automatiquement le profil d'un utilisateur en pondérant les mots-clefs extraits de sa page web et des pages liées à celle-ci. Les utilisateurs, se tenant debout physiquement devant un large écran, peuvent naviguer ensemble selon leurs profils. Chaque utilisateur porte un badge électronique permettant de l'identifier lorsqu'il s'approche de l'écran. Le système, en se fondant sur la similarité des profils des utilisateurs en co-navigation, leur recommande des liens à consulter. Ce système diffère du système de recommandation collaborative évoqué en 3.5.1 car les collaborateurs travaillent dans ce cas en temps réel et ils sont co-localisés, ce qui implique qu'ils se connaissent physiquement.

Nous avons abordé ci-dessus des exemples de systèmes de co-navigation existants. Ces systèmes ont généralement pour but d'introduire le travail collaboratif dans la navigation du Web. Ils se focalisent plus précisément sur la navigation. En revanche, un SRCI implique le partage du processus de recherche, avec la navigation collaborative intégrée, ainsi que des résultats de recherche. Nous pouvons donc dire que les systèmes de co-navigation sont un sous-ensemble des SRCI. Nous constatons que certaines fonctionnalités sont communes aux systèmes de co-navigation : la conscience du groupe (*awareness*), la synchronisation du travail et la communication entre les collaborateurs. Ces fonctionnalités sont également importantes dans la recherche collaborative d'information.

Nous allons donc, dans la section suivante, décrire certains systèmes qui sont développés avec l'intention de faciliter la recherche collaborative d'information.

## 4.4.2 Les systèmes de recherche collaborative

### 4.4.2.1 C-TORI

Hoppe et Zhaho (1994, p.103-113) ont développé le système C-TORI qui est une version collaborative de TORI (Task-Oriented Database Retrieval Interface). C'est une interface pour la recherche collaborative dans des bases de données. Le système étend la formulation interactive de requête en facilitant la coopération entre plusieurs utilisateurs. L'architecture TORI est composée de trois objets : un formulaire de requête, un formulaire de résultat et une fenêtre d'historique. La version collaborative permet d'effectuer trois opérations sur les trois objets de TORI. Les opérations sont « couplage » (*coupling*), « copiage » (*copying*) et « fusion » (*merging*). C-TORI permet à ses utilisateurs de collaborer par :

- La formulation collaborative de requête : un groupe d'utilisateurs peut exprimer conjointement son besoin informationnel sous la forme d'une requête. Les utilisateurs peuvent coupler leurs activités. Ceci permet une synchronisation des activités de recherche d'une manière WYSIWIS. C-TORI emploie un modèle de contrôle centralisé pour gérer la collaboration. Donc la collaboration est initiée par un utilisateur ayant le rôle de coordinateur. Un utilisateur peut copier une requête de l'environnement d'un autre utilisateur dans son propre environnement. Un utilisateur peut également fusionner sa requête avec celle d'un autre utilisateur ;
- Le partage d'historique des requêtes : chaque requête précédente contient la spécification de requête et, éventuellement, un sous-ensemble sélectif des résultats obtenus lors de son exécution. C-TORI fournit un mécanisme de partage d'historique qui permet d'effectuer les trois opérations de collaboration citées ci-dessus. Quand l'historique de requête d'un groupe d'utilisateurs est couplé, chaque requête formulée par un membre du groupe est automatiquement ajoutée à l'historique de chaque membre ;
- La navigation collaborative des résultats : les utilisateurs peuvent naviguer sur les résultats de recherche d'une manière collaborative en couplant leurs navigateurs (formulaire de résultat) sans coupler le formulaire de requête. Dans ce cas, il n'y a pas de formulation collaborative de la requête et celle-ci n'est pas être exécutée dans l'environnement de chacun. Ceci permet aux utilisateurs de partager les résultats sans accéder à nouveau à la base de données.

#### 4.4.2.2 ARIADNE

Le système ARIADNE (*Annotatable Retrieval of Information And Database Navigation Environment*) est parmi les premiers SRCI développés. Son objectif est d'aider à l'acquisition collaborative de compétences pour la navigation dans une base de données. C'est une interface graphique pour la visualisation du processus de recherche. Twidale et coll. (1997, p.771-775) ont remarqué à travers d'une série d'études qu'il est important de partager le processus de recherche ainsi que les résultats de recherche parmi un groupe d'utilisateurs qui collaborent. Ils observent également que lorsqu'un utilisateur dans une bibliothèque explique au bibliothécaire qu'il n'arrive pas à trouver un document, ce dernier lui demande d'expliquer le processus qu'il a suivi pour sa recherche. La plupart du temps, l'utilisateur en question n'arrive pas à expliquer ses activités de recherche car il lui est difficile de s'en souvenir parce qu'il s'est plutôt focalisé sur le but final de sa recherche. De ce fait, le système ARIADNE a été développé pour capter automatiquement les activités (les requêtes et les résultats) de recherche d'utilisateurs. ARIADNE produit une visualisation de ces activités permettant ainsi aux utilisateurs de réfléchir et de partager des techniques sophistiquées de recherche.

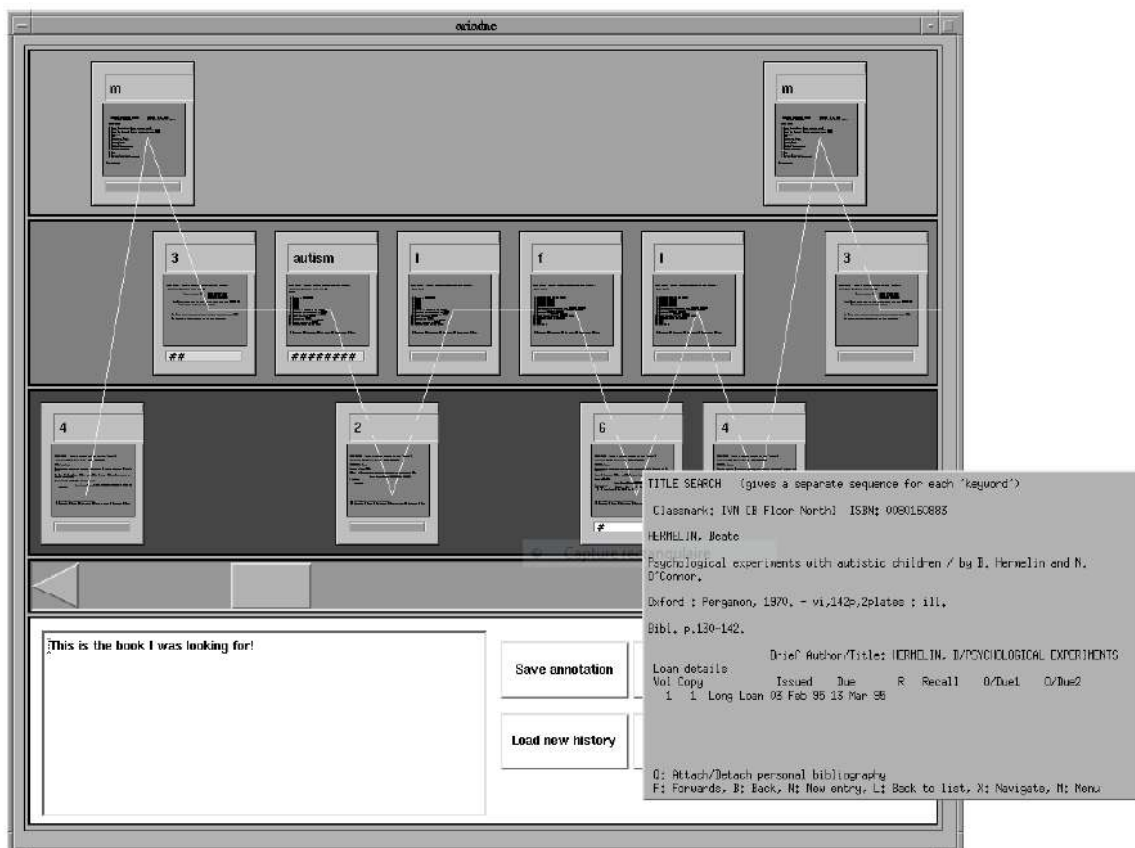


Figure 4.7 : Interface de visualisation de recherche dans ARIADNE

La visualisation du processus de recherche en ARIADNE est composée d'une séquence de « cartes à jouer » contenant les requêtes et les vignettes des copies d'écran des résultats (voir la figure 4.7). Les cartes peuvent être agrandies en plein écran par un clic de souris. Un historique dans le temps, sur les activités menées lors d'une session de recherche, peut être consulté en défilant les séquences de cartes de gauche à droite. Des informations sur la sémantique des actions sont données par la position verticale des cartes le long de trois niveaux (lignes) de visualisation: au premier niveau se trouvent le choix du menu principal, le deuxième niveau contient les requêtes de recherche et le troisième niveau comporte les vignettes de résultats (Twidale 1997, p.777-779 ; Taher 2004, p.25-26).

Ce système possède deux limitations :

- son but unique est de faciliter l'apprentissage de la navigation dans une base de données, mais pas de résoudre de manière collaborative un problème (à la différence de la RCI qui tend vers la résolution collaborative) ;
- le système est limité à une seule base de données. Or, aujourd'hui la recherche d'information se fait sur Internet et permet la consultation d'une multitude de sources d'information. Donc un vrai système de recherche collaborative d'information doit permettre aux utilisateurs de rechercher sur le web.

#### **4.4.2.3 CIRE**

Romano et coll (1999, p.1-10) ont développé le système CIRE (*Collaborative Information Retrieval Environment*) pour faciliter la RCI. Ils ont étudié les expériences d'utilisation de SRI et de systèmes de support de groupe (GSS). Ils ont ensuite fusionné les paradigmes de ces deux domaines dans un seul environnement. CIRE dispose donc des fonctionnalités de recherche d'information individuelle ainsi que celles de recherche collaborative.

CIRE utilise Alta-Vista comme moteur de recherche auquel la requête utilisateur est soumise. Le résultat de recherche est présenté avec une possibilité d'évaluer et de commenter sous forme d'annotations les pages retrouvées. Les requêtes soumises, les résultats de recherche ainsi que les annotations faites sont tous enregistrés dans une base de données. Ce système permet à plusieurs utilisateurs de partager leurs requêtes, leurs résultats de recherche et leurs

annotations. Les membres d'un groupe peuvent ainsi créer, en s'inspirant des travaux des uns et des autres, des requêtes et donc des résultats auxquels ils n'auraient pas pensé si chacun avait travaillé seul.

CIRE dispose d'une fonction de re-exécution des requêtes stockées. Cette fonction fait remonter aux utilisateurs les pages que la requête retrouve et qui n'avaient pas été retrouvées lors de la dernière exécution de la requête. Les utilisateurs peuvent également modifier ces requêtes avant de les re-soumettre. CIRE ne permet pas une collaboration synchrone.

#### **4.4.2.4 METIORE («Multimedia coopErative informaTION Retrieval systEm »).**

Amos David (1999, p.47-54) a proposé un système fondé sur la notion de communication homme-homme où les améliorations de recherche apportées au SRI ne sont plus confiées uniquement au système mais confiées également à un être humain, par exemple à un médiateur. Sa notion de médiateur englobe toute personne capable d'apporter une aide dans un contexte de recherche d'information. Le médiateur peut être un utilisateur qui possède déjà de l'expérience dans la recherche d'information sur un sujet similaire, un expert en recherche documentaire, un expert du domaine de l'information recherchée, etc...

Il a développé une architecture de collaboration qui a été utilisée par la suite pour élaborer deux SRCI. Le premier est un système de recherche d'information bibliographique appelé METIORE. Le deuxième, appelé STREEMS, est un système de recherche d'information multimédia pour gérer les données botaniques et des données d'exploitation sur les arbres.

L'architecture de collaboration de David (également appelée METIORE) est composée de quatre modules : un serveur de localisation (SL), un serveur d'application (SA), une interface de connexion (IC) et l'interface de l'application (IA). L'objectif est de permettre à deux SRI de communiquer entre eux. Cette architecture est similaire à l'architecture *broker* dans les systèmes distribués où chaque application peut être à la fois client et serveur. La communication est faite entre deux applications sans passer par un serveur central (Taylor 2005, p.5, 9, 29).

Sur le serveur de localisation se trouvent des informations telles que l'identité de chaque utilisateur, leurs états (montrant si l'utilisateur est en collaboration ou en état autonome) et

leurs partenaires. Ce serveur enregistre également les informations sur toutes les applications connectées et les envoie à chaque application enregistrée. Le serveur de localisation envoie et reçoit des messages par l'intermédiaire du serveur d'application.

Le serveur d'application reçoit tous les messages d'un serveur application partenaire, effectue les transformations nécessaires et transmet le message au module auquel il est destiné.

L'interface de connexion gère la procédure de contrôle d'entrée en communication avec une autre application. L'IA est composée d'un moteur de recherche et d'une interface de recherche.

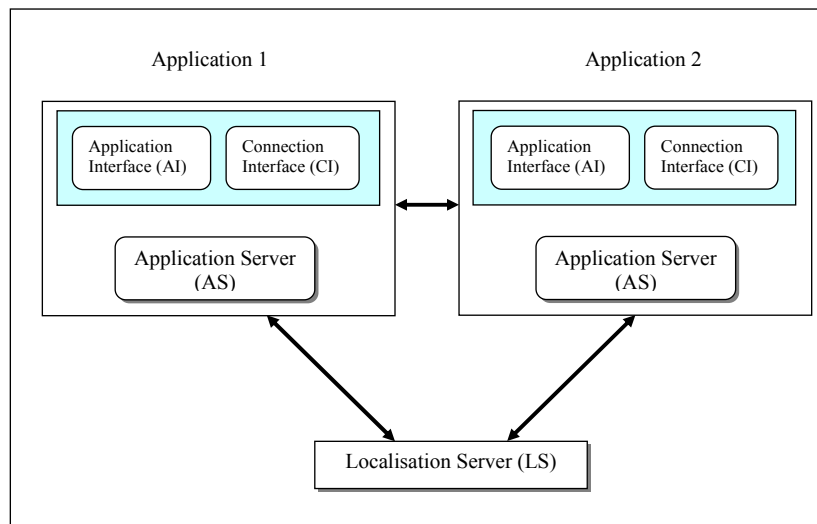


Figure 4.8 : Schéma global de l'architecture d'un SRCI (David 1999, p.50)

METIORE et STREEMS permettent une collaboration synchrone entre deux utilisateurs. Les utilisateurs peuvent partager leurs requêtes ainsi que leurs résultats de recherche.



Figure 4.9 : Interface de STREAMS (David 1999, p.108)

#### 4.4.2.5 SearchTogether

*SearchTogether* est un des systèmes de recherche collaborative d'information récent développé par Morris et Horvitz (2007, p.3-12) dans une équipe de recherche de Microsoft. Il cherche à faciliter la collaboration dans un petit groupe d'utilisateurs souhaitant satisfaire un besoin informationnel partagé. Comme d'autres systèmes expliqués ci-dessus, il permet à un groupe d'utilisateurs de partager leur processus de recherche. Ce système dispose de trois fonctionnalités principales : la conscience du groupe, la répartition des tâches et la persistance du processus de recherche.

La conscience des activités des membres du groupe permet d'éviter une duplication de travail. Elle permet l'apprentissage des méthodologies de recherche employées par les membres du groupe. *SearchTogether* implémente un mécanisme de conscience de groupe en représentant chaque membre par son nom et sa photo. Chaque fois qu'un membre effectue une recherche, les requêtes formulées sont ajoutées à une liste en dessous de sa photo. Son historique est alors synchronisé avec ceux des autres membres. Un autre membre du groupe peut voir les résultats associés à une requête en cliquant dessus. Chaque membre, en consultant une page, peut voir les autres membres qui l'ont déjà consultée. Ils partagent également, entre eux, leurs

évaluations. Les choix d'évaluation sont uniquement binaires et symbolisés par une icône en forme d'un pouce levé ou d'un pouce baissé.

*SearchTogether* fournit une fonctionnalité de répartition des tâches en implémentant plusieurs mécanismes : messagerie instantanée, mécanisme de recommandation, *split-search* et *multi-search*. La messagerie instantanée permet aux utilisateurs de coordonner leurs interactions. Leurs conversations sont enregistrées pour qu'ils puissent les consulter en mode asynchrone. Le mécanisme de recommandation permet à un utilisateur d'envoyer une page web qu'il est entrain de consulter à un autre utilisateur pour que ce dernier puisse également la consulter. La fonction *split-search* permet à un utilisateur de répartir le résultat de recherche entre les participants d'une session collaborative. La fonction *multi-search*, en revanche, permet de soumettre une requête à différents moteurs de recherche, chacun étant associé à un utilisateur différent.

La persistance du processus de recherche facilite une collaboration asynchrone en permettant à un utilisateur de voir ce que les autres membres du groupe ont fait lorsqu'ils étaient hors ligne. Elle permet à un utilisateur de se rappeler également ses dernières activités lors d'une session de recherche antérieure. Tous les aspects de *SearchTogether* sont persistants. La figure 4.10 montre l'interface de *SearchTogether*.



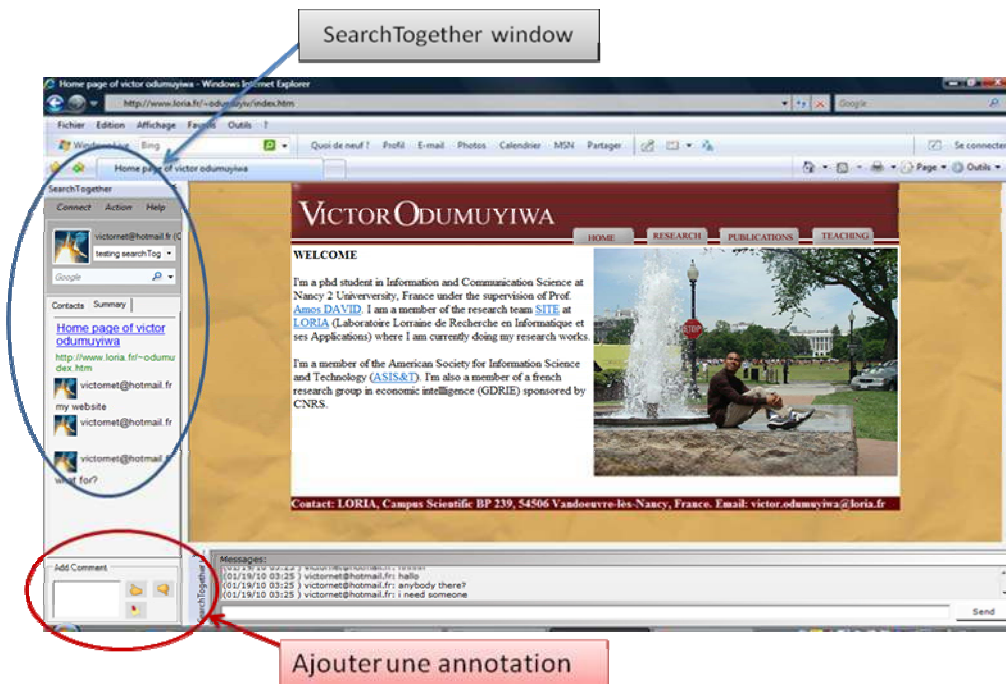


(a) messagerie instantanée, (b) conscience du groupe, (c) résultats actuels, (d) éléments recommandés, (e)(f)(g) boutons de recherche, (h) métadonnées spécifiques à une page, (i) barre d'outils, (j) navigateur (Morris et Horvitz 2007, p.3-12)

Figure 4.10 : Interface de SearchTogether.

Un des défauts de SearchTogether est de ne pas permettre aux utilisateurs de se concerter pour définir suffisamment bien l'objectif de la recherche collaborative. Sans cette phase préalable, leur recherche risque de ne pas répondre de manière suffisamment précise à leurs besoins.

SearchTogether a été retravaillé pour rendre possible son installation en *plugin* sur Microsoft Internet Explorer. Sa dernière version, que nous avons installée sous Internet Explorer, est présentée dans la figure 4.11.



La figure 4.11 Microsoft SearchTogether

#### 4.4.2.6 Le système de Foley

Foley (2008a, p.78-100) propose de combiner le partage de connaissance et la répartition des tâches dans la recherche collaborative d'information effectuée d'une manière synchrone (voir figure 4.11). L'idée est d'utiliser le bouclage de pertinence<sup>108</sup> des utilisateurs participant à une session collaborative afin d'ordonner la liste des résultats de recherche. Généralement, dans les SRI traditionnels, le bouclage de pertinence est souvent utilisé pour reformuler la requête de l'utilisateur afin d'améliorer la qualité de la recherche. Foley s'intéresse à la façon dont les bouclages de pertinence, provenant de plusieurs utilisateurs, peuvent être intégrés dans un SRCI. C'est-à-dire, comment le bouclage de pertinence collaborative peut être utilisé pour améliorer le processus de recherche d'un groupe d'utilisateurs lors d'une session collaborative ?

Foley implémente un mécanisme de répartition de tâches permettant à chaque membre d'un groupe de formuler indépendamment sa requête et d'avoir une liste de résultats de recherche qui lui est unique. Son approche est de filtrer la liste ordonnée de résultats fournie en réponse à une requête. Le filtrage est fait en enlevant les documents qui ont été consultés par d'autres membres du groupe et les documents qui apparaissent dans la liste ordonnée d'autres membres du groupe. Cette approche permet d'éviter une situation où un document serait

<sup>108</sup> Il s'agit de la connaissance partagée dans ce contexte.

consulté par plusieurs participants dans la session collaborative. Mais, si cette approche peut être vue comme une bonne méthode de répartition des tâches, nous remarquons qu'elle empêche une collaboration efficace. L'une des caractéristiques de la collaboration est l'aspect mutuel du travail. La possibilité qu'ont plusieurs utilisateurs d'évaluer et d'interpréter le même document dans une session collaborative est une richesse et un véritable partage de connaissances. Mais la proposition de Foley rend cela impossible.

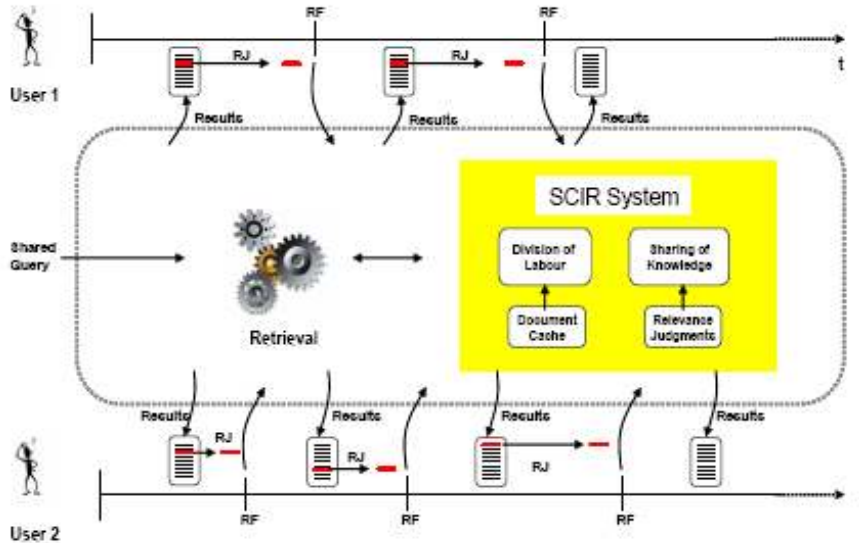


Figure 4.12 : Vue d'ensemble conceptuelle d'une session collaborative entre deux utilisateurs selon Foley (2009a, p.7)

#### 4.4.2.6 Le système RCI Web

Vivian et Dinet (2008, p.85-110) ont développé le système RCI Web qui est un outil d'aide à la RCI. Ce système permet d'une part à un ensemble de collaborateurs de fédérer leurs activités de recherche sur une thématique donnée, et d'autre part, d'améliorer la recherche individuelle. Nous nous intéressons à l'aspect collaboratif du système. Ce système consiste en un agent intégré dans un navigateur Mozilla qui fonctionne sur un moteur de recherche Google. Le responsable d'un groupe crée une thématique de recherche et les membres recherchent des informations autour de cette thématique. Chaque utilisateur peut noter les pages visitées et également visualiser les notes déjà attribuées par les membres du groupe. Ce système permet en plus d'afficher la liste de toutes les pages inventoriées sous une thématique de recherche donnée.



Figure 4.13 : Historique des pages visitées et notes (Vivian et Dinet 2008, p.95)

RCI Web dispose aussi d'une fonctionnalité permettant d'enrichir les résultats de recherche fournis par Google en intégrant à ceux-ci la moyenne des notes déjà attribuées et stockées dans l'historique de recherche pour chaque document. Ceci permet à l'utilisateur de savoir si un document a déjà été consulté (présence d'une note) par les membres du groupe. Il pourrait également inférer, sur la base de la note, si un document présente un intérêt pour la recherche en cours. La figure 4.14 montre un exemple de trois documents parmi lesquels deux sont notés. Une absence de note signifie que le document n'a pas été consulté. Le X montre que le document a déjà été consulté et qu'il ne présente que très peu d'intérêt. Une note entre 1 et 5 annonce la moyenne des notes qui a été attribuée au document.



Figure 4.14 : Enrichissement de résultats sous Google (Vivian et dinet 2008, p.96)

#### 4.5 Analyse des SRCI existants

La multiplicité des méthodes et des systèmes pour la RCI montre l'importance de faciliter la collaboration dans la recherche d'information. Nous pouvons, à partir des systèmes élaborés ci-dessus, souligner les fonctionnalités communes aux SRCI :

- Le partage du processus et du résultat de recherche est prédominant dans tous ces systèmes. Ce partage est généralement désigné par le terme « partage des connaissances ». Il comporte également le partage des commentaires et des évaluations parmi les collaborateurs.
- Le concept de conscience de groupe se manifeste dans certains de ces systèmes. Un tel mécanisme permet aux collaborateurs d'être informés des activités des uns et des autres lors d'une session collaborative.
- La synchronisation du travail sous une forme de WYSIWIS caractérise plusieurs de ces systèmes.
- La répartition des tâches permettant de diviser un travail à réaliser entre les collaborateurs.
- La communication sous forme de messagerie instantanée.
- L'intégration du tableau blanc pour faire du *brainstorming* est intégrée dans certains systèmes de navigation collaborative.



- La persistance du processus de recherche pour faciliter une collaboration asynchrone.

Remarquons qu'aucun système n'a pu intégrer toutes ces fonctionnalités dans un seul environnement. Notre but est donc de développer un SRCI qui intégrera toutes ces fonctionnalités et d'autres encore dans un seul environnement.

Nous pouvons classifier certains des systèmes présentés ci-dessus selon la taxonomie de Golovchisky et coll. (2008, p.1-3) (cf. 1.5.2). Cette classification est présentée dans le tableau 4.2.

Dimensions		
	implicite	explicite
Intention		GroupWeb, W4, CoWeb, CIRE, C-TORI, ARIADNE, METIORE, SearchTogether, RCI Web
Profondeur	algorithmique	interface
	SRCI de Foley, SearchTogether	SearchTogether METIORE
Synchronisme	synchrone	asynchrone
	ARIADNE, W4, CoWeb, C-TORI, GroupWeb, METIORE, SearchTogether, RCI Web,	CIRE, SearchTogether
Localisation	co-localisé	distribué
	Websplitter, Let's Browse	GroupWeb, W4, CoWeb, CIRE, C-TORI, ARIADNE, METIORE, SearchTogether, RCI Web

Tableau 4.2 : Classification des SRCI existants

Tous ces systèmes supportent une collaboration explicite. Certains d'entre eux permettent aux utilisateurs de coordonner leurs interactions, de répartir les tâches et d'assurer une conscience

de groupe à travers des interfaces telles qu'une interface de chat et un tableau blanc. D'autres automatisent ces fonctionnalités en implémentant des algorithmes de médiation. SearchTogether combine ces deux approches de coordination. Le système de Foley est un exemple typique d'une médiation algorithmique. La plupart de ces systèmes supporte une collaboration synchrone sauf CIRE qui est asynchrone. Certains comme SearchTogether permettent les deux. Les SRCIs facilitant une collaboration distribuée peuvent faciliter également une collaboration co-localisée. La plupart de ces systèmes peuvent donc supporter la collaboration entre les utilisateurs co-localisés ou distribués. Nos études des SRCI existants ainsi que du comportement informationnel collaboratif nous ont amené à développer un cadre conceptuel pour la gestion de la RCI. C'est ce cadre que nous présentons dans la section suivante.

#### **4.6 Le cadre conceptuel pour la gestion de la recherche collaborative d'information**

La collaboration implique le travail en commun avec un objectif partagé et nécessite beaucoup d'interactions entre les participants entraînant ainsi un haut niveau d'interdépendance entre leurs activités. Nous avons donc développé un cadre conceptuel permettant d'identifier et d'analyser tous les aspects importants pour la gestion et la facilitation de la recherche collaborative d'information à travers un environnement numérique.

Dans notre cadre, la gestion de la RCI fait intervenir quatre aspects principaux à savoir :

- la communication ;
- le mode de collaboration ;
- la coordination des interactions entre les utilisateurs ;
- la gestion des connaissances impliquées dans la collaboration.

En effet, nous considérons le cas où l'on communique pour collaborer, impliquant une coordination des interactions afin de pouvoir gérer les connaissances impliquées dans la collaboration. Nous appelons ce cadre 3CM (*Communication, Collaboration, Coordination, Management*) selon sa traduction de l'anglais. Nous présentons ci-dessous ces quatre éléments.

#### **4.6.1 La communication**

Pour qu'il y ait la collaboration dans le processus de résolution d'un problème informationnel, il doit y avoir de la communication entre les acteurs impliqués. La communication permet d'échanger des informations entre les collaborateurs. Au niveau technologique, elle nous permet d'envoyer un message provenant d'un utilisateur A à un utilisateur B. Au niveau humain, elle représente le message à faire passer d'une personne à une autre quelle que soit sa forme : gestuelle, textuelle, verbale ou audiovisuelle. Elle représente un aspect très important de la RCI. Le concept de communication a été abordé en détail dans la section 1.4.

Nous considérons la RCI comme étant une communication pour partager des connaissances. Comme le montre la figure 4.15, cette communication peut se faire selon trois modalités :

- entre deux ou plusieurs utilisateurs en passant par le SRI ;
- entre un utilisateur et le SRI ;
- entre un utilisateur et le répertoire collaboratif (un ensemble de ressources).

Un utilisateur, à partir de son interface (IU : interface utilisateur), peut envoyer une requête au SRI. Il peut également envoyer la même requête à un autre utilisateur. La requête pourrait aussi être stockée dans le répertoire collaboratif. Nous considérons les utilisateurs, le répertoire collaboratif et le SRI comme des agents ayant la possibilité de communiquer entre eux. Tout échange dans la collaboration passe par le processus de communication. La communication permet également de coordonner les interactions entre les collaborateurs. Par exemple, les collaborateurs peuvent répartir des tâches par un moyen de communication.



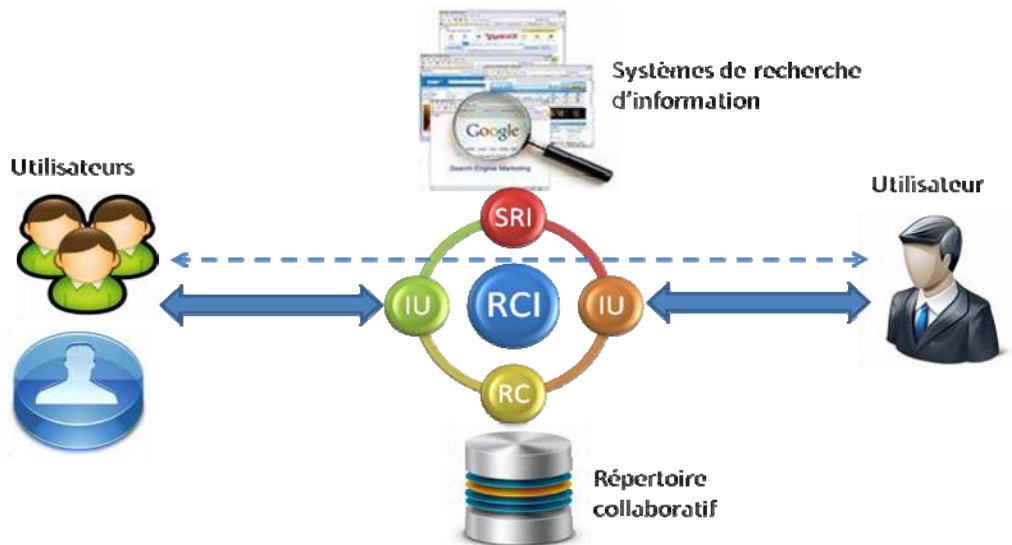


Figure 4.15 : Notre schéma global représentatif de la RCI

Nous considérons cet aspect par les questions suivantes :

- Pourquoi ? : Le « pourquoi » de la communication concerne le besoin de partager ou de collaborer qui est traduit en objectif de la communication.
- Quoi ? : Le « quoi » de la communication concerne son objet pouvant être de l'audio, du texte ou de la vidéo. Ces objets de communication sont des connaissances exprimées.
- Comment ? : Le « comment » de la communication concerne le style d'interaction pouvant être en mode synchrone ou asynchrone.
- Qui ? : Cela concerne l'expéditeur et le récepteur de l'objet de la communication (les collaborateurs).
- Quand ? : La date et l'heure de la communication pour pouvoir suivre l'évolution des utilisateurs et contextualiser les connaissances exprimées.

#### 4.6.2 Le mode de collaboration

Nous adoptons deux modes de collaboration entre les acteurs impliqués dans la RCI :

- le mode observation,
- le mode interaction.

Dans le mode observation, un ou plusieurs utilisateurs observent un autre utilisateur (expert ou non) qui recherche de l'information. Le mode observation est décrit par le graphe dans la figure 4.16.

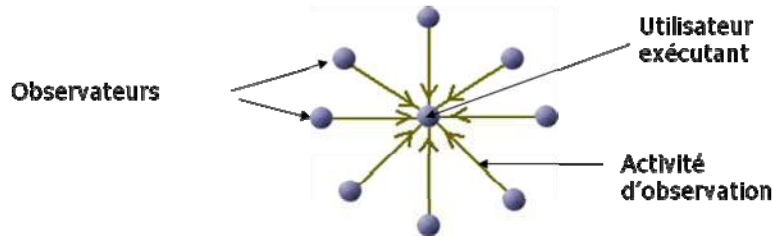


Figure 4.16 : Graphe d'observation (Odumuyiwa 2009, p.8)

Dans le mode interaction, deux ou plusieurs utilisateurs essayent conjointement de résoudre un problème informationnel. Ils partagent et échangent des informations ainsi que des compétences. Chacun contribue à la résolution du problème. Le graphe dans la figure 4.17 montre ce mode de collaboration.

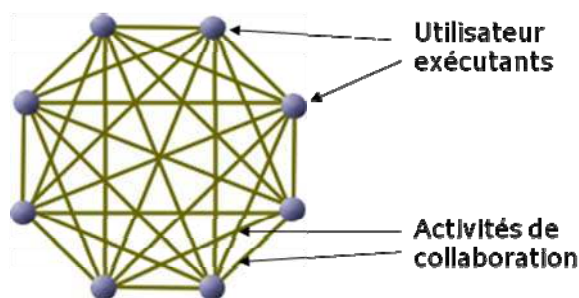


Figure 4.17 : Graphe d'interaction (Odumuyiwa 2009, p.8)

#### 4.6.3 La coordination des interactions entre les utilisateurs

La coordination dans le contexte de la RCI est la gestion de la dépendance entre les activités de recherche d'information des utilisateurs collaborant. Nous adoptons les trois catégories de dépendance selon Malone et coll. (1999, p.429-431) pour expliquer les dépendances qu'il peut y avoir entre un groupe d'utilisateurs lors d'une session de RCI:

- la dépendance de flux (*flow dependency*) ;
- la dépendance de partage (*sharing dependency*) ;
- la dépendance d'ajustement (*fit dependency*).

On parle de dépendance de flux quand les activités d'un utilisateur deviennent une ressource pour les activités d'un autre utilisateur. Par exemple, la requête formulée par un utilisateur peut devenir un point de départ pour un autre utilisateur ayant un besoin informationnel similaire.

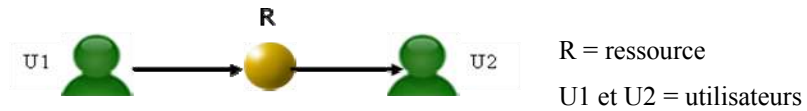


Figure 4.18 : Dépendance de flux

On parle de dépendance de partage quand deux ou plusieurs utilisateurs partagent une ressource. Si cette ressource est un résultat de recherche, les résultats de recherche contenus dans une liste ordonnée sont répartis entre eux. Cette dépendance s'exprime également quand deux utilisateurs partagent un écran d'une manière WYSIWIS.

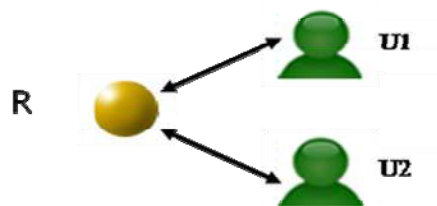


Figure 4.19 : Dépendance de partage

On parle de dépendance d'ajustement quand les activités de deux ou plusieurs utilisateurs doivent s'ajuster pour produire une ressource (par exemple : deux veilleurs travaillant ensemble pour résoudre un problème informationnel). La combinaison des requêtes formulées par deux utilisateurs en une seule requête peut être vue également comme une dépendance d'ajustement.

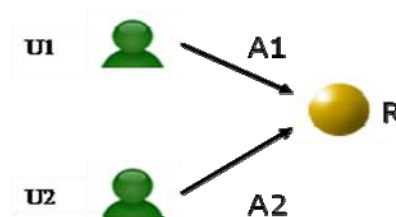


Figure 4.20 : Dépendance d'ajustement

Selon Malone et coll., ces trois catégories de dépendance ne sont pas exhaustives mais elles représentent le « noyau dur » dans lequel on peut toujours ramener tous les autres types de dépendances, en les analysant, soit par agrégation, soit par spécialisation (Malone et coll.1999, p.430 ; Lonchamp 2003, p.75). L'identification des dépendances possibles entre les activités des collaborateurs permet de développer des mécanismes de coordination pour les gérer. Dans notre prototype, nous avons implémenté plusieurs règles de coordination afin de gérer les interactions entre les collaborateurs.

#### **4.6.4 La gestion des connaissances impliquées dans la collaboration**

Lors de la RCI, les utilisateurs partagent des connaissances et en produisent de nouvelles. Un SRCI doit permettre de gérer les différents types de connaissances impliquées dans la RCI. Pour nous, les connaissances à partager incluent les requêtes, les résultats de la recherche effectuée, les évaluations faites, les annotations créées, les dialogues entre les utilisateurs et des documents partagés. Pour gérer ces connaissances, trois processus sont nécessaires :

- la modélisation de l'utilisateur ;
- l'acquisition des connaissances ;
- l'exploitation des connaissances.

Afin de capter les connaissances exprimées par un utilisateur, il faut d'abord modéliser ce dernier. Nous réalisons cette modélisation de l'utilisateur à partir de son profil et de ses activités. L'acquisition de connaissances concerne le stockage des différentes formes de connaissances exprimées lors d'une session collaborative. Le but principal dans la gestion des connaissances est de faciliter l'exploitation et la réutilisation de connaissances captées. Pour exploiter ces connaissances, nous adoptons le modèle EQuA<sup>2</sup>te (David 2002, p.34-36) signifiant en anglais :

- *explore* (explorer) : pour découvrir les objets du domaine d'étude ;
- *query* (interroger) : pour accéder à des objets du domaine d'étude à partir des connaissances déjà acquises sur les objets désirés ;
- *analyze* (analyser) : pour obtenir des informations à valeur ajoutée afin de découvrir des phénomènes du domaine d'étude ;

– *annotate* (annoter) : pour créer de nouvelles connaissances. Une annotation est vue comme une valeur ajoutée à l’information.

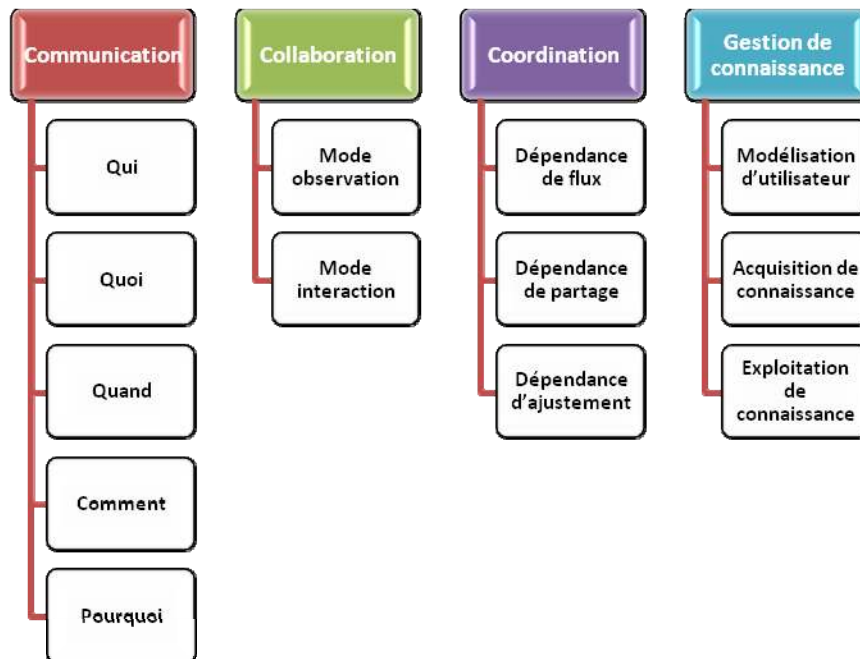


Figure 4.21: Cadre 3CM

#### 4.7 Conclusion du quatrième chapitre

Nous avons tenté, dans ce chapitre, de traiter le concept de la RCI. Nous sommes partis du comportement informationnel collaboratif pour démontrer la tendance de l'utilisateur de collaborer lors de la RI. Nous avons ensuite présenté ce que nous entendons par la RCI. Certains systèmes de navigation collaborative et systèmes de RCI ont été également présentés suivi par la présentation d'un cadre conceptuel pour la gestion de RCI. Ce cadre (3CM) a servi comme fondement pour notre compréhension de la RCI. Nous avons pu développer deux modèles de collaboration à la suite du cadre 3CM afin de faciliter la gestion de la RCI et le développement de SRCI. Ces deux modèles seront présentés dans le chapitre suivant.





# Chapitre 5

---

## Propositions en terme de modèles

### 5.1 Introduction

Dans le chapitre précédent, nous avons expliqué ce que nous entendions par la recherche collaborative d'information et notre cadre conceptuel pour la compréhension de la RCI. Nous avons vu jusqu'à présent que la RCI ne peut pas être confiée uniquement à une machine ou à un système quelconque, mais qu'elle nécessite des interactions explicites synchrones ou asynchrones entre les collaborateurs. De ce fait, nous avons développé un modèle de pyramide de collaboration qui montre les éléments nécessaires pour la réussite de la RCI et un modèle de communication pour le partage de connaissances lors de la RCI. Ces deux derniers modèles sont indépendants du domaine d'application et peuvent donc être utilisés ou implémentés pour faciliter la collaboration lors d'un processus de recherche d'informations dans tout domaine.

### 5.2 Le modèle de pyramide de collaboration

Nous nous sommes inspirés de la pyramide des besoins relevant du travail d'Abraham Maslow<sup>109</sup> sur la théorie de la motivation humaine pour élaborer un modèle de collaboration que nous appelons la pyramide de collaboration. Le modèle montre les différentes phases et composants nécessaires pour la réussite d'un processus de collaboration surtout dans le cadre de la recherche d'informations.

---

<sup>109</sup> Abraham Maslow est un psychologue américain. Il a élaboré une théorie de la motivation humaine à partir des observations réalisées dans les années 1940. Il parle de cinq niveaux de besoins des êtres humains. A l'origine, ces cinq niveaux ne furent pas représentés sous la forme d'une pyramide mais la représentation qui est désignée sous le terme « pyramide de besoins » s'est imposée dans le domaine de la psychologie du travail pour sa commodité. (Maslow 1943, p.370-396)



Pour que la RCI soit réussie, il faut un niveau minimum de confiance de la part des collaborateurs. Cette confiance est le socle de la collaboration. Pour qu'un utilisateur puisse passer de la recherche individuelle à la recherche collaborative, malgré les facteurs déclencheurs soulignés par Reddy et Jansen (2008, p.266-268) (cf. section 4.2.1), il lui faut avoir une confiance de départ lui permettant de partager son problème informationnel. Les détails qu'il livre à son collaborateur concernant son problème informationnel sont étroitement liés à la confiance qu'il manifeste à ce dernier.

La deuxième phase dans la collaboration est constituée de l'élaboration et de la clarification du problème informationnel autour duquel les collaborateurs travaillent. Il ne suffit pas seulement de partager un problème, mais il est important de faire passer les collaborateurs d'un problème partagé à une compréhension partagée du problème. Cette phase est une phase de concertation et de dialogue autour du problème informationnel afin que les collaborateurs puissent arriver à un consensus concernant leurs représentations du problème.

La communication qui est le troisième élément du modèle est très importante, car elle dépeint la phase d'échange des informations au sein et autour du groupe. Le partage de connaissances et la complémentarité de celles-ci permettent la mise en synergie des compétences. C'est le quatrième élément de la pyramide de collaboration. Quand des individus travaillent ensemble, ils cherchent toujours à avoir des connaissances sur les activités des uns et des autres, ceci afin d'assurer une conscience de groupe. La conscience du groupe est très importante dans une collaboration et elle constitue le cinquième élément de la pyramide. La collaboration comme nous l'avons expliqué dans le premier chapitre nécessite la répartition horizontale des tâches (cf. section 1.5.1). Pour réussir une collaboration, il faut un bon mécanisme de répartition des tâches qui permette aux collaborateurs de répartir les rôles quand ceci est nécessaire. La répartition des tâches est le dernier élément de la pyramide.



Figure 5.1 : La pyramide de collaboration

Notons également que la réussite d'une collaboration dépend à la fois de la bonne volonté des collaborateurs à partager et des technologies employées. Ce modèle souligne que la technologie doit être développée pour faciliter chaque élément de la pyramide. Nous précisons par la suite chaque élément de la pyramide.

### 5.2.1 La phase de confiance de départ

Tout acte de collaboration nécessite un climat de confiance dès le départ afin de pouvoir entamer la collaboration. Mais cette question de confiance repose sur différents facteurs. Selon le dictionnaire « Trésor de la Langue Française », la confiance est définie comme « la croyance spontanée ou acquise en la valeur morale, affective, professionnelle... d'une autre personne, qui fait que l'on est incapable d'imaginer de sa part tromperie, trahison ou incompetence ».

Pour faire confiance à quelqu'un, il faut déjà qu'il soit connu. Il faut également qu'il renvoie des signes rassurants. Ces signes peuvent être transmis par la parole, par l'apparence physique ou vestimentaire. C'est-à-dire que ce qui est en jeu à ce moment-là, ce sont les croyances et les représentations. La croyance est liée à l'idée qui est faite des qualités que doit posséder l'autre pour que l'on puisse lui accorder une crédibilité. Les représentations correspondent à l'idée qui est faite de l'autre dans une situation donnée. Ces éléments permettront d'évaluer la possibilité d'établir une collaboration avec telle ou telle personne. La notion de confiance repose aussi sur des critères identifiables tels que la compétence, la réputation, le fait que la personne fasse autorité dans un domaine. Si ces facteurs ne sont pas réunis, il ne sera probablement pas possible d'engager un processus de collaboration.

Dans un cadre de travail assisté par ordinateur, il est possible de s'appuyer sur des mécanismes qui permettent de repérer les compétences d'un acteur. Par exemple, il est possible de consulter l'historique des activités de recherche d'informations d'un utilisateur, pour mesurer son degré de compétence et donc se développer un niveau de confiance nécessaire à la collaboration.

Il nous semble important de préciser que la confiance est évolutive et qu'il convient de l'entretenir en permanence des éléments qui la fortifient ou la consolident. Dans le domaine qui nous préoccupe, l'un des éléments conditions indispensables pour maintenir le niveau de confiance concerne le fait de maintenir un haut degré de compétence et de communication entre les collaborateurs. Il est également important de pouvoir mettre les acteurs en synergie et les fédérer autour d'un problème partagé tout en répartissant les différentes tâches lorsque c'est utile.

### **5.2.2 La phase de compréhension partagée du problème informationnel**

Nous constatons que la plupart des travaux sur la RCI évoquent le fait que les collaborateurs travaillent autour d'un problème partagé. Malheureusement, un problème partagé n'implique pas forcément une compréhension partagée du problème. Toutefois, quelques articles ont montré qu'une même consigne pouvait donner lieu à de multiples représentations mentales selon les individus appartenant à un même groupe de travail (Dinet et coll. 2008, p.138 ; Dinet et Rouet 2002, p.133-161). C'est-à-dire qu'un problème informationnel partagé peut avoir des significations différentes pour chaque membre d'un groupe de travail. Si les collaborateurs ne

se retrouvent pas autour d'un consensus sur la représentation du problème partagé, il sera difficile de les mettre en synergie, car chacun travaillera selon la compréhension qu'il a du problème. Il nous apparaît donc comme primordial d'avoir une étape d'analyse et de clarification du problème par l'ensemble des collaborateurs, pour harmoniser les représentations. Nous considérons cette étape comme incontournable pour la réussite de la collaboration. Dans certains systèmes de collaboration existants, le processus de consensus est facilité par un mécanisme de tableau blanc. Ce mécanisme ne dispose pas d'une fonctionnalité qui permet de structurer les connaissances produites lors du processus et par conséquent une exploitation future de telles connaissances n'est pas possible. Dans le cadre de la RCI, puisque nous sommes intéressés par la production des connaissances et la capitalisation de celles-ci pour une réutilisation future, nous proposons une interface de définition et de clarification du problème qui permet aux collaborateurs d'exprimer leurs connaissances sur le problème à partager par le biais des attributs que nous définissons. Cette interface leur permet d'intégrer et de différencier leur compréhension du problème afin d'arriver à une compréhension partagée. Les attributs permettent également aux collaborateurs d'avoir un ensemble de propriétés sur lesquels ils peuvent concilier leurs connaissances.

### **5.2.3 La phase de communication**

Comme nous l'avons expliqué dans le premier et le quatrième chapitre, la communication est essentielle dans la collaboration. Toutes les autres phases de la pyramide de collaboration dépendent de la phase de communication pour échanger des informations. Pour partager des connaissances, il faut les exprimer, puis les communiquer. Pour assurer une conscience de groupe parmi les collaborateurs, un moyen de communication est nécessaire pour véhiculer les activités des utilisateurs des uns vers les autres. Pour la répartition des tâches, la communication est nécessaire pour faire connaître à chaque utilisateur la tâche qui lui a été attribuée. La communication est aussi très importante dans la phase de concertation et de dialogue autour du problème informationnel partagé. La confiance entre les collaborateurs peut être entretenue par un processus de communication efficace. Compte tenu de la centralité de la communication dans la collaboration, nous avons développé un modèle de communication pour la RCI. Ce modèle sera élaboré dans la section 5.3.

### 5.2.4 La phase de partage de connaissances

L'un des principaux objectifs de la RCI est de faciliter le partage de connaissances parmi les collaborateurs. Les utilisateurs peuvent partager leurs processus de recherche et leurs résultats de recherche. Pour réussir une collaboration, un bon mécanisme de partage de connaissances est nécessaire. Le partage de connaissances permet de mettre en synergie les compétences des acteurs collaborant. Deux utilisateurs possédant le même niveau de connaissance peuvent toujours partager des connaissances, mais le partage de connaissances sera plus efficace lorsque chaque utilisateur peut apporter quelque chose de différent. Ceci veut dire qu'une collaboration a une plus forte chance de réussir lorsque les collaborateurs possèdent des connaissances complémentaires. Par exemple, un utilisateur peut posséder un bon niveau de connaissances du domaine sur un sujet de recherche et un autre très compétent dans la méthodologie de recherche. La collaboration de ces deux utilisateurs autour d'un sujet de recherche produira une collaboration plus efficace, car leurs compétences se transformeront en connaissances complémentaires.

La connaissance comme expliquée dans la section 1.3.1 peut être tacite ou explicite. Dans la RCI le partage de connaissances concerne ces deux formes de connaissances. Donc, les quatre processus de conversion de connaissances selon Nonaka (cf. 1.3.2) s'appliquent dans ce domaine. Nous soulignons dans le tableau 5.1 les activités de la RCI qui permettent la conversion de connaissance d'une forme à l'autre.

<b>Tacite à Tacite (socialisation)</b>	<b>Tacite à Explicite (externalisation)</b>
<ul style="list-style-type: none"> <li>• communication interpersonnelle</li> <li>• partage synchrone d'interface de recherche</li> </ul>	<ul style="list-style-type: none"> <li>• a définition et la clarification du problème</li> <li>• notation</li> </ul>
<b>Explicite à Tacite (internalisation)</b>	<b>Explicite à Explicite (combinaison)</b>
<ul style="list-style-type: none"> <li>• visualisation de l'historique</li> <li>• consultation du résultat de recherche</li> </ul>	<ul style="list-style-type: none"> <li>• formulation de requête,</li> <li>• classification,</li> <li>•</li> </ul>

	<ul style="list-style-type: none"> <li>• création de métadonnées, étiquetage</li> </ul>
--	---

Tableau 5.1 : Les activités de la RCI facilitant la conversion de connaissance

- **Socialisation** dans la RCI : la communication interpersonnelle permet aux utilisateurs d’interagir d’une manière synchrone lors de la RI en échangeant des informations audio, visuelles ou textuelles. Elle facilite le partage d’expériences, ainsi favorisant l’acquisition de connaissances tacites. Le partage synchrone de l’interface de recherche permet à un utilisateur d’observer en temps réel les activités d’un autre utilisateur par le moyen de la technologie WYSIWIS. Ceci permet de communiquer ses compétences en temps réel.
- **Externalisation** dans la RCI : des connaissances tacites sont transformées en connaissances explicites lors de la définition et la clarification du problème informationnel par le processus d’annotation. La définition du problème permet d’élucider celui-ci. Le processus d’annotation facilite la création de connaissance dans la RI.
- **Combinaison** dans la RCI : lorsqu’un utilisateur formule une requête, cette dernière est une représentation de sa conception de son problème informationnel ainsi que celle de sa conception du fonctionnement de SRI qu’il utilise. Nous pouvons alors considérer la requête comme une connaissance explicite sur un problème informationnel défini. La création de métadonnées et d’étiquettes permet de créer des connaissances explicites à partir des connaissances explicites déjà exprimées dans un document. La classification et l’indexation de documents font partie des activités qui permettent la combinaison des connaissances explicites.
- **Internalisation** dans la RCI : la visualisation de l’historique de recherche permet un utilisateur d’acquérir de nouvelles connaissances à partir des expériences d’autres utilisateurs. Ces connaissances sont intégrées dans sa structure cognitive et deviennent ses propres connaissances tacites. La consultation du résultat de recherche facilite également le processus.

Il nous semble important de souligner les différents types de connaissances à partager lors de la RCI et les activités de recherche qui nécessitent leur partage.

#### ***5.2.4.1 Connaissances et compétences à partager***

- La connaissance du domaine : elle reflète le degré de compréhension qu'a un utilisateur sur son sujet de recherche. Elle décrit la connaissance des faits, des concepts, des terminologies dans un domaine spécifique. Une très bonne connaissance du domaine sur un sujet de recherche peut permettre à l'utilisateur de mieux clarifier son problème informationnel et de mieux juger la pertinence du résultat de recherche (Sutcliffe et Ennis 1998, p.330).
- La compétence en méthodologie de recherche d'information : elle concerne la capacité d'un utilisateur à planifier sa recherche en fonction de sa compétence en méthodologie de recherche d'information, ce qui relève de la stratégie de recherche dont les éléments sont les suivants : la sélection du système, la formulation, la modification et l'expansion de la requête de recherche (Zhang et Li 2005, p.1).
- La connaissance du système d'information : elle concerne la connaissance que l'on possède sur les fonctionnalités d'un système de recherche d'information. Les fonctionnalités peuvent être l'intégration d'un thésaurus, une liste de mots-clés, une bibliothèque des requêtes réutilisables, un support pour la formulation de requêtes booléennes, une visualisation du résumé du résultat, une visualisation du processus de recherche, etc.
- La connaissance des sources d'information : elle concerne la capacité d'un utilisateur à trouver des bases de données en ligne, des ressources web ou des systèmes de recherche d'information qui peuvent être utiles dans la résolution d'un problème informationnel.
- La connaissance des collaborateurs : il nous faut souligner que les utilisateurs disposent des types de connaissances ci-dessus à des degrés différentes. Pour collaborer, la première question est de savoir avec qui. En consultant les historiques et les modèles utilisateurs, on peut déterminer les différents niveaux de connaissance des utilisateurs. La capacité à explorer la base de connaissances (dans laquelle sont stockés les historiques et les modèles utilisateurs) pour découvrir des collaborateurs potentiels

n'est pas aisée. Cette capacité est aussi une connaissance que l'on peut partager. Nous retenons l'idée que la meilleure façon de trouver une bonne information est d'identifier la personne qui est susceptible de la connaître.

#### ***5.2.4.2 Activités de recherche d'information qui nécessite le partage***

- L'identification et la clarification du problème informationnel : la première activité dans la RCI est d'identifier le problème informationnel suivi de sa clarification. Tenant compte du fait que le niveau de connaissance du problème qu'a un utilisateur augmente avec le temps, un utilisateur, à un moment donné, peut se rendre compte de son état insatisfaisant de connaissance par rapport à son problème informationnel. Il peut, par conséquent, établir une collaboration avec un autre utilisateur afin de palier son manque de connaissance. Dans cette activité de collaboration, ce sont leurs connaissances du domaine sur le sujet de recherche qui sont partagées.
- L'articulation du besoin informationnel : l'identification et la clarification d'un problème informationnel mènent à l'articulation du besoin informationnel en formulant les buts de la RI à effectuer. Le partage de connaissances du domaine sur le sujet de recherche est aussi très important dans cette activité.
- Le choix de système d'information à utiliser : la spécification des buts de la RI, mène au choix des sources d'information à utiliser pour la recherche. Cette activité nécessite le partage de la connaissance des sources d'information ainsi que la compétence en méthodologie de recherche.
- La formulation, la clarification et la reformulation de requête consistent en l'expression du besoin informationnel en termes de requête pour interroger le SRI choisi. Ces activités nécessitent le partage de connaissances du domaine sur le sujet de recherche, le partage de connaissances du système d'information et le partage de compétences en méthodologie de recherche.
- L'évaluation des résultats : elle reflète le jugement de l'utilisateur sur la pertinence des informations retrouvées pour son besoin informationnel. Ce jugement est étroitement lié à sa connaissance du domaine sur le sujet de recherche. Un document peut être jugé pertinent à un moment par un utilisateur et non-pertinent à un autre moment par le même utilisateur vis-à-vis du même problème. Cela peut s'expliquer par le fait que le



niveau de connaissance du problème augmente avec le temps. Cela nous montre aussi l'importance de la collaboration dans l'évaluation des résultats de recherche d'informations, par exemple dans une situation où un expert du domaine aide un utilisateur à déterminer la pertinence d'un document par rapport à son besoin informationnel.

- La communication : pour partager des connaissances dans toutes les activités soulignées ci-dessus, il faut des activités de communication entre les utilisateurs collaborateurs. Cela est une réponse à la question « comment partager ces connaissances ».

### 5.2.5 La phase de conscience de groupe

La conscience de groupe permet de fédérer les acteurs sur la résolution du problème. Lorsque deux ou plusieurs personnes travaillent ensemble autour d'un problème, chacun génère une multitude de signaux<sup>110</sup>, soient explicites ou implicites. La possibilité de percevoir les signaux des uns des autres dans un groupe, soit traditionnel soit virtuel, permet une compréhension des actions et des intentions des membres du groupe. Dans le domaine de travail collaboratif assisté par ordinateur (TCAO), la « conscience de groupe » est le terme utilisé, pour désigner la connaissance qui résulte de la perception des signaux émis par les membres de groupe (Lonchamp 2003, p.199). La conscience de groupe permet aux collaborateurs de gérer le processus de collaboration tout en s'adaptant et planifiant leur comportement en fonction de leur compréhension des actions des uns et des autres. Dourish et Bellotti la définissent comme « une compréhension de l'activité des autres qui procure un contexte pour sa propre activité. Ce contexte permet d'assurer que les activités individuelles s'insèrent dans l'activité globale du groupe. Il permet également d'évaluer les actions individuelles à l'égard des buts et de l'évolution du groupe<sup>111</sup> » (Dourish 1992, p.1).

Gutwin et Greenberg (1999, p.9), à partir des travaux antérieurs existants (Adams et coll. 1995, p.85-104 ; Norman 1993 ; Endsley 1995, p.32-64), identifient quatre caractéristiques du concept de conscience de groupe :

<sup>110</sup> Les signaux dans ce contexte correspondent aux actions et inactions des membres d'un groupettes qu'un clic sur un lien, un manque d'activité pendant une période de temps, un défilement d'une page web etc.

<sup>111</sup> *Awareness is an understanding of the activities of others, which provide a context for your own activity. This context is used to ensure that individual contributions are relevant to the group's activity as a whole, and to evaluate individual actions with respect to group goals and progress* (Dourish 1992, p.1).

- la conscience de groupe est la connaissance de l'état d'un environnement délimité dans l'espace et le temps<sup>112</sup> ;
- les connaissances constituant la conscience de groupe doivent être mises à jour en fonction des transformations de l'environnement ;
- la conscience de groupe est entretenue par l'interaction des individus avec l'environnement<sup>113</sup> ;
- la conscience de groupe est pratiquement toujours liée à une activité. Le but principal de l'activité n'est pas d'entretenir la conscience de groupe, mais cette dernière contribue à l'accomplissement d'une tâche collective.

Dans une collaboration médiatisée au travers d'un système logiciel, les informations qui permettent d'assurer la conscience de groupe sont fournies par le système. Les mécanismes qui fournissent ses informations sont désignés sous le terme de mécanismes de conscience de groupe. Ces mécanismes accroissent la synergie du groupe lors d'une session collaborative. Lors d'une session collaborative synchrone, un bon mécanisme de conscience de groupe permettra aux membres du groupe de connaître les requêtes formulées par les uns et les autres. Celui-ci leur permettra également de connaître les documents consultés par leurs partenaires, les systèmes de recherche d'information utilisés, les annotations et les évaluations faites.

Plusieurs auteurs tels que Greenberg et coll. (1996, p.30), Steinfield et coll. (1999, p.83-85) et Liechit (2000, p.3-7) ont tenté de définir des typologies de la conscience de groupe. En nous inspirant de ces typologies, nous élaborons les différents types de conscience du groupe qui nous semblent pertinents dans une situation de recherche collaborative d'information :

- la conscience de la présence et de la disponibilité des autres (*presence awareness*)

---

<sup>112</sup> Traduction de l'anglais prise de Lonchamp (2003, p.201)

<sup>113</sup> Traduction de l'anglais prise de Lonchamp (2003, p.201)

permet d'avoir des connaissances sur la disponibilité et la présence des membres du groupe à un moment donné ;

- la conscience de l'activité des autres (*activity awareness*) recouvre toutes les connaissances sur les activités en cours dans l'espace de collaboration ;
- la conscience de l'espace de travail (*workspace awareness*) permet d'avoir des connaissances sur l'état de l'espace de travail partagé et sur les évolutions de celui-ci. La conscience de l'espace de travail peut être réalisée dans une collaboration synchrone par un tableau blanc ou un *éditeur* partagé. Lorsqu'il s'agit d'une collaboration asynchrone, elle peut être réalisée par un mécanisme de persistance de l'espace de travail ;
- la conscience du contexte (*contextual awareness*) s'applique dans une situation de collaboration là où les membres du groupe utilisent différents dispositifs de travail tels qu'un ordinateur portable et un téléphone portable et qu'ils disposent de différentes bandes passantes. Elle permet de fournir aux membres du groupe des informations selon leur contexte technologique ou tout autre contexte. Elle permet de savoir quelle information fournir à un individu et comment elle doit lui être fournie.

### 5.2.6 La phase de répartition des tâches

Lors d'une collaboration, on trouve des situations qui nécessitent que les collaborateurs changent de rôles. Comme expliqué dans la section 1.5.1, il s'agit d'une répartition momentanée des tâches qui s'applique temporairement. Reprenons l'exemple utilisé en 1.5.1 : lors de la recherche collaborative d'informations, un partenaire consulte les résultats issus d'une requête et un autre réfléchit sur la reformulation de la requête ou sur le besoin d'utiliser un autre système d'information. Après une suggestion pour essayer un autre système de recherche d'information, les deux partenaires continuent les activités de recherche. Il n'y a pas de rôle fixe dans ce type de répartition de rôle. Les partenaires peuvent changer de rôle toutes les cinq minutes. La répartition des tâches est importante dans la RCI car elle permet de réduire la redondance des activités. Elle permet aussi de gagner du temps en permettant à chaque collaborateur à un moment donné de se concentrer sur un aspect spécifique des tâches.

A l'heure actuelle, quand deux ou plusieurs utilisateurs, co-localisés ou distribués, recherchent des informations autour d'un problème partagé, la répartition des tâches est faite manuellement par les utilisateurs par le biais de la communication. Cette façon de répartir les tâches peut susciter une surcharge cognitive, car les collaborateurs en essayant de résoudre un problème doivent gérer et coordonner eux-mêmes leurs interactions. Le temps mis pour gérer ces interactions peut dépasser le temps réellement passé pour rechercher des informations. Développant un bon mécanisme de répartition des tâches peuvent accroître la réussite d'une collaboration. Morris et Horvitz (2007, p.8) dans leur système « SearchTogether » implémentent plusieurs mécanismes de répartition : messagerie instantanée, mécanisme de recommandation, *split-search* et *multi-search* (cf. 4.4.2.5).

### 5.3 Le modèle de communication pour la recherche collaborative d'information (COCIR)

Au cours de ce travail de thèse, nous réalisons que la communication, qu'elle soit interpersonnelle, communication homme-machine ou communication interpersonnelle par machine interposée, est l'un des éléments les plus importants dans la recherche collaborative d'information surtout quand il s'agit d'une collaboration synchrone. En fait, nous avons pris comme slogan : « l'on communique pour collaborer ce qui nécessite une coordination des interactions afin de pouvoir gérer les connaissances impliquées dans la collaboration ». Nous avons donc modélisé le processus de communication dans le cadre de la RCI. Ce modèle que nous nommons COCIR (*Communication model for Collaborative Information Retrieval*) est une représentation du contexte collaboratif pour le partage de connaissances. Il est composé de quatre éléments :

- l'utilisateur-expéditeur,
- l'objet informationnel,
- le contexte d'échange,
- l'utilisateur-destinataire.

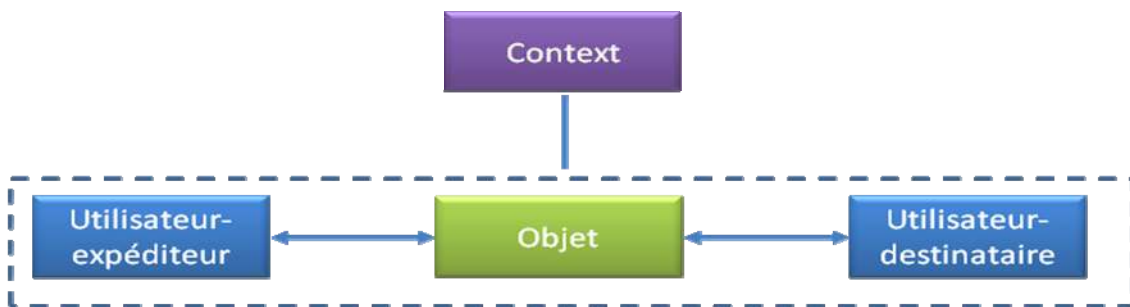


Figure 5.2 : Modèle COCIR

Chaque élément dans ce modèle est composé d'un ensemble de sous-éléments qui le caractérisent. L'expéditeur et le destinataire sont tous les deux des utilisateurs à qui nous pouvons attribuer les mêmes sous-éléments. Donc, un utilisateur peut jouer les deux rôles dans une session de RCI. De ce fait le modèle est composé de composants qui sont présentés dans la figure 5.3. Rappelons que dans la collaboration, tout échange passe par un processus de communication : pour partager des connaissances, il faut les exprimer sous forme d'objets informationnels et puis les communiquer ; pour assurer une conscience de groupe, il faut capter les activités des acteurs et puis les communiquer à chacun d'entre eux. Puisqu'il s'agit de la communication, il est aussi important de savoir quel objet informationnel est à communiquer et dans quel contexte est-il communiqué ?

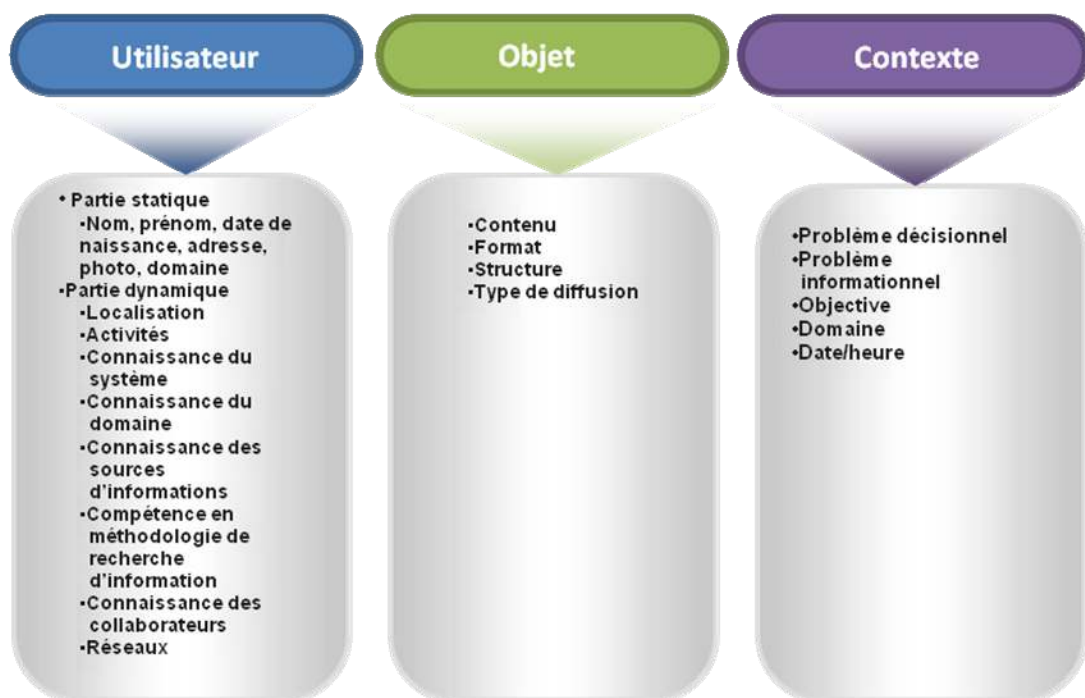


Figure 5.3 : Trois éléments du COCIR et leurs sous-éléments

L'une de nos préoccupations majeure dans ce travail est de réaliser un système de RCI qui facilite une collaboration synchrone et asynchrone dans laquelle les utilisateurs sont co-localisés ou distribués. De ce fait, deux niveaux de communication s'installent : la communication interpersonnelle par machine interposée et la communication inter-processus. D'une part les utilisateurs s'échangent des informations en utilisant le système – ceci correspond à ce que nous appelons la communication interpersonnelle par machine interposée. D'autre part, les applications s'échangent également des informations et cela correspond à ce que nous appelons la communication inter-processus. Pour assurer l'échange entre les machines, nous avons développé un protocole de communication entre les applications. Ce protocole sera expliqué dans le chapitre 6 (cf. section 6.2.2.2.2).

Nous détaillerons ainsi les trois éléments du modèle COCIR impliqués dans la communication humaine : utilisateur, objet et contexte. Notons également que notre modèle emprunte ces trois éléments des modèles de communication existants (cf. chapitre 1). Ce sont les sous-éléments que nous attribuons à chaque élément qui différencient notre modèle des autres modèles de communication. Ces sous-éléments reflètent particulièrement un contexte collaboratif.

### **5.3.1 Utilisateur**

Comme mentionné précédemment, l'utilisateur peut être l'expéditeur ou le récepteur de l'objet informationnel. La modélisation de l'utilisateur dans la RI a été évoquée dans la section 3.4. Nous avons également indiqué son importance dans la RCI (cf. 4.6.4). La modélisation de l'utilisateur permet de gérer les connaissances exprimées lors d'une session collaborative. Elle permet aussi de mettre en relation les utilisateurs ayant une représentation similaire. Comme nous l'avons élaboré dans la pyramide de collaboration (cf. 5.2.1), la confiance de départ nécessaire afin qu'un utilisateur entame un processus de collaboration avec un autre utilisateur est tributaire de la représentation qu'il fait de ce dernier. Un utilisateur peut s'inspirer du modèle de son partenaire stocké sur un SRCI afin de former sa propre représentation de celui-ci.

Pour pouvoir bien modéliser l'utilisateur, nous combinons l'approche implicite et l'approche explicite (cf. 3.4). Certaines informations sont explicitement fournies par l'utilisateur et d'autres sont déduites de ses activités. La partie statique de son modèle est aussi combinée avec la partie dynamique de son modèle pour avoir une représentation plus fine. Les sous-éléments que nous utilisons pour représenter l'utilisateur vont contenir eux-mêmes des sous-éléments.

Le modèle de l'utilisateur va contenir des informations telles que son identité, sa date de naissance, son adresse, sa photo, son domaine. L'attribut « localisation » concerne la localisation géographique où se trouve l'utilisateur lors de la collaboration. Elle peut être obtenue de l'adresse IP de l'ordinateur utilisé par l'utilisateur. Les « activités » de l'utilisateur comportent ses requêtes, les documents qu'il consulte, ses annotations, etc. La connaissance du système d'information, la connaissance du domaine, la connaissance des sources d'informations, la compétence en méthodologie de recherche d'information et la connaissance des collaborateurs sont obtenues à partir des activités de l'utilisateur. Le réseau va contenir des informations sur les utilisateurs avec qui l'utilisateur a déjà collaboré.

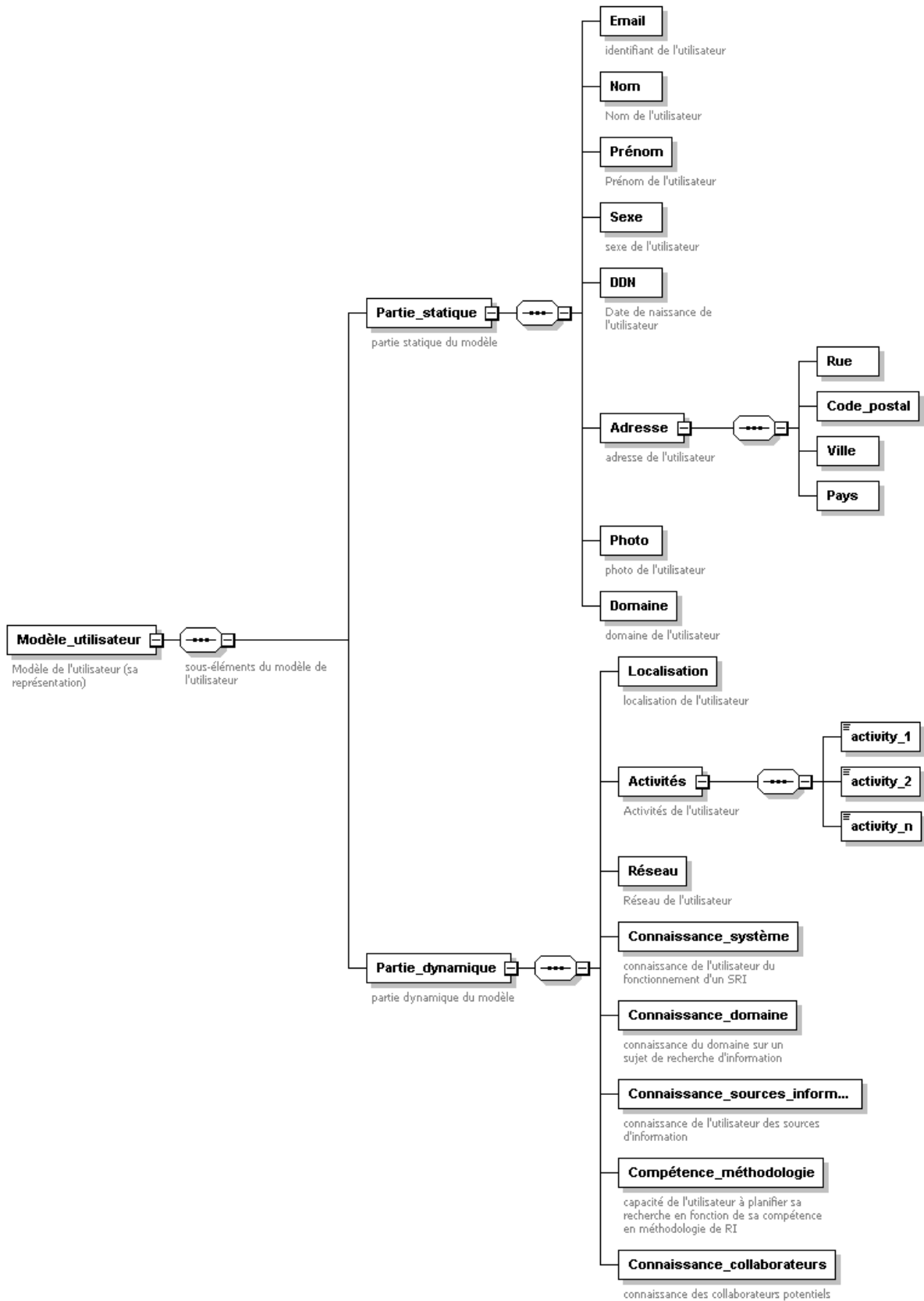


Figure 5.4 : Représentation de l'utilisateur



### 5.3.2 Objet informationnel

Le deuxième élément du COCIR est l'objet informationnel. Il représente toutes les informations échangées entre les collaborateurs qu'elles soient exprimées par eux ou qu'elles viennent d'autres artefacts existants. Quatre sous-éléments caractérisent l'objet informationnel : contenu, format, structure et type de diffusion. Le contenu représente le message à communiquer. Ce message peut prendre des *formats* différents. Il peut être de l'audio, du texte ou de la vidéo. Le *type de diffusion* peut être unidirectionnel quand il s'agit d'un objet envoyé par un expéditeur à un ou plusieurs destinataires sans attente de retour. Il peut également être conversationnel quand il s'agit d'un objet d'un expéditeur à un ou plusieurs destinataires avec un retour suite à l'expédition.

Dans la section 4.6.1, nous avons expliqué que la communication dans le cadre de la RCI peut se faire entre l'utilisateur et le répertoire collaboratif. Dans ce cadre, une information communiquée à un destinataire est aussi stockée dans le répertoire collaboratif. Notons également que toutes les connaissances exprimées lors de la collaboration sont stockées pour une réutilisation future. De ce fait, l'objet informationnel doit être structuré pour en faciliter le stockage et l'exploitation. L'attribut « structure » concerne donc la représentation dans la base de connaissances des informations sur l'organisation de l'objet informationnel. La *structure* varie selon le type d'objet à communiquer. Par exemple, l'un des objets informationnels à communiquer dans la RCI est l'annotation. Cette dernière peut porter sur un document web ou sur un problème informationnel. Comme le montre la figure 5.5 et la figure 5.6, la *structure* de l'annotation varie selon l'objet sur lequel elle est faite. La figure 5.7 montre une représentation d'un autre objet informationnel : un message envoyé par la messagerie instantanée. Cette représentation est différente de celles de la figure 5.5 et 5.6.

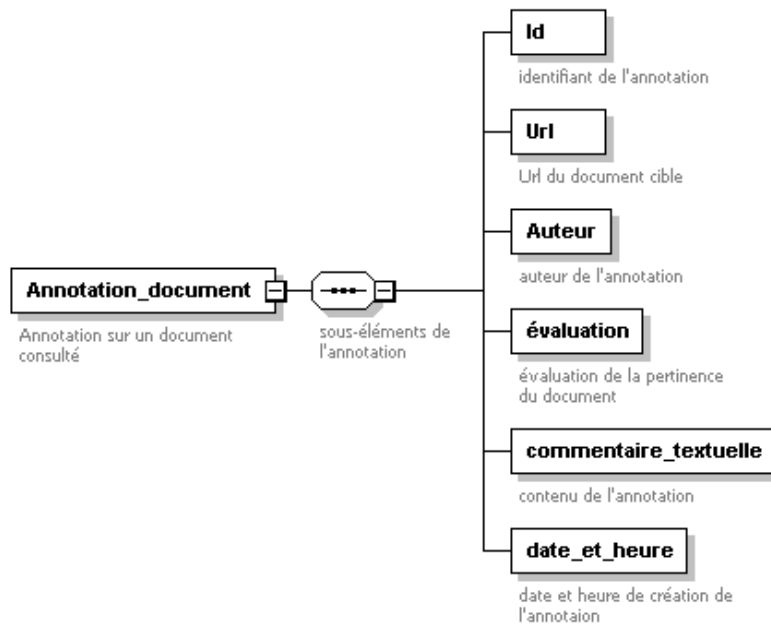


Figure 5.5 : Structure d’annotation sur un document web

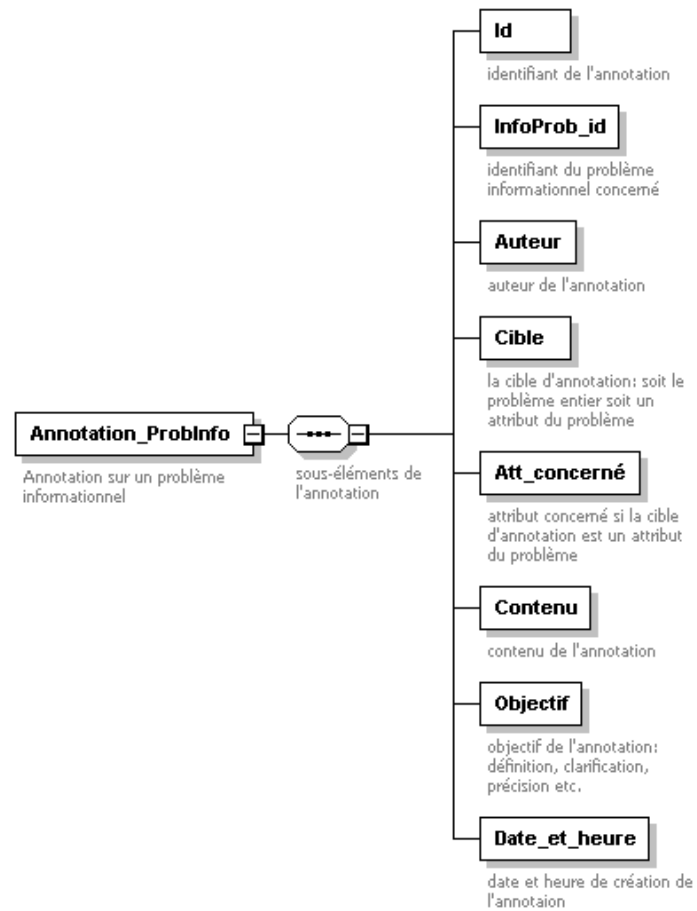


Figure 5.6 : Structure de l’annotation sur un problème informationnel

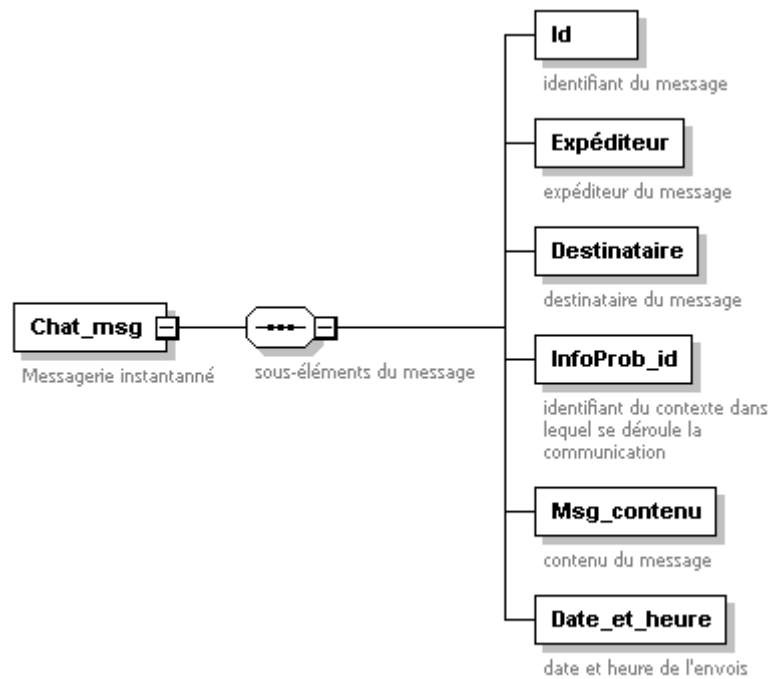


Figure 5.7 : Structure d'un message envoyé par la messagerie instantanée

### 5.3.3 Contexte

Le troisième élément du modèle COCIR concerne le *contexte* dans lequel se déroule la collaboration. Chaque échange dans la collaboration s'effectue dans un contexte particulier. Parce que notre domaine d'application est l'intelligence économique, le premier sous-élément de « contexte » est le problème décisionnel. Comme nous l'avons évoqué dans le deuxième chapitre, un problème décisionnel peut être traduit en problèmes informationnels. Les activités dans la RCI sont centrées sur le problème informationnel. Donc le deuxième sous-élément est le problème informationnel (PI) qui peut être un ensemble de PI.

Pour modéliser le problème décisionnel, nous adoptons le modèle MEPD (modèle pour l'explicitation d'un problème décisionnel) de Bouaka (2004, p.104-130). Nous sommes particulièrement intéressés par la représentation de l'enjeu d'un problème décisionnel, selon le modèle MEPD. L'enjeu représente ce que l'on peut gagner ou perdre dans une entreprise (Encyclopaedia universalis). L'enjeu est représenté par trois éléments dans MEPD :

- objet de l'enjeu, la cible sur laquelle le décideur souhaite agir ;

- signal de l'enjeu émis par l'objet, qui représente ce qui pousse le décideur à formuler un problème décisionnel ;
- hypothèse de l'enjeu, qui correspond à ce que nous pouvons déduire de la détection du signal. Elle comporte le(s) risque(s) perçu(s) et les conséquences (en termes de gains ou de pertes) envisagées si le décideur réagit ou ne réagit pas au signal.

Les trois autres éléments que nous utilisons pour représenter un problème décisionnel sont :

- l'expression du problème, il représente la demande que le décideur fait du problème perçu. C'est l'expression du problème dans le langage du décideur ;
- domaine, pour classer l'expression du problème décisionnel en sous-domaines éventuellement ;
- date/heure, comme éléments du contexte temporel au problème décisionnel.

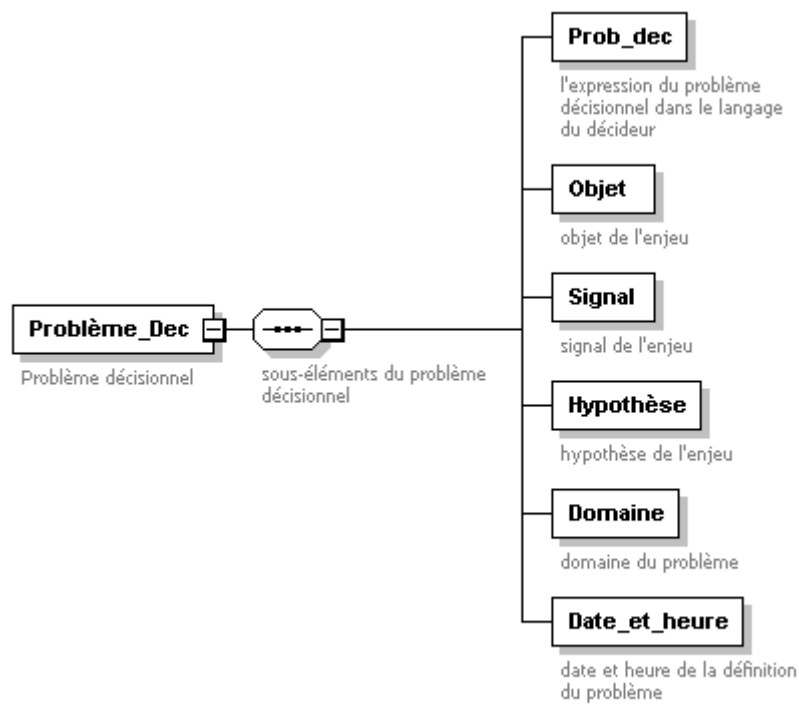


Figure 5.8 : Représentation d'un problème décisionnel

Tout problème décisionnel peut être traduit en problèmes informationnels. Il est possible qu'un problème informationnel soit exprimé sans une expression explicite du problème décisionnel duquel il est issu. Rappelons que notre objectif est de pouvoir appliquer le modèle COCIR dans la conception des SRCI quel que soit le domaine d'application. Ainsi, notre préoccupation majeure est de modéliser le problème informationnel autour duquel la collaboration est centrée. Cette modélisation sert à donner le contexte pour l'échange d'informations parmi les collaborateurs. Elle sert également à aider les collaborateurs d'élaborer mutuellement le problème à résoudre à travers des attributs prédéfinis dans le modèle. Les collaborateurs auront également la possibilité de spécifier leurs propres attributs lors de l'élaboration pour mieux clarifier le problème.

Un problème informationnel est représenté par les attributs suivants :

- identifiant  
du problème décisionnel : si le problème informationnel découle d'un problème décisionnel, un lien vers ce dernier est intégré par le biais de son identifiant ;
- identifiant  
du problème informationnel : permet d'attribuer un identifiant au problème informationnel ;
- problème  
informationnel : permet de décrire le problème en langage naturel. Il représente le besoin informationnel qui suscite une quête d'information. Celui-ci peut être flou et imprécis au début de la recherche. Il doit être clarifié et analysé. Un problème peut nécessiter sa décomposition en plusieurs problèmes informationnels lors du processus de clarification ;
- objectif :  
représente le « pourquoi » de la recherche. Ceci représente sur quoi les informations retrouvées seront appliquées. L'objectif de recherche aide à la clarification d'un problème informationnel. Il donne une direction suivant laquelle la recherche sera effectuée ;
- mots-clefs :  
les collaborateurs peuvent fournir des mots-clefs pour décrire le problème informationnel. Les mots-clefs fournis pourront être éventuellement utilisés pour formuler des requêtes ;

- sources  
d'informations : pour représenter les sources d'informations à consulter. Les sources d'informations peuvent être des systèmes ou des être humains ;
- domaine :  
permet aux collaborateurs d'attribuer le problème informationnel à un ou plusieurs domaines. Ceci aide à la clarification du problème informationnel pour restreindre l'interprétation sémantique des concepts employés tout en réduisant le problème lié à l'interprétation des concepts qui se trouvent dans plusieurs domaines, mais portant des significations différentes. Il facilite également la réutilisation des connaissances captées lors d'une collaboration.
- indicateurs :  
les collaborateurs peuvent définir d'autres attributs qui leur sont propres pour clarifier leur problème informationnel. Ces attributs et les valeurs associées sont désignés sous le terme indicateur. Les indicateurs peuvent inclure également des éléments que les utilisateurs recherchent dans des documents pour pouvoir juger leur pertinence. Les indicateurs sont définis en tant qu'attributs valeur ;
- date/heure :  
cet attribut permet de suivre l'évolution d'un problème informationnel car la définition de ce dernier peut évoluer au fil du temps grâce à l'augmentation de connaissances du domaine des collaborateurs.

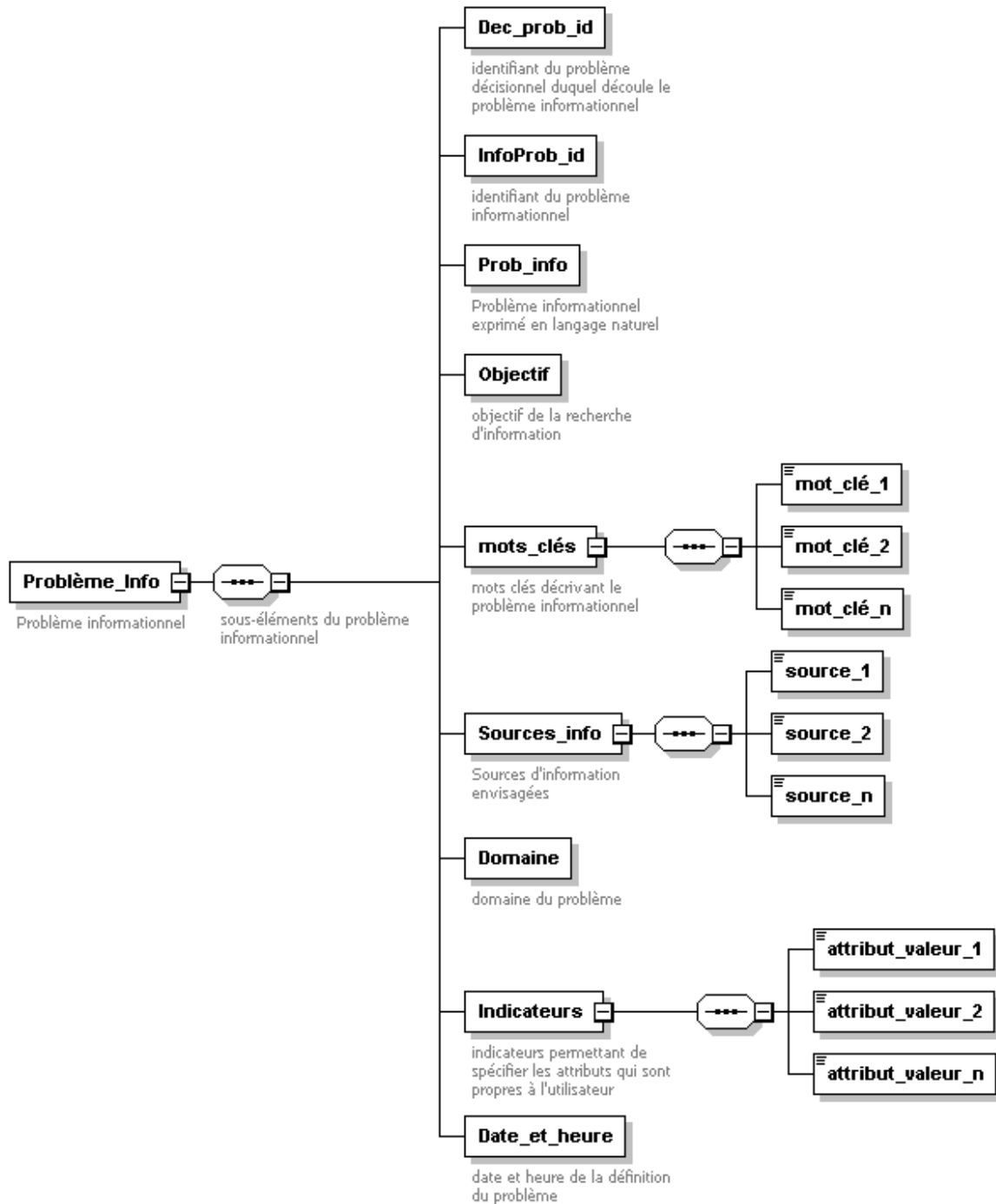


Figure 5.9 : Représentation d'un problème informationnel

#### 5.4 Conclusion du cinquième chapitre

Dans ce chapitre, nous avons présenté nos deux modèles : le modèle de pyramide de collaboration qui montre les phases nécessaires pour la réussite de la RCI et le modèle COCIR qui est une représentation du contexte collaboratif pour le partage de connaissances. Certains des éléments du modèle COCIR sont explicites et l'on peut leur attribuer des valeurs tandis

que d'autres éléments sont conceptuels et ne peuvent pas porter des valeurs explicites. Nous ne prétendons pas être exhaustifs par rapport à tous les attributs possibles pour représenter le processus d'échange entre des collaborateurs. En fait, notre but est de présenter des modèles qui peuvent être adaptés par n'importe quel concepteur de système de RCI. Les sous-éléments de trois éléments principaux du modèle COCIR peuvent être modifiés selon le besoin. Nous avons implémenté nos deux modèles dans le prototype MECOCIR que nous expliquerons dans le chapitre six.





# Chapitre 6

---

## Réalisation d'un SRCI et Expérimentation

### 6.1 Introduction

Dans le chapitre précédent, nous avons présenté nos apports en termes de modèles de collaboration. Pour valider nos propositions, nous avons développé un SRCI qui est une implémentation des méthodes pour gérer les activités de RCI. Ce système est fondé sur le cadre 3CM, les éléments de la pyramide de collaboration ainsi que le modèle COCIR. Nous commencerons donc ce chapitre par la présentation de l'architecture qui nous a permis de développer notre prototype que nous nommons MECOCIR. Nous présenterons également les composants logiciels du prototype. Le fonctionnement du prototype sera présenté tout en expliquant comment les collaborateurs recherchent des informations à l'aide de ce système. Nous décrivons enfin les deux premières séries d'expérimentations effectuées.

### 6.2 Développement du système MECOCIR

Le nom MECOCIR est dérivé du modèle COCIR et le système METIORE (cf. 4.4.2.4) qui est le premier SRCI développé au sein de l'équipe de recherche SITE au laboratoire LORIA. Les deux lettres « ME » sont tirées de METIORE, suivi de COCIR pour former le nom MECOCIR. MECOCIR est développé principalement en Java. Le choix de Java est lié à la capacité de ce dernier à faciliter la programmation réseau.

#### 6.2.1 Architecture du MECOCIR

MECOCIR est composé de plusieurs modules pour gérer les activités de collaboration parmi les utilisateurs. La figure 6.1 décrit une architecture générale du MECOCIR en montrant l'interaction entre les différents composants du système. Un utilisateur se connecte au système et puis définit son problème informationnel. Il dispose d'une interface de recherche lui permettant de se connecter à un SRI de son choix. Il peut envoyer des requêtes au SRI et ensuite recevoir des résultats.



Dans notre approche pour la recherche collaborative d'information, nous permettons aux utilisateurs d'utiliser des systèmes d'informations externes et stocker leurs résultats de recherche sur une base de connaissances interne. Puisque la RCI implique le partage du processus de recherche ainsi que du résultat de recherche, la base de connaissances est alimentée par les activités de recherche d'informations des utilisateurs. Le problème informationnel défini ainsi que les annotations sont stockés sur la base de connaissances. Les modèles des utilisateurs sont également stockés sur la base de connaissances. Toutes les requêtes formulées et les documents consultés ainsi que les évaluations faites sont captées et stockées dans la base de connaissances.

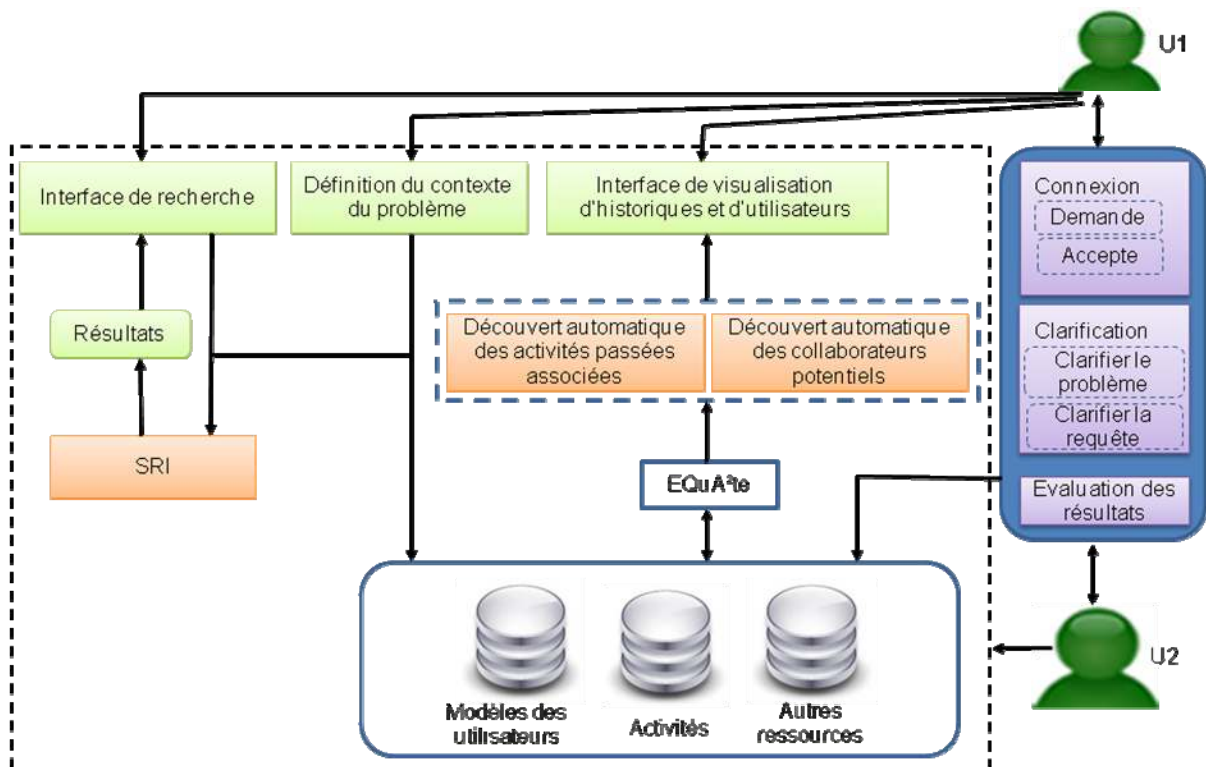


Figure 6.1 : L'architecture générale de MECOCIR

Nous avons défini un module qui permet une découverte automatique des activités passées associées au problème informationnel d'un utilisateur A comme le montre la figure 6.1. Ce ne sont pas seulement les activités passées qui nous intéressent mais aussi les utilisateurs qui les ont effectués. Cela permet aux utilisateurs, dans un premier temps de savoir avec qui ils peuvent collaborer et dans un deuxième temps d'avoir des connaissances sur les façons dont les autres utilisateurs ont pu aborder des problèmes similaires dans le passé. MECOCIR

facilite la visualisation d'historiques en implémentant le modèle EQuA<sup>2</sup>te afin de permettre aux utilisateurs d'exploiter la base de connaissances.

Un utilisateur peut travailler en autonomie, mais le but principal du système est de lui permettre de collaborer d'une manière synchrone avec un autre. Un utilisateur peut envoyer une demande de collaboration à un autre utilisateur en ligne. Si ce dernier accepte la demande, les deux utilisateurs entament une collaboration autour du problème informationnel défini par le premier utilisateur. Ils élaborent et clarifient ensemble le problème. Ils s'engagent dans un processus d'interaction pour aboutir à une compréhension similaire du problème informationnel que nous considérons comme « une compréhension partagée du problème informationnel ». À partir de là, les deux utilisateurs lancent des requêtes sur le problème. Ils partagent en temps réel leurs requêtes et les résultats de recherche. Ils évaluent mutuellement les documents retrouvés. Tous les échanges entre les utilisateurs pendant la collaboration sont stockés dans la base de connaissance. Ce sont les intelligences collectives<sup>114</sup> que l'on peut réutiliser.

Nous présenterons dans la section suivante le langage de programmation que nous avons utilisé pour réaliser MECOCIR.

### 6.2.2 Langage de programmation Java

Java est un langage de programmation orienté objet, inventé en 1995. C'est est un langage distribué, interprété, robuste, sécurisé, portable, à haute performance, *multi-threaded*, dynamique et plateforme-neutre. Selon son créateur, James Gosling, « la chose que Java essaie de faire et qu'il a remarquablement réussi est de s'étendre à beaucoup de domaines différents... »<sup>115</sup> (Cadenhead et Lemay 2007, p.9). Java a été conçu dès le début avec un but principal de faciliter la programmation réseau. Cet objectif est resté le même jusqu'à aujourd'hui. On trouve que la capacité de programmer en Java devient indispensable pour les

---

<sup>114</sup> L'intelligence collective renvoie à une intelligence partagée ou à l'intelligence du groupe ou à l'intelligence des connexions et des relations qui résultent d'une coopération ou d'une compétition de plusieurs individus (Zara 2004, p.8).

<sup>115</sup> "The thing that Java tries to do and is actually remarkably successful at is spanning a lot of different domains, so you can do app server work, you can do cell phone work, you can do scientific programming, you can write software, do interplanetary navigation, all kinds of stuff..." – le créateur de Java, James Gosling, interviewé par SearchWebServices.com (Cadenhead et Lemay 2007, p.9).

ingénieurs informatiques d'aujourd'hui. Nous avons choisi Java comme langage de programmation pour développer notre prototype parce que :

- il est plateforme-neutre : une application développée sur Windows peut s'exécuter sans problème sur Linux ;
- il est orienté réseau : il dispose d'un package « java.net » permettant de communiquer sur un réseau.

Ayant présenté le langage de programmation de notre choix, nous présenterons par la suite les composants logiciels du système MECOCIR qui ont été développés au cours de ce travail de recherche.

### **6.2.3 Les composants logiciels du système MECOCIR**

Le schéma des composants fonctionnels du système MECOCIR est décrit par la figure 6.2. Ce schéma est composé de trois couches montrant les éléments du système à trois niveaux d'opération. La première couche est l'espace de travail pour les utilisateurs. Cet espace comporte toutes les interfaces qui permettent aux utilisateurs d'interagir avec le système et avec d'autres utilisateurs. La deuxième couche représente le noyau du système. Elle comporte tous les modules constituant le *business logic* du système. Elle sépare les interfaces d'utilisateur de la couche de données. La troisième couche est celle de données et de connaissances. C'est cette couche qui assure la gestion des connaissances produites lors des collaborations. Nous détaillerons chacune de ces couches dans les sections suivantes.

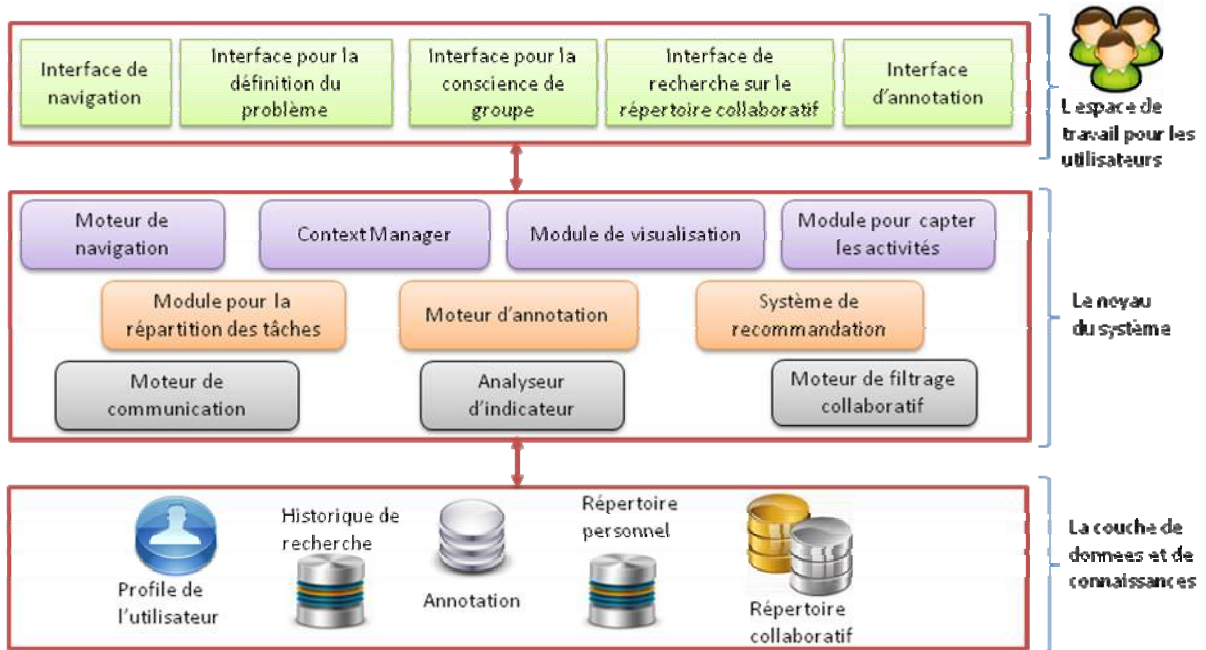


Figure 6.2 : Schéma des composants fonctionnels du système MECOCIR

### 6.2.3.1 L'espace de travail pour les utilisateurs

L'espace de travail pour les utilisateurs contient cinq interfaces :

- *interface de navigation*, permet de naviguer dans les pages web et de se connecter aux SRI en ligne ;
- *interface pour la définition du problème*, permet aux collaborateurs de définir, d'élaborer et de clarifier le problème informationnel autour duquel ils collaborent. Si ce problème découle d'un problème décisionnel défini, l'interface permet également aux collaborateurs de le spécifier. Cette interface implémente tous les sous éléments caractérisant le *contexte* de collaboration du modèle COCIR ;
- *interface pour la conscience de groupe*, permet aux utilisateurs de savoir en temps réel les activités de leurs partenaires. Toutes les informations nécessaires pour assurer la conscience de groupe sont affichées sur cette interface ;
- *l'interface de recherche sur le répertoire collaboratif*, permet d'interroger la base de connaissances et de rechercher des informations de la base d'historiques des utilisateurs ;
- *l'interface d'annotation*, permet de créer des annotations sur les objets informationnels.

### 6.2.3.2 Le noyau de MECOCIR

Dix composants logiciels constituent le noyau du système MECOCIR. Ces composants interagissent pour fournir aux collaborateurs les fonctionnalités soulignées dans la pyramide de collaboration.

#### 6.2.3.2.1 Le moteur de navigation

Ce composant est l'un des plus importants dans ce système. Dès le début de ce travail de thèse, nous remarquons que la plupart des SRCI existants permettent une collaboration centrée autour d'une base d'informations ou d'un moteur de recherche spécifique. Mais nous constatons que les utilisateurs ne se limitent pas à un seul SRI ou à un seul moteur de recherche ou à une base d'informations dans leur quête pour répondre à leur besoin informationnel. Ce comportement utilisateur nous a fait réfléchir sur une possibilité d'intégrer un navigateur web dans notre système. Un navigateur web permettra à un utilisateur de se connecter à autant de SRI en ligne qu'il le souhaite. Ceci lui permettra donc de choisir des sources d'informations convenables.

En revanche, ce choix nous a fait rencontrer beaucoup de difficultés dans le développement. Ces difficultés sont liées à l'intégration d'un navigateur web dans une application développée en java. Notons que Java ne dispose pas à l'heure actuelle d'un moyen permettant d'afficher correctement les pages web dynamiques. Le *JEditorPane* dont dispose Java est limité aux pages html. Une solution qui nous est apparue possible est de permettre à notre système d'afficher les pages web en se servant du moteur de navigateur web (Mozilla, Internet Explorer, Safari) existant sur l'ordinateur de l'utilisateur. Pour réaliser cela, une « communication inter-processus » est nécessaire entre notre application et le navigateur web. Nous avons eu recours à quelques API en java permettant de gérer cette communication inter-processus, mais ces API ont également leurs limitations.

Nous avons commencé le développement avec JDIC<sup>116</sup> qui est une API java pour afficher les pages web. Le problème majeur rencontré est lié au phénomène de *lightweight* et *heavyweight* en GUI (*Graphical User Interface*). Le GUI de notre application est développé avec *JSwing* qui est un composant *lightweight* mais JDIC est un composant *heavyweight*. Suite à l'échec de JDIC, nous avons utilisé une autre API qui s'appelle MozSwing. Cette dernière nous a permis

<sup>116</sup> <http://javadesktop.org/articles/jdic/index.html>



de résoudre le problème de *lightweight* et *heavyweight* mais d'autres difficultés ont émergé. Par exemple avec MozSwing, nous avons eu du mal à capter les activités de l'utilisateur.

Nous avons enfin découvert une API qui fonctionne très bien avec JSwing. Cette API nous permet de faire autant de manipulations que nous souhaitons. La seule contrainte est liée au fait que celle-ci n'est pas gratuite. Cette dernière s'appelle WebRenderer. Nous avons développé notre propre navigateur web en nous servant de WebRenderer pour gérer la communication inter-processus entre le navigateur par défaut et notre application. Ce composant est le moteur dorsal pour l'interface de navigation expliqué dans la section 6.2.2.1 ci-dessus.

#### 6.2.3.2.2 Module de communication

Ce module gère la communication entre deux environnements MECOCIR. Lorsqu'un utilisateur envoie une information à son partenaire, les deux ignorent comment l'information est transportée. Comme nous l'avons mentionné dans le chapitre cinq (cf. 5.3), le deuxième niveau de communication qui nous préoccupe dans ce travail est la communication inter-processus entre deux ou plusieurs lancements du système. Pour assurer cette communication, nous avons développé un protocole de communication pour gérer les échanges entre des processus lors de la RCI. Dans le cadre de communication entre deux individus humains, pour une communication effective, il faut qu'ils partagent le même langage, avec les mêmes sens attachés aux mots. Même s'ils communiquent avec un langage de signe, il faut que chaque signe signifie la même chose pour chacun. Par exemple, si un français écrit le mot « manger » sur un papier et puis le donne à un anglais qui ne comprend pas le français. L'anglais va le traduire comme « mangeoire ou crèche ». La même contrainte s'applique dans la communication interprocessus. Il faut que deux machines (ou deux applications ou deux processus) partagent les mêmes règles de communication afin de se comprendre. Ces règles de communication sont appelées « protocoles de communication ». Dans le cadre de notre travail, le protocole que nous avons développé sert d'une part pour la communication entre deux applications clients et d'autre part pour la communication entre une application client et une application serveur. Nous appelons ce protocole CPC<sup>117</sup> (Protocole de communication pour la RCI).

---

<sup>117</sup> *Communication Protocol for Collaborative information retrieval*

Deux types de prototypes de MECOCIR ont été développés. Le premier prototype est basé sur une architecture pair-à-pair où deux applications communiquent sans passer par un serveur. Dans ce prototype chaque application est à la fois un serveur et un client. Avec cette approche, nous avons été confrontés au problème du pare-feu<sup>118</sup>. Ceci nous a amené à développer le deuxième type de prototype qui est basé sur une architecture client-serveur. Dans les deux cas, le même protocole de communication est implémenté sauf que pour une architecture client-serveur, d'autres règles de communication sont ajoutées.

Les messages-types pour la communication entre l'application client et l'application serveur de MECOCIR sont :

- *ELEID*, permet à un client de se connecter au serveur et d'enregistrer ses paramètres d'identification sur le serveur ;
- *PRIVE*, permet d'envoyer un message au serveur qui le transmettra à un partenaire en intégrant l'identifiant de l'expéditeur;
- *GRPCO*, permet d'envoyer un message au serveur qui le transmettra à un groupe d'utilisateurs ;
- *REMCL*, permet à un client de se déconnecter du serveur.

Les messages-types pour la communication entre deux applications clients de MECOCIR sont :

- *MSGCL* signifie que le message envoyé est un message destiné à la messagerie instantanée ;
- *DIALOG* signifie que le message envoyé doit être interprété comme un dialogue<sup>119</sup> ;
  - *DEMAND* envoie une demande de collaboration à un autre utilisateur ;
  - *ACCEPT* montre qu'une demande a été acceptée ;

---

<sup>118</sup> « Pare-feu (*firewall* en anglais) désigne un dispositif de sécurité destiné à protéger un réseau interne informatique (entreprise, administration...) des attaques extérieures par un filtrage des informations provenant d'un réseau public (Web, messagerie électronique, news, etc.). Ce dispositif est paramétré en fonction d'une politique interne de sécurité qui définit une typologie des services externes accessibles de l'intérieur et, réciproquement, à quels services internes on peut accéder de l'extérieur » (Cacaly et coll. 2004, p.178).

<sup>119</sup> En informatique, une **boîte de dialogue** est un composant d'interface graphique constitué d'une fenêtre affichée par un programme ou par le système d'exploitation pour informer l'utilisateur d'un événement ou obtenir une information de l'utilisateur. Ces fenêtres sont appelées *boîte de dialogue* parce qu'elles établissent un dialogue entre l'ordinateur et l'utilisateur [Wikipédia consulté le 6 Mai, 2010].

- *CONTEXTCHANGE* est envoyé à chaque fois qu'un utilisateur change son contexte<sup>120</sup> lors d'une collaboration, pour signaler ce changement à son partenaire ;
- *STOPCOL* est envoyé pour signaler qu'un partenaire a arrêté la collaboration ;
- *SPLITSEARCH* signifie que ce message concerne la répartition des tâches ;
- *QUERY* signifie que ce message contient une requête formulée par un partenaire ;
- *DOCUMENT* signifie que ce message contient un document qu'un partenaire est en train de consulter.

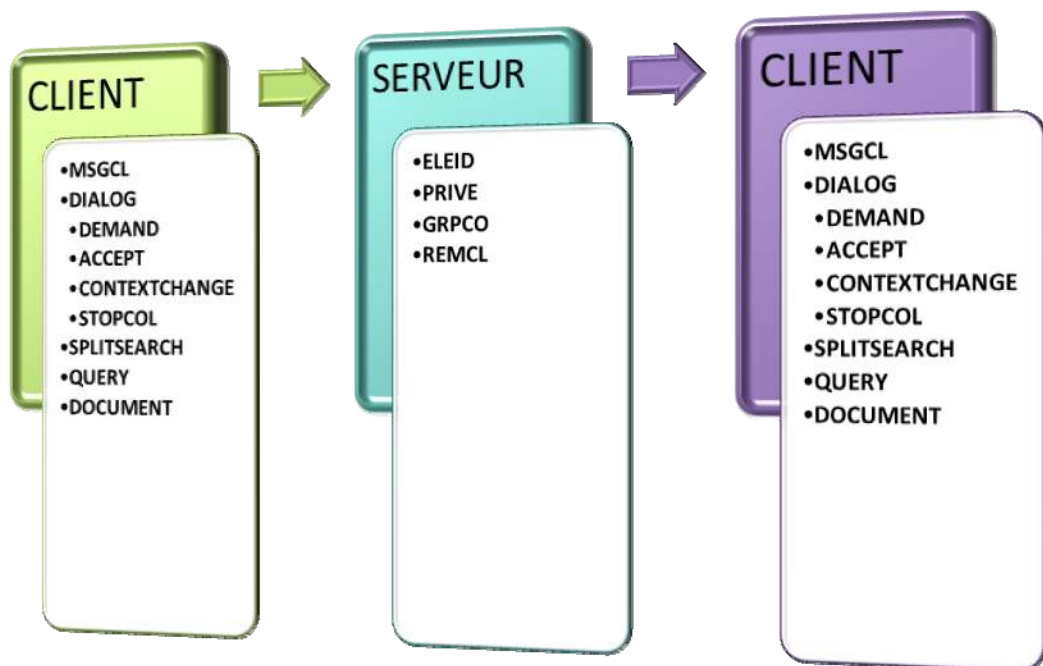


Figure 6.3 : protocole de communication pour MECOCIR

Tous les échanges entre deux applications passent par le module de communication, assurant ainsi la communication des connaissances exprimées par les collaborateurs et la conscience de groupe.

#### 6.2.3.2.3 Contexte manager

Ce module gère tout ce qui concerne le contexte de collaboration. Le *contexte manager* ne permet pas à un utilisateur de commencer la recherche d'informations sans définir d'abord le contexte du problème informationnel. Même si l'utilisateur veut reprendre un contexte qui a

<sup>120</sup> Contexte correspond au problème informationnel et l'objectif de recherche de l'utilisateur

été défini auparavant, il faut qu'il choisisse explicitement ce contexte pour que le système puisse lui permettre de commencer les activités de recherche d'informations. Le contexte manager gère également les informations qui apparaissent sur « l'interface pour la conscience de groupe »: lors d'une collaboration, si un utilisateur change son contexte, les informations qui seront affichées sur l'interface pour la conscience de groupe changeront.

#### *6.2.3.2.4 Module pour capter les activités*

Ce module permet de capter les activités et leurs attributs (qui, quand, quoi, pourquoi). Les activités sont les requêtes formulées, les documents consultés, le temps passé sur un document, le fait d'avoir imprimé un document, les messages instantanés échangés, les évaluations faites et les annotations créées. Les activités captées sont contextualisées en ajoutant des informations sur les utilisateurs qui les ont émises ainsi que le contexte dans lequel elles sont émises. Toutes ces informations sont stockées dans la base de connaissances.

#### *6.2.3.2.5 Moteur d'annotation*

L'annotation est l'un des moyens de communication par lequel les collaborateurs communiquent en utilisant le système MECOCIR. Les annotations créées doivent être structurées. Le moteur d'annotation dans notre système gère la contextualisation de la fenêtre d'annotation selon le type d'objet à annoter. Il s'occupe également du stockage de l'annotation et assure leur partage d'une manière synchrone entre les collaborateurs.

#### *6.2.3.2.6 Module de visualisation*

Ce module est le moteur dorsal de « l'interface pour la conscience de groupe ». Il permet de visualiser les informations sur la présence et la disponibilité des utilisateurs (*presence awareness*). Il gère également la visualisation des informations sur les activités des collaborateurs en temps réel.

#### *6.2.3.2.7 Module pour la répartition des tâches*

La répartition des tâches est l'une des fonctionnalités importantes pour la réussite d'une collaboration. Dans le système MECOCIR, nous permettons aux collaborateurs de se répartir les résultats de recherche. Par exemple quand deux utilisateurs se mettent d'accord sur une requête à lancer, un des deux peut lancer la requête en choisissant l'option « SplitSearch ». Ceci permet au système de répartir automatiquement les résultats de recherche entre ses deux utilisateurs pour que chacun puisse consulter une partie des documents retrouvés. Nous avons

développé un algorithme semblable à « round-robin strategy<sup>121</sup> » pour gérer la répartition des tâches entre les collaborateurs.

#### 6.2.3.2.8 *Analyseur d'indicateur*

Comme expliqué dans la section 5.3.3, les utilisateurs ont la possibilité de définir des attributs qui leur sont propres pour définir ou clarifier leur problème informationnel. Ces attributs et les valeurs associées sont désignés sous le terme indicateur. Les utilisateurs peuvent également définir des éléments ou les mots qu'ils souhaitent trouver dans les documents issus du résultat de la recherche. Chaque document consulté par un utilisateur est analysé par ce module afin de déterminer la fréquence et l'occurrence des indicateurs dans le document. Ce module traverse le DOM<sup>122</sup> de chaque document afin de trouver les indicateurs présents et de calculer leur fréquence. Ceci peut aider les utilisateurs dans le jugement de la pertinence d'un document.

#### 6.2.3.2.9 *Moteur de filtrage collaboratif et système de recommandation*

Ces deux modules fonctionnent ensemble. Le module de filtrage collaboratif est un moteur dorsal pour le module de recommandation. Il permet de découvrir un collaborateur potentiel en rapport avec un problème informationnel donné. Ce module calcule la similarité entre deux problèmes et la similarité entre deux utilisateurs. Ainsi, nous tentons d'identifier le réseau d'utilisateurs et les problèmes informationnels, en considérant le lien entre les utilisateurs et leur participation à la résolution d'un problème informationnel. Un utilisateur qui participe à la résolution d'un problème informationnel appartient au réseau du problème considéré. Nous nous posons les questions suivantes :

- Qui a exprimé un problème informationnel ?
- Qui a participé à la résolution du problème informationnel ?
- Quel rapport y a-t-il entre un nouveau problème exprimé et les problèmes résolus dans le passé ?
- Quel rapport y a-t-il entre les utilisateurs ?

---

<sup>121</sup> Le « round robin » est une manière de choisir les éléments d'un groupe tout en assurant que chaque élément est choisi et qu'aucun élément ne dispose d'une priorité plus élevée que les autres. En informatique, le round robin est un algorithme utilisé dans les systèmes d'exploitation pour répartir le temps du processeur entre chaque processus de manière à ce que chaque processus soit affecté du même temps

<sup>122</sup> Le Document Object Model (DOM) est un standard, recommandé par le W3C, qui permet à des programmes et à des scripts informatiques d'accéder aux éléments de documents HTML, XHTML et XML d'une manière standardisée et indépendante du langage de programmation ou de la plateforme utilisée (Wikipédia, consulté le 7 septembre 2010).

Dans la figure 6.4, U1, U2.....U9 représentent les utilisateurs du système ; P1, P2, P3 sont les problèmes informationnels résolus dans le passé ; P4 est un nouveau problème exprimé par U6. Il y a une similarité entre P2, P3 et P4. U5 est également similaire à U6. Pour rechercher un collaborateur potentiel qui peut aider U6 dans la résolution de son problème, nous calculons l'autorité de l'utilisateur dans le réseau ainsi que son expertise supposée dans le domaine du problème. Par exemple, dans le graphe, U3 a participé à la résolution de deux problèmes similaires<sup>123</sup> à P4 et la représentation de U3 est également similaire à celles de deux autres utilisateurs (U1 et U2) donc on peut considérer qu'il possède une certaine autorité dans le domaine et dans le réseau. U5 est similaire à U6 et il a participé à la résolution de P3 qui est similaire à P4 (le problème exprimé par U6) donc U3 et U5 peuvent être classés parmi les deux premiers collaborateurs potentiels pour la résolution de P4.

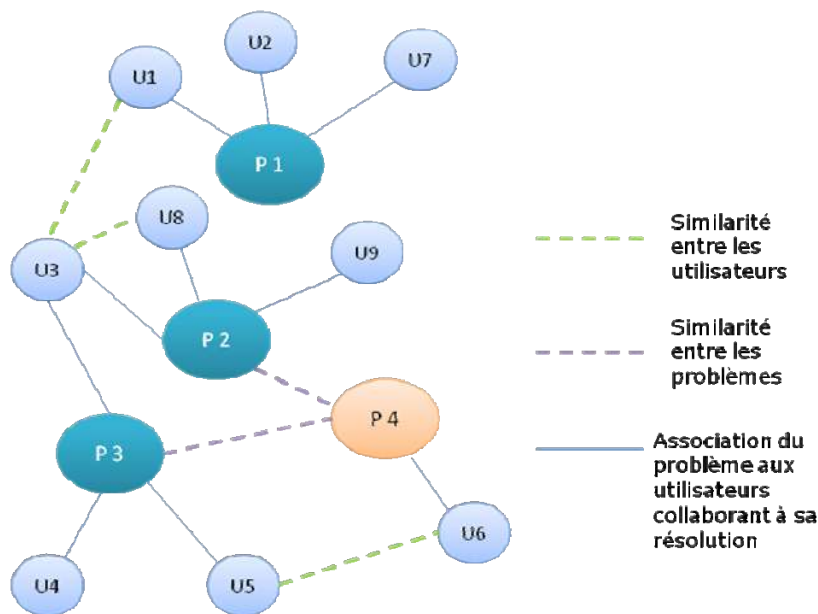


Figure 6.4 : Graphe montrant l'association entre les utilisateurs et les problèmes

### 6.2.3.3 La couche de données et de connaissances

Dans cette thèse, nous avons utilisé le terme « base de connaissances » pour décrire cette troisième couche de composants du système MECOCIR. Elle est composée de cinq bases de données :

<sup>123</sup> Le calcul de similarité est effectué en implémentant le modèle vectoriel traité dans la section 3.3.3.2

- Base de représentations d'utilisateurs : contient des informations sur la représentation d'utilisateurs.
- Base d'historique de recherche : les activités de recherche des utilisateurs sont stockées dans cette base. Ces activités comportent les requêtes formulées et les documents consultés.
- Base d'annotation : contient toutes les annotations créées par les utilisateurs du système.
- Répertoire personnel : lorsque les collaborateurs se répartissent les tâches, certaines informations sur les tâches attribuées à chaque utilisateur sont stockées sur sa machine en local. Le répertoire personnel permet également à un utilisateur de stocker les informations qu'il ne veut pas mettre à la disposition du groupe.
- Répertoire collaboratif : toutes les autres connaissances exprimées lors d'une collaboration sont stockées sur le répertoire collaboratif. Par exemple la communication interpersonnelle à travers de la messagerie instantanée est stockée sur cette base.

#### 6.2.3.3.1 MySQL

Pour la gestion de nos bases de données, nous avons choisi d'utiliser MySQL qui est un système de gestion de bases de données relationnelles (SGBD). Il existe également d'autres SGBD tels que Oracle, Sybase, Microsoft SQL server etc. mais les avantages associés à MySQL nous ont conduit à le choisir. Ces avantages sont liés au fait qu'il est gratuit, un logiciel source ouvert et rapide à deployer et à utiliser. Il est très souvent utilisé avec PHP dans une architecture 3-tiers pour créer et gérer des sites web dynamiques. Dans le cadre de notre travail nous l'avons utilisé avec java et PHP.

#### 6.2.3.3.2 JDBC (Java Database Connectivity)

Java ne communique pas directement avec un SGBD. La communication est faite par le biais de JDBC, une API fournie avec Java, permettant d'accéder à une base de données supportant SQL (*Structured Query Language*). L'avantage majeur de l'API JDBC est qu'il est conçu pour permettre à un programme de se connecter à n'importe quelle base de données en utilisant la même syntaxe. Ainsi, quel que soit le SGBD utilisé, la syntaxe pour accéder à une base de données ne change pas. JDBC est ainsi indépendant du SGBD. Ceci est un avantage

ajouté à la portabilité du code en java : une application développée en java est indépendante de la plateforme sur laquelle elle s'exécute. Nous pouvons migrer nos bases de données d'un SGBD à un autre sans avoir besoin de modifier l'application qui s'en sert.

L'API JDBC permet trois opérations :

- Etablir une connexion à une base de données,
- Envoyer des instructions SQL,
- Traiter les résultats.

L'architecture de JDBC décrite par la figure 6.4 montre que l'accès aux bases de données peut s'effectuer de deux façons :

- Un accès direct à la base de données : cette façon nécessite que l'on dispose d'un pilote JDBC adéquat. Les instructions SQL sont directement envoyées à la base de données qui renvoie, en retour, aussi directement les résultats. La base de données peut être installée sur la machine sur laquelle l'application java fonctionne. Elle peut être aussi bien installée sur une autre machine d'un réseau intranet ou internet.
- Un accès intercepté par un intergiciel : cette façon permet d'intégrer un intergiciel (une autre application ou serveur d'application) entre l'application java et la base de données. L'intergiciel reçoit les instructions SQL venant du JDBC et les traduit en protocole d'un SGBD. Les résultats en retour d'une interrogation transitent également par l'intergiciel. Il permet aussi d'accéder à plusieurs bases de données et dispose d'un avantage d'intégrer un contrôle d'accès.



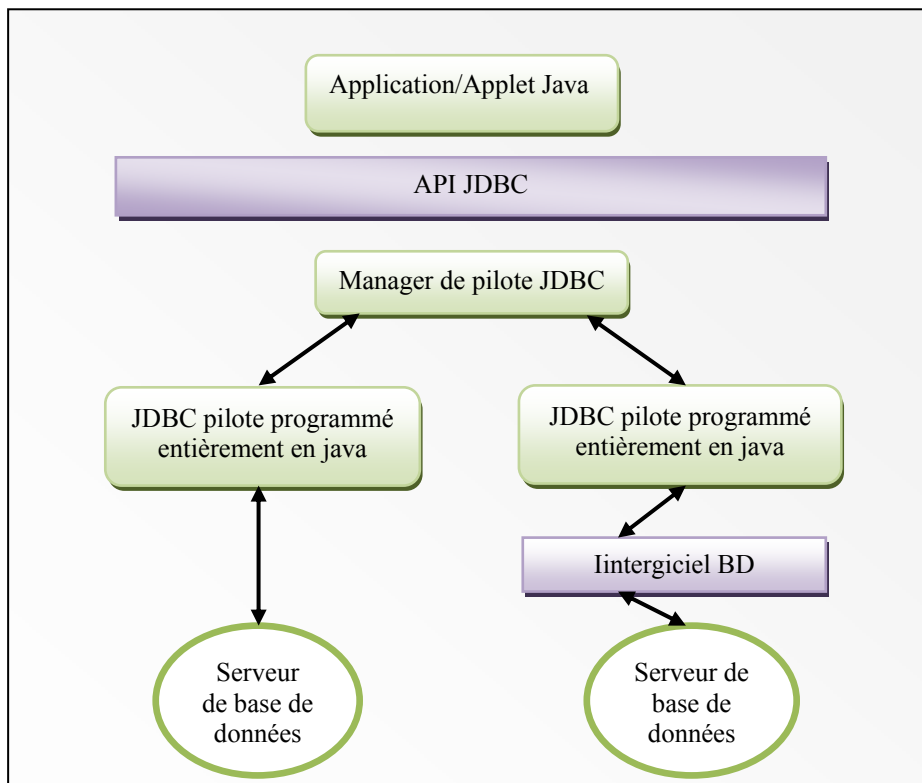


Figure 6.5 : Architecture JDBC (<http://java.sun.com/products/jdbc/overview.html>)

### 6.3 Le cadre d'application

MECOCIR permet deux cadres de fonctionnement pour collaborer dans la recherche d'information. Le premier cadre correspond à une situation où un utilisateur cherche quelqu'un avec qui il peut collaborer pour résoudre son problème informationnel. Le deuxième cadre concerne une situation où un groupe prédéfini d'utilisateurs travaille ensemble pour résoudre un problème informationnel partagé.

Pour le premier cadre, un utilisateur se connecte au SRCI, ensuite définit son problème informationnel selon les attributs donnés dans l'interface de définition du problème. Cette définition est fonction de la compréhension de l'utilisateur de son problème informationnel. Il peut passer à l'activité de recherche d'information en se connectant à un SRI et commencer à saisir des requêtes. Mais puisqu'il s'agit de la recherche collaborative et qu'il ne sait pas forcément qui choisir comme collaborateur, il peut exploiter la base de connaissance pour trouver des problèmes similaires à son problème et les utilisateurs qui ont eu ces problèmes et

qui ont pu les résoudre. Cette démarche consiste d'une part en la réutilisation des connaissances déjà exprimées pouvant être utiles et d'autre part à aider l'utilisateur à connaître les utilisateurs qui ont exprimé ces connaissances. Le but est de permettre à l'utilisateur d'entamer une collaboration directe avec les collaborateurs potentiels qu'il a trouvés et ce, afin de partager des connaissances tacites entre eux. Cet intérêt peut être justifié par cet extrait du cours de Le Dantec Christophe (2007) « *La mise en œuvre des fonctions et des processus cognitifs ne se fait pas à l'identique pour tous les sujets. Cependant, on observe régulièrement au niveau des performances, une relative stabilité intra-individuelle intra-tâche. De même, on peut observer une certaine stabilité intra-individuelle inter-tâches (notamment lorsque les tâches sont de nature semblable) alors que la mise en œuvre des fonctions et processus cognitifs ne se fait pas nécessairement à l'identique pour un même sujet dans ces différentes tâches* »<sup>124</sup>.

En trouvant un utilisateur avec qui il peut collaborer, l'utilisateur ayant le problème informationnel fait une demande de collaboration à l'autre utilisateur. Si celui-ci accepte la demande, le problème informationnel devient un problème partagé entre les deux utilisateurs et ils commencent alors un processus d'intégration et de différenciation de leur compréhension du problème pour aboutir à une compréhension partagée du problème informationnel. Dès que le problème est partagé, les processus de collaboration qui suivent sont identiques dans tous les deux cadres présentés ci-dessus. La différence qu'il peut y avoir entre les deux cas est liée aux cognitions sociales des collaborateurs. Nous supposons que les cognitions sociales sont plus développées dans un groupe prédéfini là où les membres du groupe se connaissent.

Partant d'une compréhension partagée du problème informationnel, les utilisateurs choisissent et se connectent aux sources d'informations. Cela pourrait être fait de trois manières :

- chaque utilisateur se connecte à différentes sources d'information et, après, partage les informations retrouvées. Leurs activités (les navigations et les requêtes) sont mémorisées, ce qui implique là aussi un possible partage.

---

<sup>124</sup> Le Dantec C., Psychologie différentielle PSP2DF4\_S4 - Interaction cognition / conation. Cours à l'université de Rouen 2006-2007. <http://www.univ-rouen.fr/servlet/com.univ.utils.LectureFichierJoint?CODE=1172696611874&LANGUE=0>. Site visité le 15 Avril, 2009.

- la deuxième façon est de permettre à un utilisateur d'observer un autre utilisateur en temps réel pendant la navigation ou la saisie de requête. Il suit les activités de l'autre et s'engage dans une communication interpersonnelle pour avoir des explications et clarifications de ses façons de procéder dans la recherche d'information. Il peut aussi donner son avis pour contribuer à ce que l'autre fait. Cela est un processus d'apprentissage dans la collaboration. Par exemple, un utilisateur plus aguerri dans l'utilisation d'une base de données spécialisée prend la conduite pendant que les autres observent.
- la troisième façon est de permettre à tous les collaborateurs de formuler simultanément des requêtes de manière collaborative. Il s'agit ici d'interaction plus prononcée. Ils se connectent au même système de recherche d'information, ils formulent, clarifient et reformulent les requêtes ensemble. Ils évaluent la pertinence des informations retrouvées ensemble. Ils fusionnent leurs requêtes et suggèrent des synonymes aux mots-clés pour pouvoir trouver plus d'informations pertinentes. Ils s'engagent dans une communication interpersonnelle et en même temps créent des annotations afin d'exprimer leurs connaissances sur des objets (une requête, un document retrouvé, une url d'une source d'information, etc.).

Dans les deux cadres de fonctionnement de MECOCIR, trois catégories de connaissances peuvent être captées afin d'aider les utilisateurs :

- à mieux connaître les autres utilisateurs (qui peuvent être des collaborateurs potentiels) ;
- à développer la confiance de départ nécessaire pour engager une collaboration ;
- à améliorer leurs cognitions sociales<sup>125</sup> ;
- à trouver des connaissances déjà exprimées qu'ils peuvent réutiliser dans la résolution de leurs problèmes ;
- à décrire leurs capacités métacognitives.

Ces trois catégories de connaissances consistent en :

- des connaissances et compétences initiales ;

---

<sup>125</sup> Le terme « cognition sociale » désigne un domaine de la psychologie qui étudie les processus cognitifs (perception, mémorisation, traitement, émotions, raisonnement, etc.) employés par un individu pour traiter les informations sociales. Ces processus sont impliqués dans la gestion des interactions sociales. La cognition sociale renvoie également à la façon dont un individu représente son environnement social et lui-même. Dans le cadre de la collaboration, elle renvoie aux représentations que l'on fait de soi et de ses partenaires.

- des connaissances et compétences mentales ;
- des connaissances et compétences mises en œuvre.

Les connaissances et compétences initiales sont celles spécifiées ou déduites des réponses de l'utilisateur à une série de questions lors de l'inscription à MECOCIR.

Les connaissances et compétences mentales sont exprimées lors de la définition, l'analyse et la clarification du problème informationnel. Elles montrent le niveau de compréhension d'un utilisateur de son problème informationnel. Elles décrivent aussi comment un utilisateur pense aborder et résoudre son problème. Chaque contribution à la définition, l'analyse et la clarification du problème à travers un processus d'annotation et de communication interpersonnelle fait partie de cette catégorie de connaissance.

Les connaissances et compétences mises en œuvre sont des connaissances associées aux activités de RI et aux interactions entre les utilisateurs pendant la résolution du problème informationnel. Elles portent sur les sources d'information utilisées, les requêtes saisies, les liens visités, les annotations faites, les utilisateurs contactés, les réseaux exploités, les documents retrouvés, les évaluations portées sur ces documents, etc. Ce sont des connaissances qui montrent comment les utilisateurs ont pu réellement résoudre le problème informationnel partagé.

#### 6.4 Utilisation de MECOCIR

Lors du lancement de MECOCIR, le premier écran qui s'affiche invite l'utilisateur à s'identifier. Si l'utilisateur est déjà inscrit, cette étape permet au système de récupérer de la base de données, des informations sur l'utilisateur pour contextualiser son environnement de recherche. Dans le cas où l'utilisateur ne dispose pas d'un compte MECOCIR, il peut en créer un en s'inscrivant sur le système.

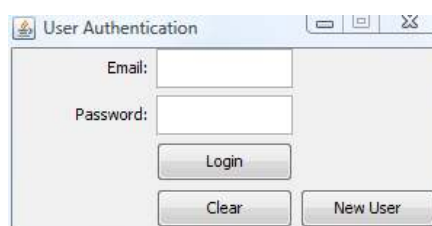


Figure 6.6 : Fenêtre d'authentification

### 6.4.1 Inscription

Pour s'inscrire au système, l'utilisateur remplit une succession de formulaires permettant au système d'acquiescer des informations sur lui. Le choix d'une succession de formulaires est fait afin de mettre l'utilisateur à l'aise et réduire le stress lié au fait de remplir un formulaire trop chargé. La plupart des attributs qui se trouvent dans les formulaires représentent les éléments du modèle COCIR. L'utilisateur commence en saisissant ses informations personnelles (figure 6.7) et puis son adresse (figure 6.8). Il remplit également son (ou ses) domaine(s) de compétence (figure 6.9) et enfin répond à une série de questions afin de capter ses connaissances initiales en méthodologie de recherche (figure 6.10).

The screenshot shows a 'User registration' window with the following fields and controls:

- Personal Information**
  - Surname:
  - First name:
  - Sex:
  - Date of birth**
    - Day:
    - Month:
    - Year:
  - Photo:
- Login Information**
  - E-mail:
  - Password:  Exactly eight characters
  - Confirm password:  Exactly eight characters
  - Secret question:
  - Answer to secret question:

Buttons:

Figure 6.7 : Formulaire d'inscription (saisir des informations personnelles)

The screenshot shows a 'User registration' window with the following fields and controls:

- Address**
  - Street address:
  - Complementary information:  (optional)
  - Postal code:
  - Town:
  - Country:

Buttons:

Figure 6.8 : Formulaire d'inscription (saisir d'adresse)

The screenshot shows a window titled "User registration" with a "Domain" section. On the left, there is a box labeled "Your chosen domain(s):" which is currently empty. In the center, there are two buttons: "Select" and "Remove". On the right, there is a list box titled "Choose your domain(s):" containing the following items: Accounting, Agriculture, Airline/Aviation, Alternative Dispute Re, Alternative Medicine, Animation, and Apparel and Fashion. Below the list box are "Add" and "Remove" buttons. At the bottom of the window, there is a text input field labeled "Add your domain if not included above:" followed by an "Add" button. Finally, there are "Back" and "Next" buttons at the very bottom.

Figure 6.9 : Formulaire d'inscription (saisir des domaines de compétence de l'utilisateur)

The screenshot shows a window titled "User registration" with a "Search Competence" section. It contains several multiple-choice questions:
 

- Are you familiar with internet search engine? (Radio buttons: No, Yes)
- Do you often use advance features of internet search engines? (Radio buttons: No, Yes)
- Are you familiar with online databases? (Radio buttons: No, Yes)
- Do you know the importance of quotation marks in query formulation on the internet? (Radio buttons: No, Yes)
- Do you use other information source apart from internet search engine? (Radio buttons: No, Yes)
- Have you other domain specific information ressource that you use in retrieving information? (Radio buttons: No, Yes)

 Below these questions is a section for query formulation:
 

- Will this query broaden or narrow the response result set? (Radio buttons: Narrow, Broaden, No idea). The "Query:" field contains "Information retrieval".
- Which of the following is a boolean query? (Radio buttons: chinese japanese food, chinese+japanese+food, chinese||japanese&&food, No idea)
- Which of the following is a truncation query? (Radio buttons: Victor\*, victor hugo, -victor -hugo, No idea)

 The final part of the form asks: "Given the information problem below, please list the keywords to represent the problem and form a query to capture them". It includes:
 

- An "Information problem:" text area containing: "I want to know about cassava procesing in africa but I don't want information from south africa"
- A text input field labeled "List your keywords here separated by comma:"
- A text input field labeled "Type your query here:"
- "Back" and "Next" buttons at the bottom.

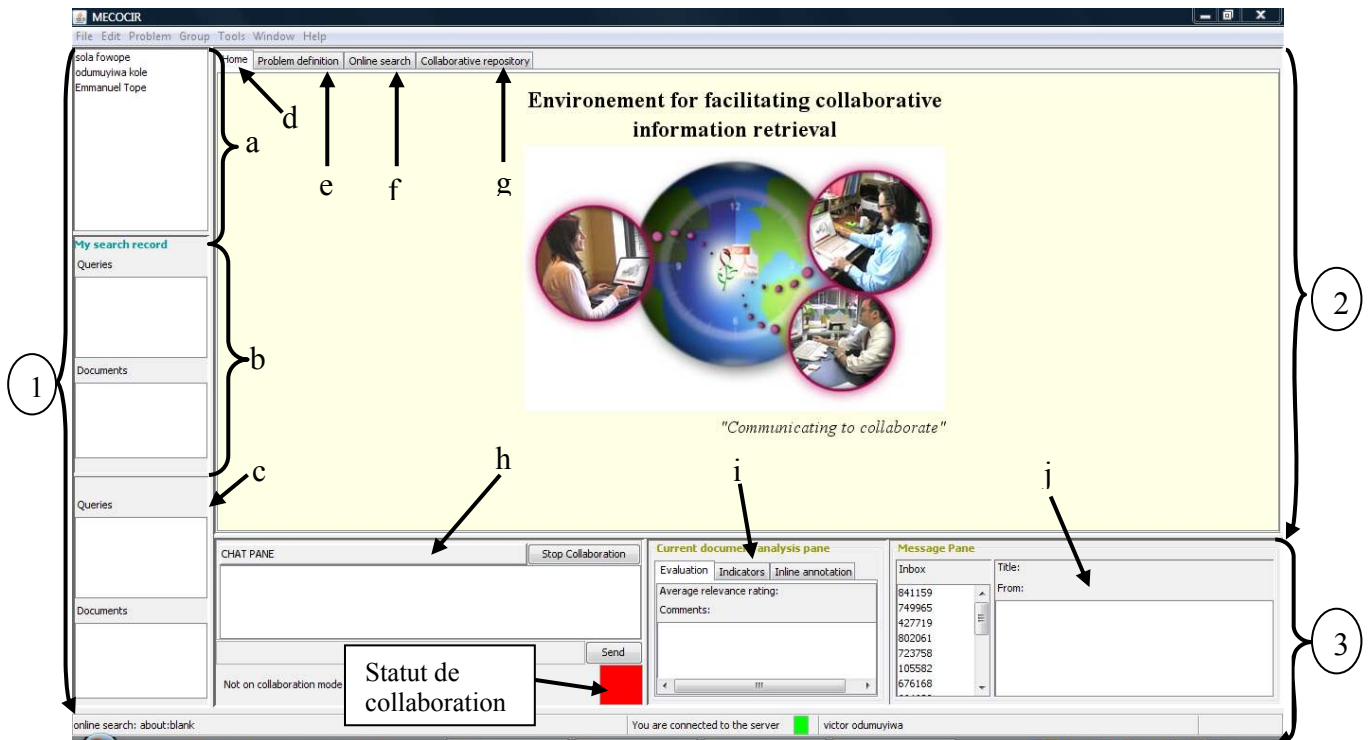
Figure 6.10 : Formulaire d'inscription captant des connaissances initiales de l'utilisateur

## 6.4.2 L'interface graphique principale de MECOCIR

Comme nous l'avons expliqué dans la section 6.2.2.1, l'espace de travail dans MECOCIR comporte plusieurs interfaces. Certaines de ces interfaces sont intégrées comme *bloc* dans

l'interface graphique principale présentée par la figure 6.11. D'autres s'affichent sur l'interface principale en fenêtres superposées.

Après l'identification de l'utilisateur, l'interface principale s'affiche avec la page d'accueil au milieu. Cette interface comporte une barre de menus et trois cadres (*frames*). La barre de menus permet à l'utilisateur d'accéder aux différents éléments du modèle COCIR ainsi qu'aux fonctionnalités fournis par le système. Par exemple, le menu « Problème<sup>126</sup> » donne accès à l'interface de définition du problème décisionnel et du problème informationnel ; le menu « Groupe » permet de spécifier le cadre de collaboration, soit en groupe ouvert ou soit en groupe fermé. Le menu « Outil » donne accès aux outils intégrés.



(1) L'interface pour la conscience de groupe : (a) bloc pour la conscience de la présence des membres de groupe, (b) bloc d'historique d'un utilisateur (c) bloc pour la conscience de l'activité d'un partenaire ; (2) Le cadre haut : (d) l'onglet de la page d'accueil, (e) l'onglet de la définition et de la clarification du problème informationnel, (f) l'onglet de l'interface de navigation, (g) l'onglet de l'interface de recherche sur le répertoire collaboratif

Figure 6.11 : L'interface principale de MECOCIR et la page d'accueil

La zone de gauche qui représente « l'interface pour la conscience de groupe » est composée d'un ensemble de trois blocs :

<sup>126</sup> Notons que les interfaces de MECOCIR sont en anglais. Pour faciliter la lecture de ce mémoire, nous utilisons les équivalents des termes en français pour présenter le système.

- bloc pour la conscience de la présence des membres de groupe (*presence awareness*) permettant de voir les utilisateurs en ligne ;
- bloc permettant à un utilisateur de voir son propre historique dans une session collaborative ;
- bloc permettant de voir l'historique d'un partenaire dans une session collaborative.

La zone en bas est composée également de trois blocs, chacun constituant une interface :

- l'interface de messagerie instantanée (*chat interface*) permet aux collaborateurs d'engager dans une communication synchrone lors de la collaboration ;
- l'interface d'analyse du document actuel, comporte trois onglets :
  - l'onglet « Evaluation » permet de voir les évaluations faites sur le document actuel que l'utilisateur est en train de consulter ;
  - l'onglet « Indicateurs<sup>127</sup> » permet de voir le résultat d'analyse des indicateurs faite par le module « analyseur d'indicateur » ;
  - l'onglet « Annotation<sup>128</sup> » permet de voir les annotations faites sur le document actuel sélectionné.
- l'interface de message asynchrone, permet de consulter des messages asynchrones qui sont envoyés à l'utilisateur. Lorsqu'un utilisateur est en pleine collaboration avec un autre, le système ne permet pas à d'autres utilisateurs de lui envoyer une demande de collaboration. En revanche, il peut recevoir des messages asynchrones provenant d'autres utilisateurs. Ce mécanisme fonctionne comme un système de courrier électronique.

La zone en haut est composée de quatre onglets :

- « Page d'accueil<sup>129</sup> », est la page d'accueil de MECOCIR. On y trouve des informations sur l'évolution du système, sur les versions disponibles, et sur le développeur ;
- « Problème informationnel<sup>130</sup> », permet d'avoir accès à l'interface de clarification du problème informationnel ;

---

<sup>127</sup> *Indicators*

<sup>128</sup> *Inline annotation*

<sup>129</sup> *Home*

<sup>130</sup> *Information problem*



- « Navigateur web<sup>131</sup> », permet d'accéder à l'interface de navigation web ;
- « Répertoire collaboratif<sup>132</sup> », donne accès à l'interface de recherche sur le répertoire collaboratif.

### 6.4.3 Définition du contexte de recherche

Pour commencer une activité de recherche d'information, l'utilisateur doit définir un contexte de travail pour commencer le processus que nous considérons une session de recherche. Ce contexte de travail correspond à l'élément « Contexte<sup>133</sup> » dans le modèle COCIR. Rappelons que le contexte comporte le problème décisionnel et le problème informationnel. Au démarrage, l'utilisateur peut créer un nouveau contexte ou reprendre un contexte existant. Pour créer un nouveau contexte, il clique sur le menu « Problème<sup>134</sup> » (voir figure 6.12) et puis sélectionne « Nouveau problème décisionnel<sup>135</sup> » pour définir son problème décisionnel. Une interface semblable à la figure 6.13 s'affiche permettant à l'utilisateur de définir son problème décisionnel selon les éléments du modèle COCIR. Il peut ensuite cliquer sur « Nouveau problème informationnel<sup>136</sup> » pour définir son problème informationnel. La figure 6.14 montre l'interface de définition du problème informationnel. Pour terminer, il clique sur « Définir les indicateurs<sup>137</sup> » pour définir les indicateurs associés au problème informationnel. Une fenêtre semblable à celle de la figure 6.15 s'affiche lui permettant de les spécifier.

---

<sup>131</sup> *Online search*

<sup>132</sup> *Collaborative repository*

<sup>133</sup> *Context*

<sup>134</sup> *Problem*

<sup>135</sup> *New decision problem*

<sup>136</sup> *New information problem*

<sup>137</sup> *Define indicators*

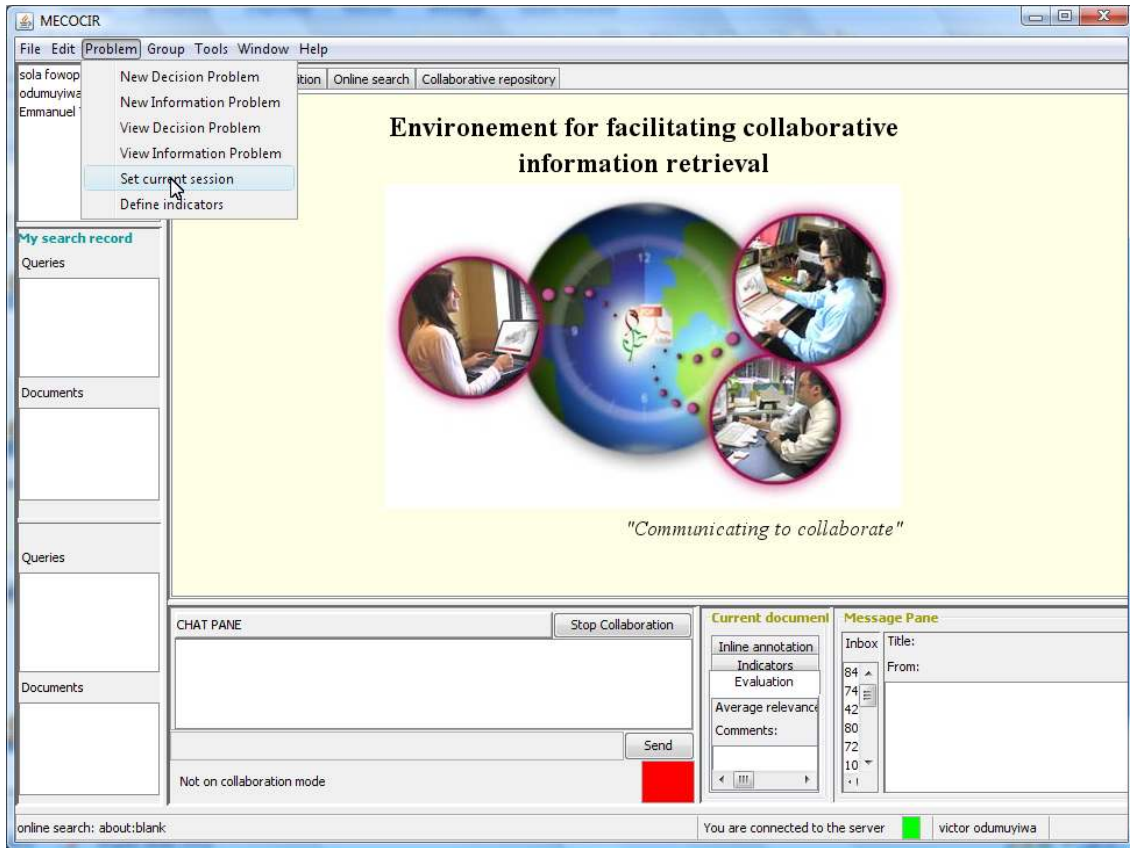


Figure 6.12 : Menu « problème »

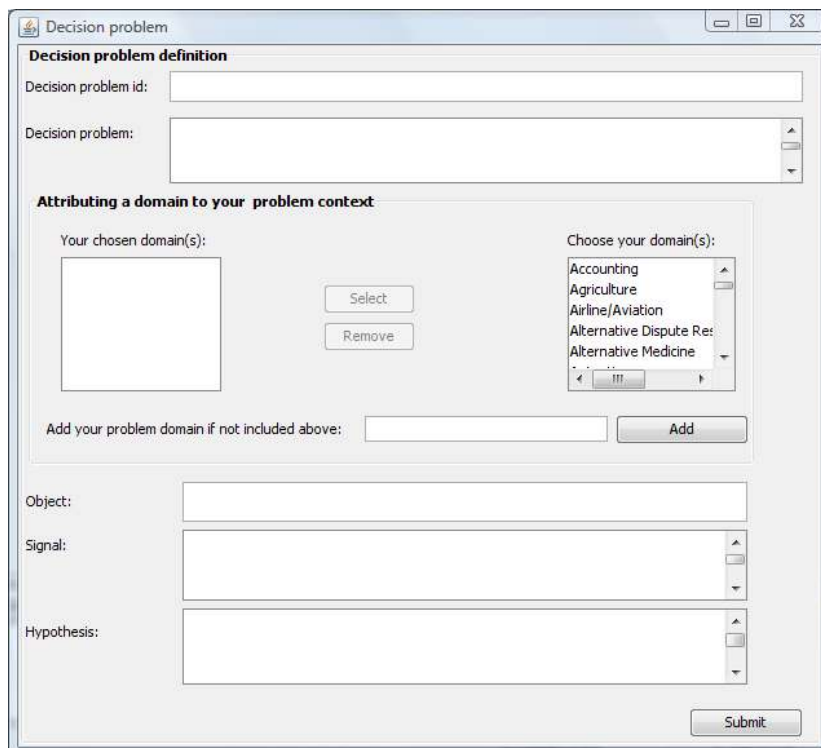


Figure 6.13 : Interface pour la définition du problème décisionnel

The screenshot shows a window titled "Information problem" with two main sections:

- Information Problem Context Definition:** This section contains several input fields:
  - Decision problem id: A dropdown menu currently showing "null".
  - Information problem id: A text input field.
  - Information problem: A text input field with a list icon on the right.
  - Objective: A text input field with a list icon on the right.
  - Keywords to be used for searching: A text input field with a list icon on the right.
  - Information ressources to be used: A text input field with a list icon on the right.
- Attributing a domain to your problem context:** This section includes:
  - A box labeled "Your chosen domain(s):" which is currently empty.
  - Buttons labeled "Select" and "Remove".
  - A list box labeled "Choose your domain(s):" containing the following items: Accounting, Agriculture, Airline/Aviation, Alternative Dispute Res, and Alternative Medicine.
  - An input field labeled "Add your problem domain if not included above:" with an "Add" button next to it.
  - A "Submit" button at the bottom right.

Figure 6.14 : Interface pour la définition du problème informationnel

The screenshot shows a window titled "Information problem" with the following elements:

- Information problem id: A text input field containing "Odeur épouvantable".
- A table with two columns: "Attributes" and "Values". The table is currently empty.
- Buttons labeled "Add more rows" and "Submit".
- A dropdown menu showing "1" and a "Create" button.

Figure 6.15 : Interface pour définir les indicateurs

Une fois le problème informationnel défini, l'utilisateur peut l'associer à la session actuelle. Il peut également à la session actuelle un problème informationnel parmi ceux définis antérieurement, auxquels il a participé. Pour définir la session de recherche, l'utilisateur

clique sur « Définir la session actuelle<sup>138</sup> ». Une fenêtre semblable à celle de la figure 6.16 s'affiche. Elle comporte une liste de tous les problèmes informationnels que l'utilisateur a définis ainsi que les problèmes informationnels auxquels il a participé mais qu'il n'en est pas l'initiateur. Il en choisit un pour la session actuelle.

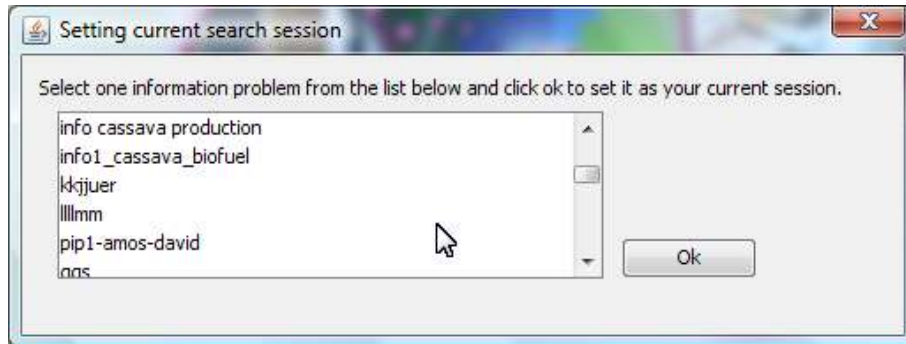


Figure 6.16 : Démarrage d'une session de recherche

#### 6.4.4 Demande de collaboration

Si l'utilisateur travaille dans un mode « groupe ouvert », il peut faire une demande de collaboration à un des utilisateurs en ligne. Il fait un clic droit sur le nom de l'utilisateur désiré et puis sélectionne « Collaborer<sup>139</sup> » du menu contextuel (voir figure 6.17). Le système envoie un message de demande à l'utilisateur concerné. Ce dernier apercevra un écran semblable à la figure 6.18. S'il accepte de collaborer, le système envoie un autre message à l'utilisateur qui a demandé la collaboration pour lui informer de l'acceptation. La figure 6.19 montre un message de notification d'une acceptation de collaborer. Après l'acceptation de collaborer, le système met à jour automatiquement le statut des deux utilisateurs de rouge en vert montrant qu'ils sont en mode collaboration.

<sup>138</sup> Set current session

<sup>139</sup> Collaborate

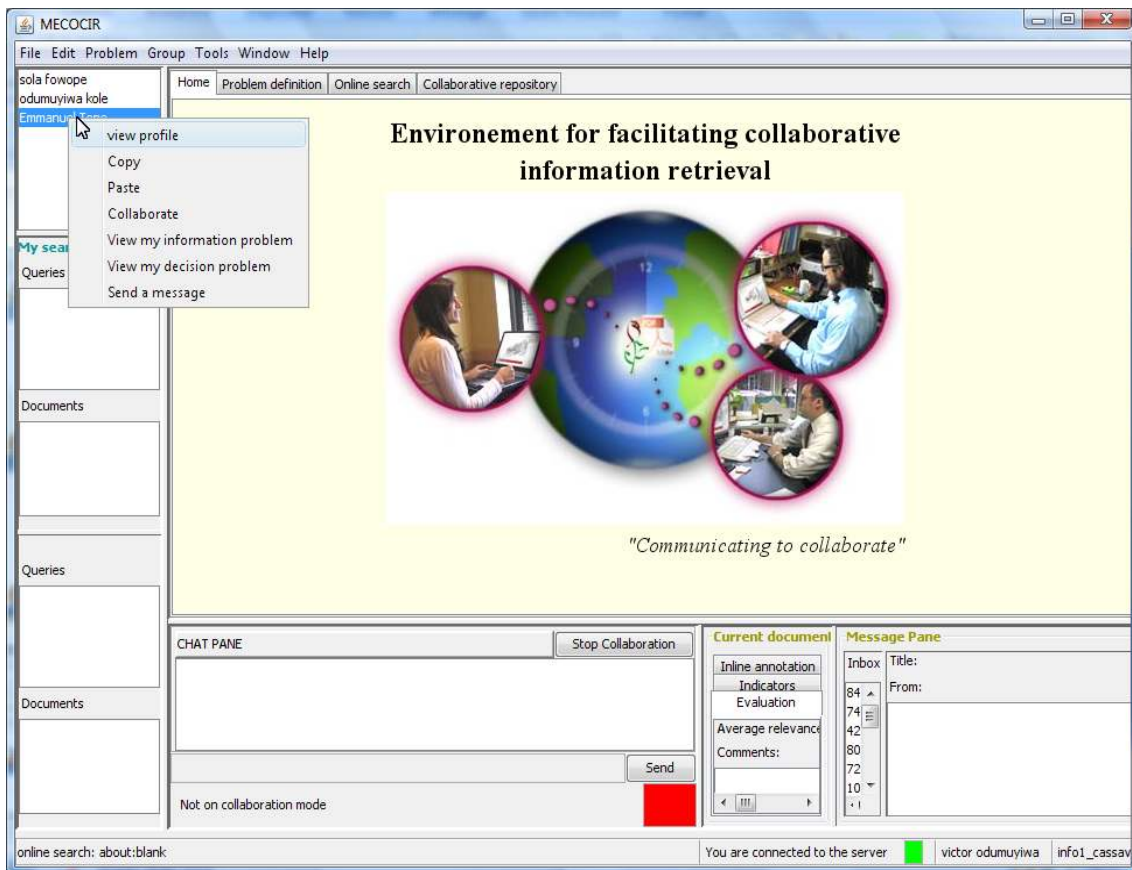


Figure 6.17 : Demande de collaboration

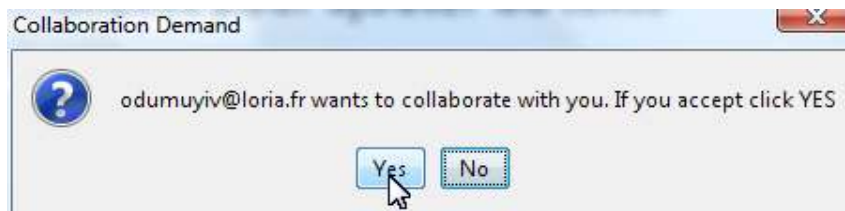


Figure 6.18 : Notification de la demande de collaboration

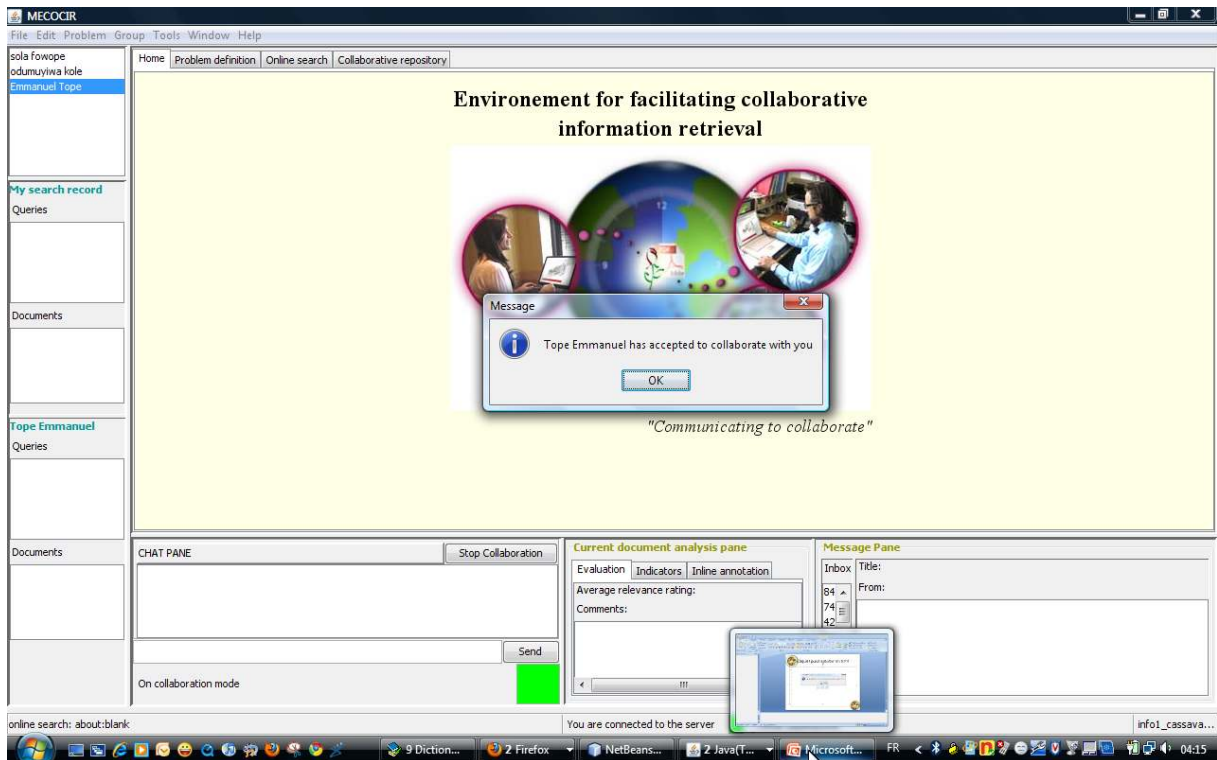


Figure 6.19 : Notification d'acceptation de collaborer

#### 6.4.5 Clarification du problème informationnel

Comme nous l'avons évoqué dans cette thèse, la réussite d'une activité de recherche collaborative d'information dépend de la compréhension du problème informationnel par les collaborateurs. Rappelons à nouveau qu'un problème partagé n'implique pas forcément une compréhension partagée du problème. Dans la RCI, la phase de clarification du problème informationnel permet aux collaborateurs de *s'entendre* sur une représentation du problème partagé. La figure 6.20 montre l'interface pour la clarification d'un problème informationnel. Cette interface est composée de deux zones :

- à gauche se trouve la définition initiale du problème tel qu'il est défini par l'initiateur ou par un membre d'un groupe s'il s'agit d'un groupe fermé ;
- à droite se trouve les annotations faites sur le problème. Elles correspondent aux connaissances exprimées pour clarifier le problème. Cette zone permet aux collaborateurs d'avoir des connaissances des activités des uns et des autres sur l'espace de travail d'une manière synchrone.



Cette interface fournit également une fonction de persistance de l'espace de travail car à chaque fois que l'utilisateur entame une session sur un problème informationnel, toutes les connaissances exprimées autour de ce problème antérieurement sont affichées permettant ainsi un suivi de l'évolution du problème.

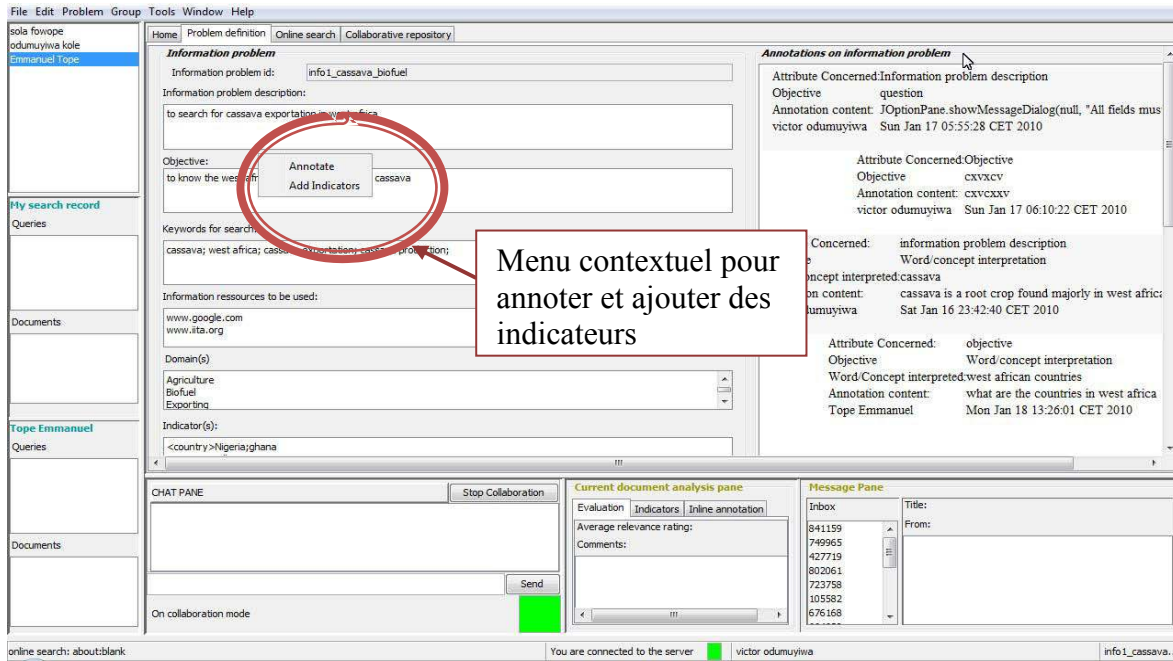


Figure 6.20 : Interface pour la clarification du problème informationnel

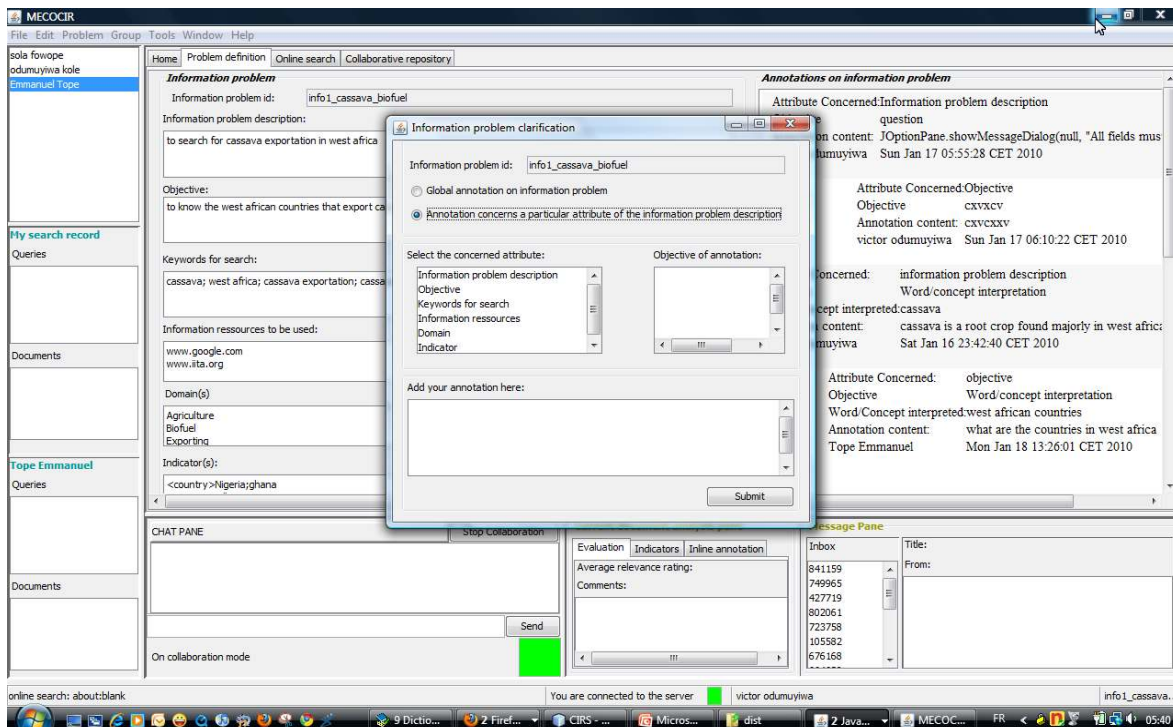


Figure 6.21 : L'interface d'annotation du problème informationnel

#### 6.4.6 La recherche d'information

Après la clarification du problème informationnel les collaborateurs peuvent commencer les activités de recherche des informations à propos du problème. L'onglet « Navigateur web<sup>140</sup> » donne l'accès à l'interface de navigation qui permet aux utilisateurs de se connecter à un système de recherche d'information de leur choix. Chaque SRI utilisé et les requêtes formulées sont captés automatiquement. Ces informations sont automatiquement partagées entre les partenaires lors de la collaboration. Chaque document consulté est également capté et partagé entre eux. MECOCIR permet aussi aux collaborateurs d'alterner entre l'interface de clarification et l'interface de recherche. En retrouvant des informations utiles, leurs connaissances du domaine augmentent. Ils peuvent ainsi revenir à l'interface de clarification afin de modifier le problème informationnel. La figure 6.22 montre l'interface de navigation avec le document actuel consulté par l'utilisateur.

Ce système analyse à travers le module « analyseur d'indicateur » chaque document affiché sur l'interface de navigation pour déterminer la fréquence d'occurrence des valeurs associées aux attributs définis comme indicateurs lors de la définition et de la clarification du problème informationnel. Ceci aide l'utilisateur dans le jugement de la pertinence d'un document avant de le lire. Si un utilisateur trouve qu'un document contient tous les indicateurs qu'il a défini, ceci peut l'inciter à lire le document. Dans la partie entourée de la figure 6.22, l'onglet « Indicateurs » de l'interface d'analyse du document actuel montre deux valeurs des indicateurs définis et leurs fréquences d'occurrence entre parenthèses.

---

<sup>140</sup> *Online search*



The screenshot displays the MECOCIR web application interface. The main content area shows a document titled "2. CASSAVA PRODUCTION" with a sub-section "2.1 CURRENT STATUS". Below the text is a bar chart showing cassava production in '000 MT. The chart has two bars: a taller one for Nigeria (21) and a shorter one for Ghana (1). A red circle highlights the "Current document analysis pane" which lists "Nigeria (21)" and "ghana (1)". A callout box with an arrow points to this pane with the text "L'analyse d'indicateurs". The interface also includes a left sidebar with search records and a bottom chat pane.

Figure 6.22 : L'interface de navigation et l'interface d'analyse du document actuel

#### 6.4.7 La conscience de groupe et le partage de connaissances

Dans la zone de gauche, se trouve le bloc des activités de groupe et le bloc des utilisateurs en ligne. Le bloc des activités de groupe (figures 6.23 et 6.24) permet à chaque utilisateur de voir son propre historique de recherche. En mode collaboration, un utilisateur peut voir en temps réel les activités de son partenaire. Il peut cliquer sur un document que son partenaire est en train de voir afin de le consulter. Il peut également annoter les documents et voir les annotations de son partenaire. Ce bloc permet d'avoir connaissance des activités de groupe. Par ce bloc, un utilisateur peut connaître le moteur de recherche préféré de son partenaire. Il permet également d'éviter la redondance des activités. Par exemple, en remarquant que son partenaire a lancé une requête précise sur un moteur de recherche, si l'utilisateur avait l'idée de formuler une requête identique, cette information sur l'activité de son partenaire qu'il vient d'apercevoir peut l'amener à changer sa requête ou lancer la même requête sur un autre moteur de recherche.

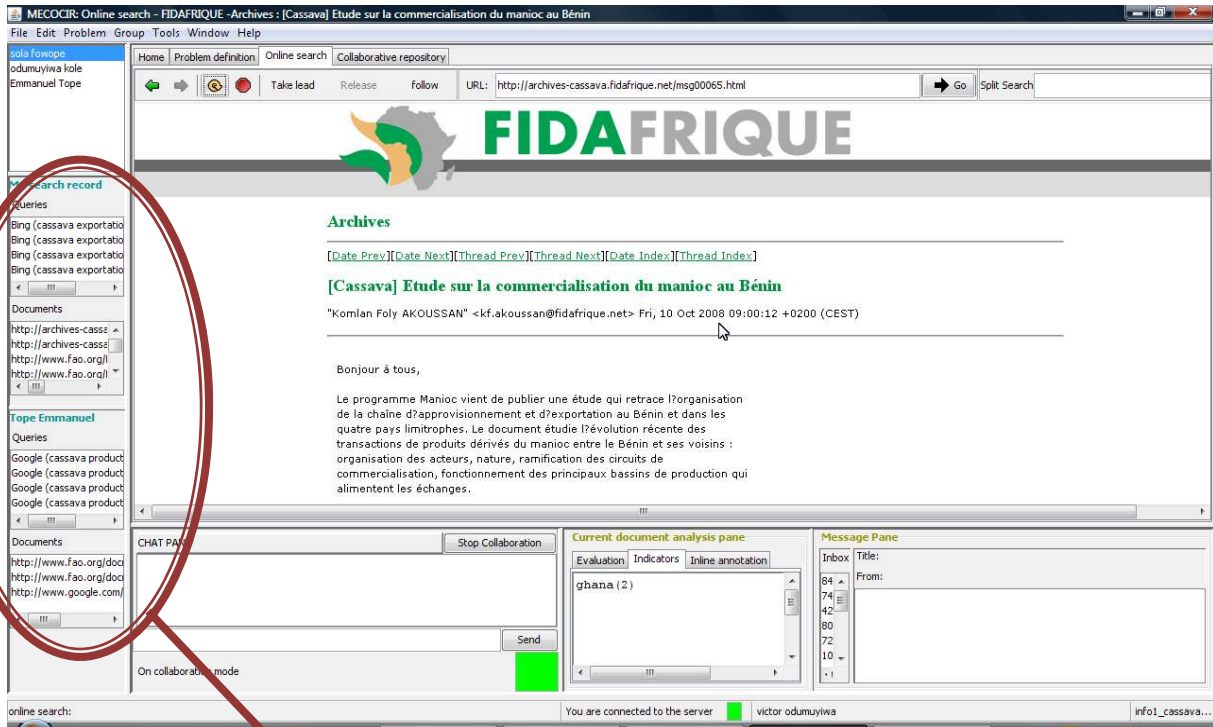


Figure 6.23 : Interface de navigation et l'interface pour la conscience des activités de groupe

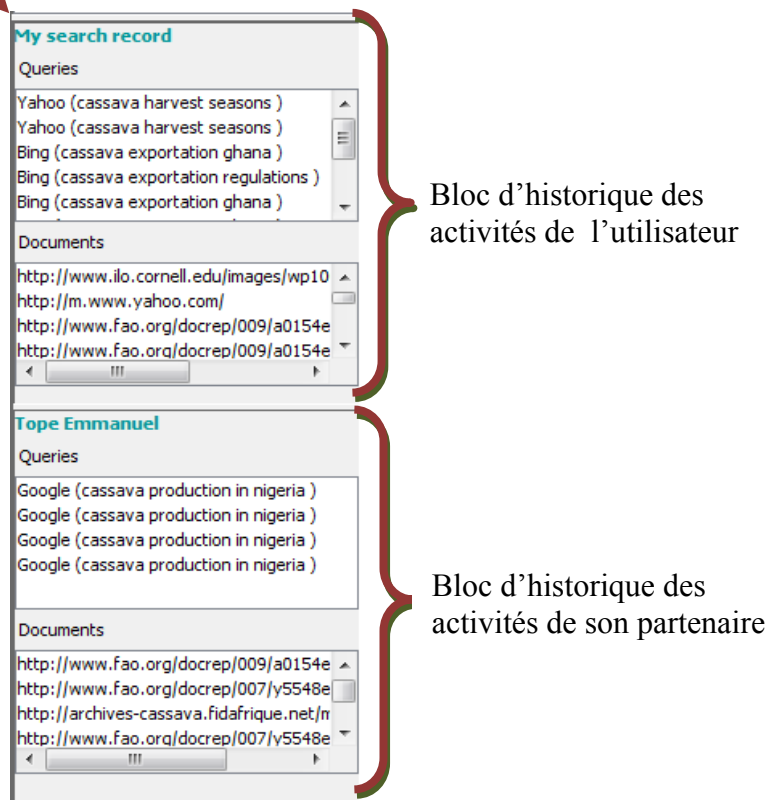


Figure 6.24 : L'interface pour la conscience des activités de groupe

Ce bloc permet également aux collaborateurs de partager leurs connaissances en temps réel cas leurs requêtes et les systèmes de recherche d'information utilisés font partie des connaissances à partager en RI. Les documents consultés et les annotations faites sur eux sont également des connaissances à partager. Un utilisateur (A) peut demander à son partenaire (B) de regarder le document qu'il est en train de voir et de donner son avis. Cette communication est faite à travers l'interface de chat. Le bloc des activités de groupe est conçu de manière à permettre l'affichage des activités des utilisateurs dans un ordre décroissant, c'est-à-dire que l'activité actuelle est toujours le premier élément de la liste. Ceci veut dire que B va cliquer sur le premier document dans la liste d'historique de A et voir le document s'afficher sur son écran. Il peut consulter le document ainsi que l'évaluation déjà faite là-dessus par son partenaire. Il peut donner son avis en annotant ce document. Pour faire cela, il fait un clic droit sur le document, puis un menu contextuel s'affiche. Il clique sur « annoter » et un écran semblable à celui de la figure 6.25 s'affiche, le permettant d'évaluer la pertinence du document par rapport au problème informationnel et aussi de donner son avis en texte libre.

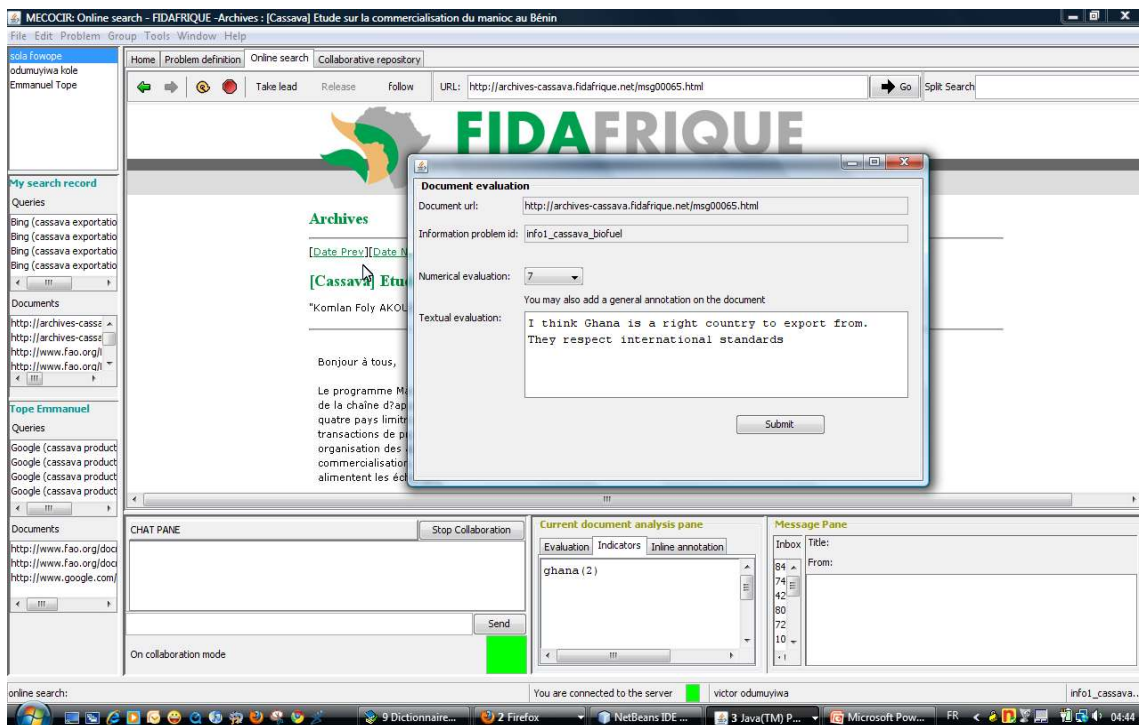


Figure 6.25 : L'évaluation et l'annotation d'un document

De la même manière, sur l'écran de A, l'annotation faite par B s'affiche dans l'interface d'analyse du document actuel. La figure 6.26 montre l'évaluation et l'annotation faite par B sur le document actuel de A.

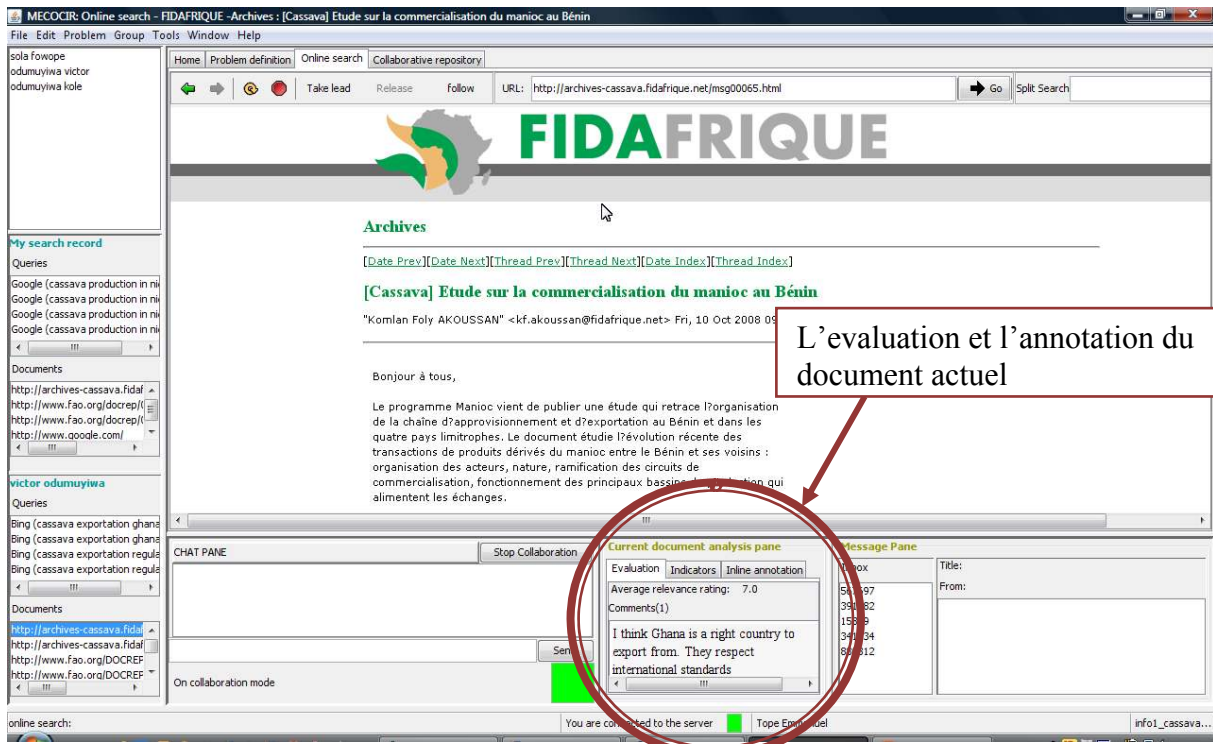


Figure 6.26 L'onglet d'évaluation dans l'interface d'analyse du document actuel

### 6.4.8 Synchronisation de l'interface de navigation

Dans le quatrième chapitre (cf. 4.6.2) sur le cadre conceptuel 3CM, nous avons évoqué deux modes de collaboration dans la RCI : mode observation et mode interaction. Jusqu'ici nous avons présenté comment MECOCIR permet aux utilisateurs de collaborer dans un mode interaction. Le système dispose également d'une fonctionnalité permettant de collaborer dans un mode observation en synchronisant les interfaces de navigation des collaborateurs. Pour synchroniser les interfaces, un utilisateur A prend la main et un utilisateur B le suit. Pour réaliser cela, A envoie un message instantané à B en le sollicitant d'entrer dans un mode synchronisé. Figure 6.27 montre la communication entre les deux utilisateurs pour partager les rôles.

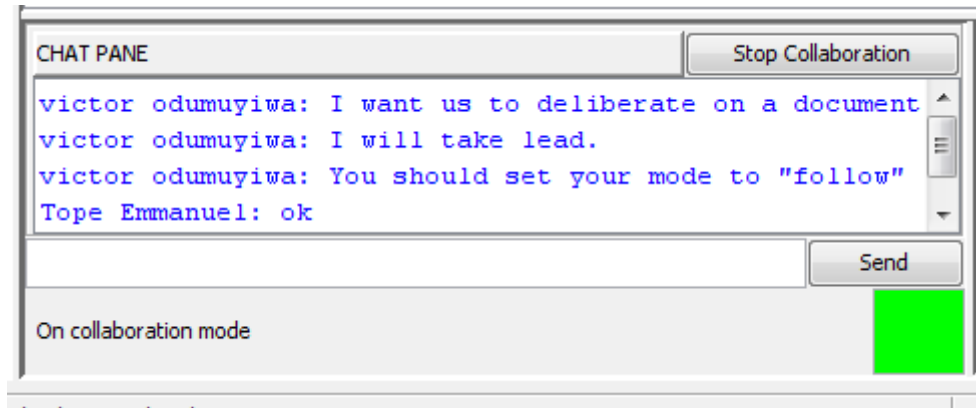


Figure 6.27 : L'interface de messagerie instantanée

Ensuite A clique sur le bouton « Prendre la main <sup>141</sup> » dans la barre d'outils (voir figure 6.28) et B clique sur le bouton « Suivre <sup>142</sup> » (voir figure 6.29). Le navigateur de B est alors synchronisé au navigateur de A. Chaque document consulté par A s'affiche automatiquement sur l'interface de B. La synchronisation peut être arrêtée en cliquant sur le bouton « Relâcher <sup>143</sup> ».

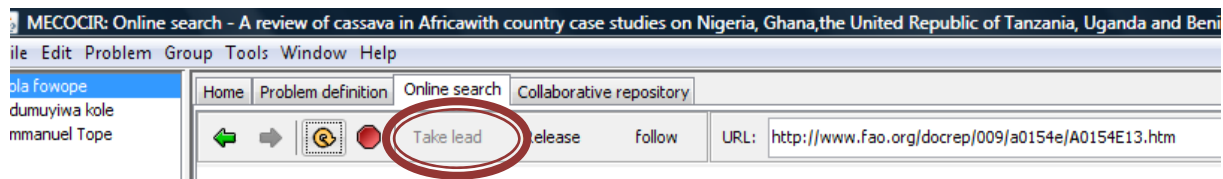


Figure 6.28 : Le bouton pour prendre la main

<sup>141</sup> *Take lead*

<sup>142</sup> *Follow*

<sup>143</sup> *Release*



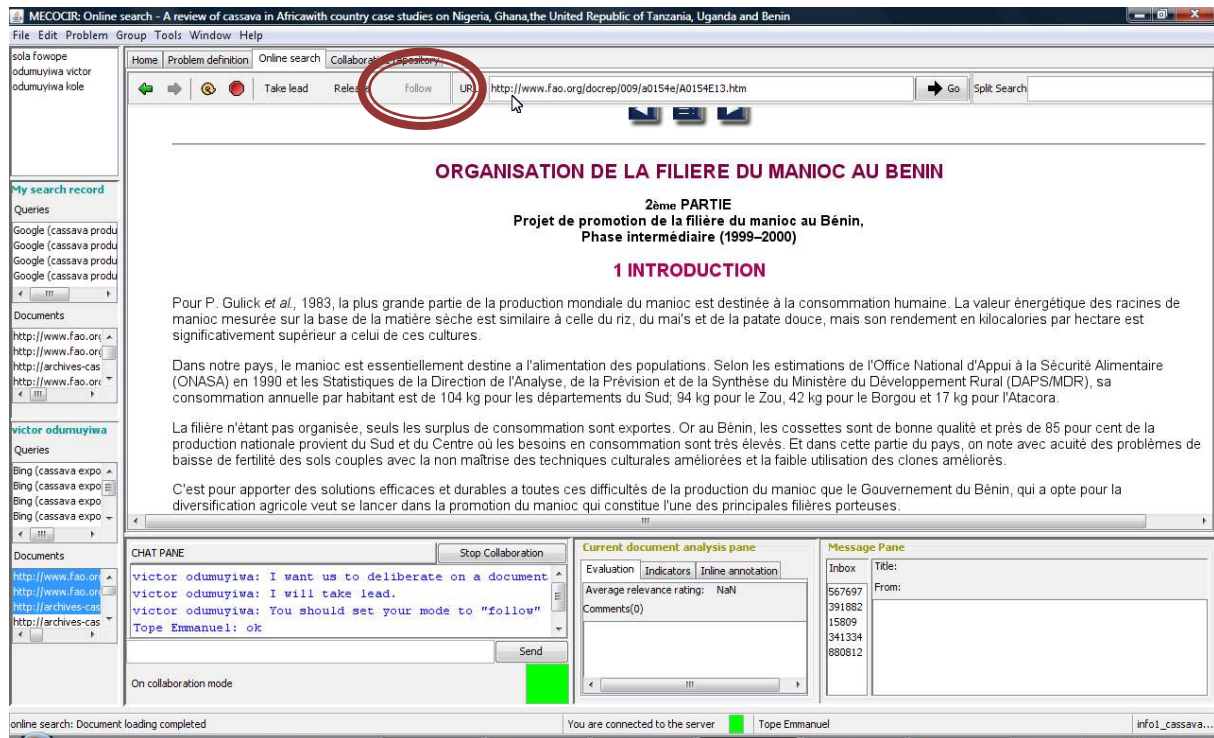


Figure 6.29 : L'interface de l'utilisateur B synchronisé avec celle de A

#### 6.4.9 La répartition des tâches

MECOCIR dispose de deux mécanismes de répartition des tâches : la messagerie instantanée et le *SplitSearch*. La messagerie instantanée permet aux collaborateurs de distribuer les tâches d'une manière non-automatisée. Le *SplitSearch*, en revanche, permet de faire la répartition des résultats de recherche parmi les collaborateurs d'une manière automatisée. Dans une session collaborative, lorsqu'un utilisateur lance une requête dans le champ « SplitSearch » comme le montre la figure 6.30, les résultats sont répartis également entre tous les participants. Les figures 6.31 et 6.32 montrent la répartition des résultats de recherche entre deux utilisateurs A et B. Notons que l'url de la page du résultat sur chaque interface commence par « file:/// ». Ceci correspond à l'explication donnée dans 6.2.3.3 que lorsque les collaborateurs se répartissent des tâches, certaines informations sur les tâches attribuées à chaque utilisateur sont stockées sur sa machine en local.

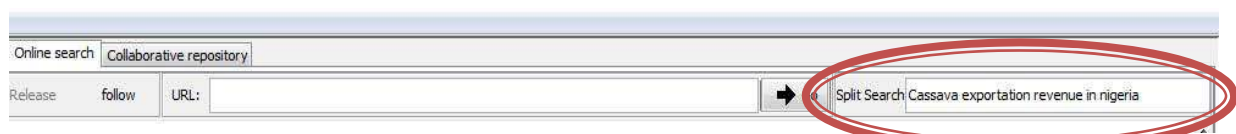


Figure 6.30 Le champ « SplitSearch »

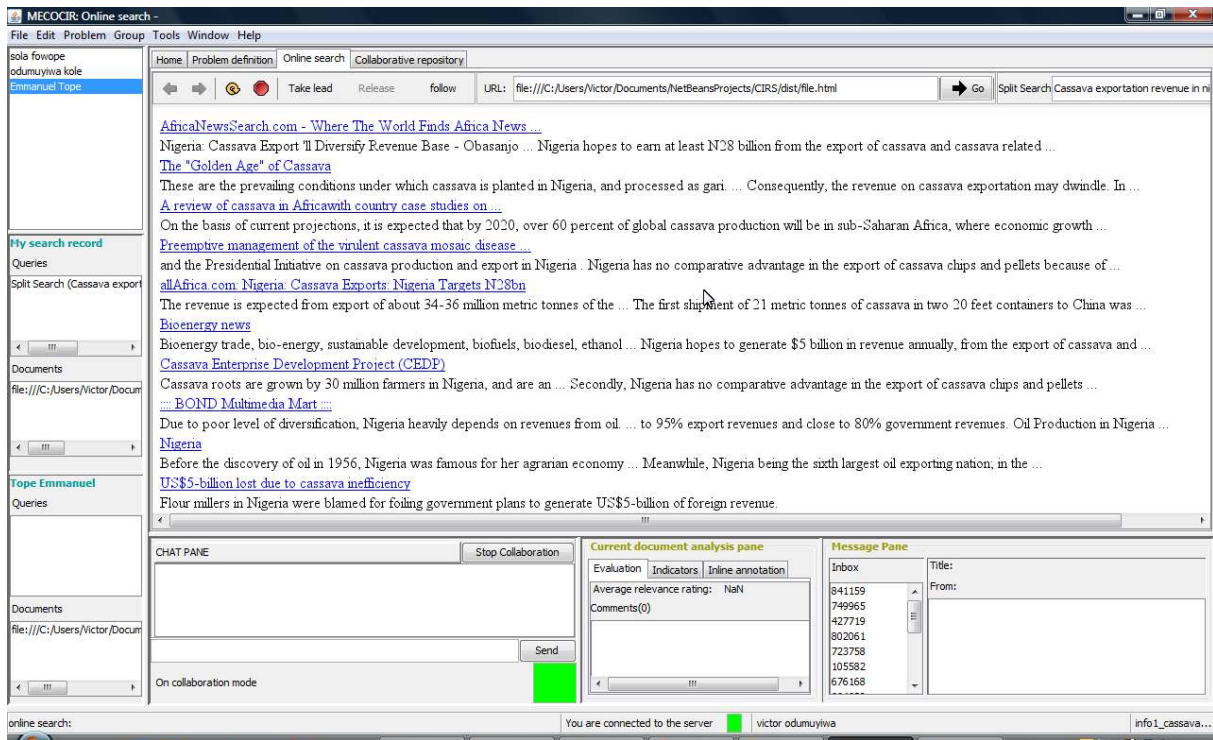


Figure 6.31 : Le résultat de la recherche affiché sur l'écran de l'utilisateur A

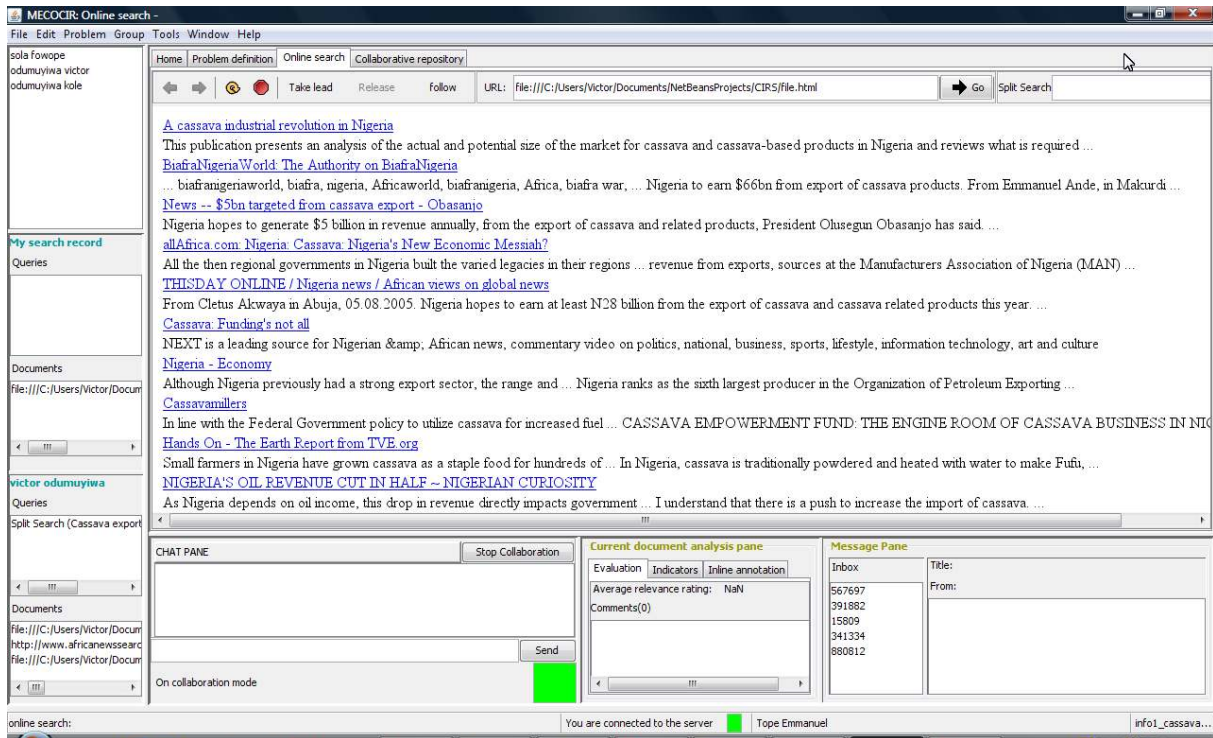


Figure 6.32 : Le résultat de la recherche affiché sur l'écran de l'utilisateur B

#### 6.4.9 Paradigme de collaboration

Dans la section 6.3, nous avons décrit deux cadres possibles de collaboration. MECOCIR dispose de fonctionnalités permettant de travailler dans ces deux cadres. Nous en avons ajouté un troisième « pair-à-pair<sup>144</sup> » et les trois représentent pour nous les paradigmes de collaboration :

- groupe  
ouvert : après l'identification d'un utilisateur sur MECOCIR, il se trouve automatiquement dans le groupe ouvert. Par défaut tous les utilisateurs appartiennent à ce groupe. Un utilisateur peut explorer le répertoire collaboratif pour découvrir les activités passées qui peuvent l'aider dans la résolution de son problème. Il peut également découvrir d'autres utilisateurs appartenant au même domaine que lui. Il peut envoyer une demande de collaboration à un autre utilisateur en ligne. Il n'est pas automatique que toute demande de collaboration reçoive une réponse positive ;
- groupe  
fermé : ceci correspond à un groupe prédéfini d'utilisateurs qui travaillent sur un problème partagé. Par exemple dans une entreprise, trois veilleurs peuvent former un groupe fermé. Un groupe fermé peut être également formé par le biais des interactions entre les utilisateurs dans un groupe ouvert, c'est-à-dire des utilisateurs qui découvrent qu'ils partagent un intérêt commun. Lorsqu'un utilisateur appartenant à un groupe fermé se connecte au système, il peut passer du mode groupe ouvert (par défaut) au mode groupe fermé en cliquant sur le menu « Groupe<sup>145</sup> », puis choisit « Groupe fermé<sup>146</sup> ». Le système s'adapte automatiquement et génère une liste des groupes dans lesquels l'utilisateur participe. Chaque groupe est représenté par son problème informationnel. L'utilisateur choisit l'un de ces problèmes pour commencer une session de recherche. De ce fait, l'interface principale du système est contextualisée et le bloc de la présence des membres de groupe n'affiche que les informations concernant les membres du groupe ;
- pair-à-pair :  
ceci correspond à un cadre de collaboration où les collaborateurs ne souhaitent pas que les connaissances exprimées soient stockées dans le répertoire collaboratif. Chaque

---

<sup>144</sup> *Peer-to-peer*

<sup>145</sup> *Group*

<sup>146</sup> *Closed group*



collaborateur dispose de sa propre base de connaissances et les échanges entre eux ne passent pas par le serveur d'application.

Le développement du système MECOCIR nous a permis d'implémenter nos modèles. Nous avons par la suite expérimenté le système auprès des utilisateurs. La section suivante présente les séries d'expérimentation effectuées.

## **6.5 Expérimentation**

Nous avons au cours de ce travail de recherche effectué deux séries d'expérimentation afin de vérifier notre hypothèse et valider nos modèles. La première série d'expérimentation a été réalisée en Janvier 2010 et la deuxième a été réalisée en Juin 2010. Dans chaque série d'expérimentation, nous avons fait intervenir six utilisateurs qui ont travaillé à deux, formant ainsi trois groupes. La procédure de conduite a été identique pour les deux séries. Nous sommes bien conscients qu'une population de 12 participants ne suffit pas pour tirer des conclusions définitives d'une expérimentation mais elle nous a permis d'observer certains comportements des utilisateurs.

Chaque expérimentation a commencé par une explication de l'objectif de l'exercice aux participants. Chaque participant a ensuite rempli un premier questionnaire (cf. annexe 1.1) qui nous a permis de recueillir des informations sur sa pratique informationnelle. A la fin du premier questionnaire, un problème informationnel a été présenté et à chaque participant a été demandé d'anticiper les démarches qu'il allait suivre pour résoudre le problème. Après cette étape, chaque participant a été amené à discuter avec son partenaire sur le problème informationnel afin qu'ils puissent élaborer ensemble comment ils allaient le résoudre.

Chaque participant s'est ensuite inscrit sur le système MECOCIR en suivant les étapes présentées dans la section 6.4. Cette étape a été suivie par l'explication du mode de fonctionnement du système et des fonctionnalités dont dispose le système. Pour chaque groupe d'utilisateurs, les membres du groupe ont été placés dans des salles différentes et ils ne disposaient pas d'autre moyen de communication que le système. Ces membres de groupe ont travaillé ensemble de manière synchrone sur le problème informationnel. Après cette activité de collaboration, chaque participant a rempli le deuxième questionnaire (cf. annexe 1.3) qui nous a permis d'avoir le retour de leurs expériences avec le système ainsi que leurs jugements

de l'activité de RCI qu'ils venaient d'effectuer et les apports de cette dernière à la résolution du problème informationnel auquel ils étaient confrontés.

Nous décrirons dans la section suivante la composition du groupe pour chaque série d'expérimentation et nos observations de celle-ci.

### **6.5.1 Première série d'expérimentation**

Dans cette série d'expérimentation, le premier groupe est composé de deux étudiants, le deuxième groupe est composé de deux médecins et le troisième groupe est composé de deux maîtres de conférences. Un exemplaire du premier questionnaire rempli par les participants est présenté en annexe 1.1.

Les réponses au premier questionnaire montrent que :

- i. 3 participants sur 6 utilisent la bibliothèque dans leurs démarches pour résoudre un problème informationnel
- ii. Tous les participants utilisent Internet dans leurs démarches de recherche d'information. Comme le montre la figure 6.34, ils utilisent tous Google comme moteur de recherche principal. D'autres moteurs de recherche utilisés de manière régulière et moins régulière sont indiqués dans le graphique
- iii. 2 participants sur 6 partagent leurs problèmes informationnels avec leur entourage avant de commencer la recherche d'information
- iv. Lors de la recherche d'information personne ne partage son problème informationnel
- v. 4 participants sur 6 partagent leur problème informationnel avec leur entourage après un effort individuel sans résultat
- vi. 5 participants sur 6 indiquent qu'ils n'ont pas peur de partager leur problème informationnel avec d'autres individus
- vii. 5 participants sur 6 indiquent qu'ils partagent leurs résultats de recherche avec d'autres individus ayant des problèmes informationnels similaires
- viii. Tous les participants désirent que des moyens de partage avec d'autres utilisateurs en temps réels soient intégrés dans les SRI qu'ils utilisent

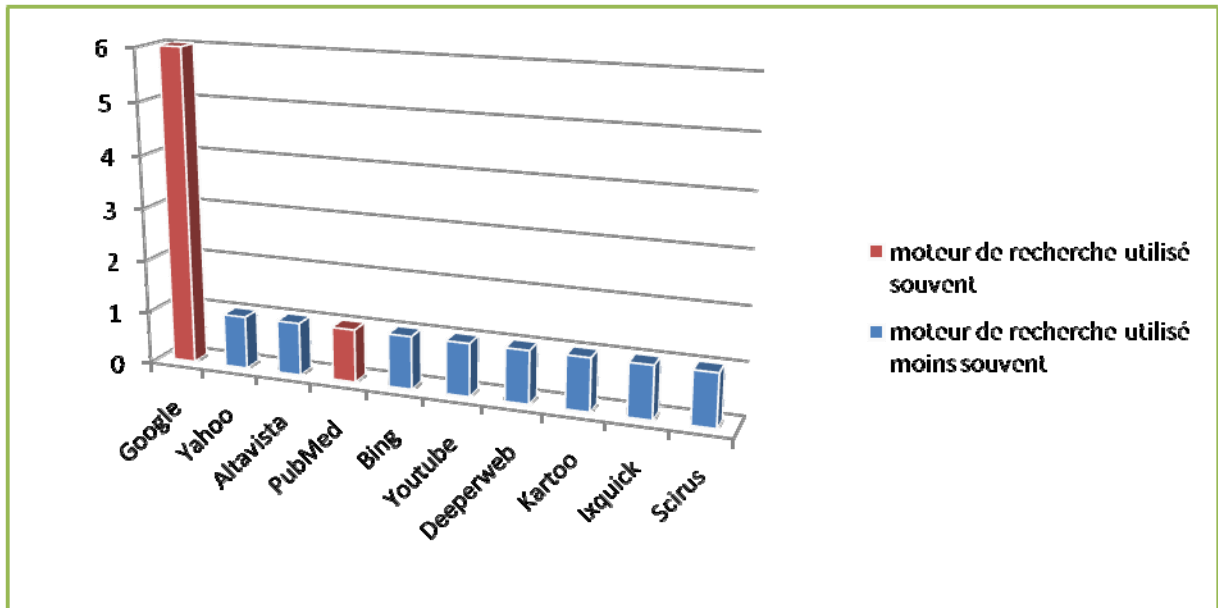


Figure 6.34 : Les moteurs de recherche utilisés par les participants

Les réponses au premier questionnaire justifient certains de nos propos concernant le comportement informationnel de l'utilisateur. Nous avons remarqué que tous les six participants utilisent Internet comme source d'information principale et qu'ils se servent de plusieurs SRI, Google étant le plus utilisé. Un SRCI doit ainsi pouvoir supporter la recherche d'information sur Internet et il doit permettre aux utilisateurs de choisir les sources d'informations qui leur conviennent. Nous avons remarqué également que les participants ont une tendance à partager leurs problèmes informationnels mais le questionnaire ne leur permet pas de donner plus de précision sur la raison pour laquelle ils le partagent et sur les types de connaissances qu'ils désirent à acquérir sur leurs problèmes informationnels. Nous avons donc modifié le premier questionnaire pour la deuxième série d'expérimentation.

Ayant rempli le questionnaire, il a été donné aux participants un problème informationnel à résoudre. Ce problème découle d'un problème décisionnel défini de la manière suivante :

- L'expression du problème : j'ai une odeur corporelle qui peut être perçue à deux mètres. Les médecins n'ont pas pu m'apporter des solutions face à cette situation. Je cherche à comprendre les causes de mon problème car je suis frustré et je vois ma vie s'effondrer.
- Objet : odeur corporelle.
- Signal :

- l'odeur est perçue depuis une distance de plus de deux mètres ;
- les médecins n'ont pas de solutions pour moi.
- Hypothèse :
  - si je connais les causes du problème, je pourrais me soigner ;
  - sinon je risque continuer à être frustré et embêter mon entourage.

Le problème informationnel issu du problème décisionnel est le suivant :

« Que peut-il se passer dans l'organisme de quelqu'un expliquant qu'il ait une odeur épouvantable ? »<sup>147</sup>

Chaque participant a d'abord anticipé les démarches qu'il allait suivre pour résoudre le problème individuellement et puis les participants de chaque groupe ont discuté sur la façon dont ils allaient diriger la recherche d'information.

Nous utilisons les activités de groupe des deux étudiants pour expliquer les interactions entre les participants. Dans la suite, U1 et U2 représentent les deux étudiants. Leurs anticipations personnelles pour la résolution du problème informationnel sont présentées ci-dessous:

U1 : « D'abord je ferai des recherches générales sur Internet puis des recherches plus précises en bibliothèque et si nécessaire je demande l'aide d'une personne qui s'y connaît, comme un professeur »

U2: « Retrouver les mots-clefs qui correspondent à cette question par exemple la physiologie de l'odeur et du système nerveux. Chercher sur Internet ou dans des livres en fonction de la disponibilité »

Après cette étape personnelle, U1 et U2 ont partagé ce qu'ils ont noté puis ils ont harmonisé leurs points de vue.

---

<sup>147</sup> Cette question a été tirée d'un site d'aide en ligne qui permet aux utilisateurs de poser des questions. Notons que le <http://www.ehelp.com/questions/10434735/what-can-be-going-on-internally-to-make-someone-smell-very-awful-nk-stds>

La deuxième étape est de passer à l'utilisation du système MECOCIR. U1 a défini le problème informationnel selon les attributs proposés par le système (voir figure 6.35). Si nous comparons l'anticipation personnelle d'U1 sur la résolution du problème avant d'utiliser le système avec la définition du problème qu'il a fait en utilisant le système, nous pouvons remarquer une augmentation dans son état de connaissance sur le problème. Nous avons remarqué que les attributs proposés par le système ainsi que la discussion entre les deux participants ont permis à U1 de donner plus de clarification sur le problème informationnel.

Après la définition du problème informationnel, U1 a invité U2 à collaborer avec lui sur le problème comme expliqué dans la section 6.4.4. Ayant accepté de collaborer avec U1, U2 est automatiquement amené à voir la définition du problème faite par U1 et il a ensuite ajouté des annotations pour clarifier le problème. Les annotations d'U2 sont présentées dans la figure 6.36. U1 et U2 ont ensuite recherché ensemble des informations sur le problème informationnel.

The screenshot shows a web application interface with a navigation bar at the top containing 'Home', 'Problem definition', 'Online search', and 'Collaborative repository'. The main content area is titled 'Information problem' and contains several input fields:

- Information problem id:** Odeur épouvantable
- Information problem description:** Que peut se passer à l'intérieur pour rendre l'odeur de quelqu'un épouvantable?
- Objective:** Je connais quelqu'un dans mon entourage qui a ce problème.
- Keywords for search:** Forte odeur, problème corporel
- Information resources to be used:** Professeur, Bibliothèque, Internet
- Domain(s):** Physiologie
- Indicator(s):** (empty field)

Figure 6.35 : la définition du problème informationnel faite par U1

**Annotations on information problem**

Attribute Concerned: Keywords for search

Objective: Le probleme contient une autre composante. La réaction de celui qui sent l'odeur.

Annotation content: Les systèmes d'intégration neuronale

Annotation added by: U2 Fri Oct 01 18:42:03 CEST 2010

Attribute Concerned: Domain

Objective: Manque de précision

Annotation content: Sémiologie

Annotation added by: i U2 Mon Jan 18 16:01:01 CET 2010

Attribute Concerned: Keywords for search

Objective: Rectification

Annotation content: Sémiologie associée aux odeurs ou sémiologie et odeur

Annotation added by: U2 Mon Jan 18 15:55:23 CET 2010

Attribute Concerned: Information ressources

Objective: Pas assez précis

Annotation content: Il est necessaire de réduire le libellé avec des mots clés

Annotation added by: U2 Mon Jan 18 15:16:32 CET 2010

Figure 6.36 : les annotations ajoutées<sup>148</sup> par U2 pour clarifier le problème informationnel

Cette série d'expérimentation a été marquée par des problèmes techniques au niveau du serveur d'application à cause de certaines erreurs de programmation. Seul le premier groupe a pu bien terminer l'expérimentation. Les deux d'autres groupes n'ont pas pu dépasser l'étape de clarification du problème informationnel. Les deux participants du premier groupe ont rempli le deuxième questionnaire (cf. annexe 1.2). Plusieurs problèmes au niveau technique et au niveau ergonomique ont été soulignés par les participants. Nous avons, à la suite de cette première expérimentation, retravaillé le système et corrigé les problèmes qui ont été soulignés. Nous avons également modifié les deux questionnaires afin d'avoir plus de précision dans les informations fournies par les utilisateurs. Ces deux versions modifiées des questionnaires et la nouvelle version de MECOCIR ont été utilisées lors de la deuxième série d'expérimentation.

<sup>148</sup> Les attributs dans cette interface sont en anglais (car MECOCIR à l'heure actuelle ne dispose que des interfaces en anglais). Voici les traductions de ces attributs : *attribute concerned* → attribut concerné ; *objective* → objectif ; *annotation content* → contenu d'annotation.

### 6.5.2 Deuxième série d'expérimentation

Six personnes, dont quatre étudiants et deux maîtres de conférences en SIC, ont participé à la deuxième série d'expérimentation. Les quatre étudiants ont formé deux groupes et le troisième groupe est composé des deux maîtres de conférences. Les deux maîtres de conférences sont ceux qui ont participé à la première série d'expérimentation. Tous les participants ont rempli la version modifiée du premier questionnaire (voir annexe 1.3).

Les réponses au premier questionnaire montrent que :

- i. 3 participants sur 6 utilisent la bibliothèque dans leurs démarches pour résoudre un problème informationnel :
  - 2 participants sur 6 utilisent l'outil informatique dans la bibliothèque pour rechercher des informations ;
  - 1 participant dialogue avec les bibliothécaires sur leur problème informationnel ;
  - tous les participants qui utilisent la bibliothèque recherchent des livres directement dans les rayons.
- ii. Tous les participants utilisent Internet dans leurs démarches de recherche d'information. Ils utilisent tous Google comme moteur de recherche principal. D'autres moteurs de recherche utilisés de manière régulière et moins régulière sont indiqués dans la figure 6.37.
- iii. Avant de commencer une recherche d'information :
  - 4 participants sur 6 indiquent qu'ils demandent de l'aide de leur entourage pour comprendre leur problème informationnel ;
  - personne n'indique qu'il demande de l'aide pour savoir quelle source d'information à utiliser ;
  - personne ne demande de l'aide sur le fonctionnement du SRI à utiliser ;
  - 1 participant demande de l'aide sur comment procéder dans la recherche d'information.
- iv. Lors d'une recherche d'information:
  - 3 participants sur 6 demandent de l'aide pour comprendre leur problème informationnel ;

- 1 participant demande de l'aide pour savoir quelle source d'information à utiliser ;
  - 2 participants sur 6 demandent de l'aide sur le fonctionnement du SRI à utiliser ;
  - 1 participant demande de l'aide sur comment procéder dans la recherche d'information.
- v. Après un effort individuel de recherche d'information sans résultat :
- 3 participants sur 6 demandent de l'aide pour comprendre leur problème informationnel ;
  - 4 participants sur 6 demandent de l'aide pour savoir quelle source d'information à utiliser ;
  - 3 participants sur 6 demandent de l'aide sur le fonctionnement du SRI à utiliser ;
  - 3 participants sur 6 demandent de l'aide sur comment procéder dans la recherche d'information.
- vi. 4 participants sur 6 indiquent qu'ils se méfient du fait de demander de l'aide à une autre personne par rapport à leur problème informationnel. Les raisons qu'ils ont données sont les suivantes :
- « j'ai peur que les gens m'embrouillent » ;
  - « mauvaise foi de certains » ;
  - « problème de compréhension et d'interprétation et de perte de temps »
  - « ça dépend de la personne et du sujet :
    - si la personne est un concurrent, je me méfie de lui demander de l'aide ;
    - si la personne est indiscreète, je me méfie de lui demande de l'aide ;
    - si la personne est un collègue ou un collaborateur, je lui demande de l'aide ;
    - si le sujet est confidentiel, je me méfie de demander de l'aide ;
    - si il n y a aucun enjeu, je ne me méfie pas de demander de l'aide. ».
- vii. 5 participants sur 6 indiquent qu'ils partagent leurs résultats de recherche avec d'autres individus ayant des problèmes informationnels similaires.



- viii. Tous les participants désirent que des moyens de partage avec d'autres utilisateurs en temps réel soient intégrés dans les SRI qu'ils utilisent.

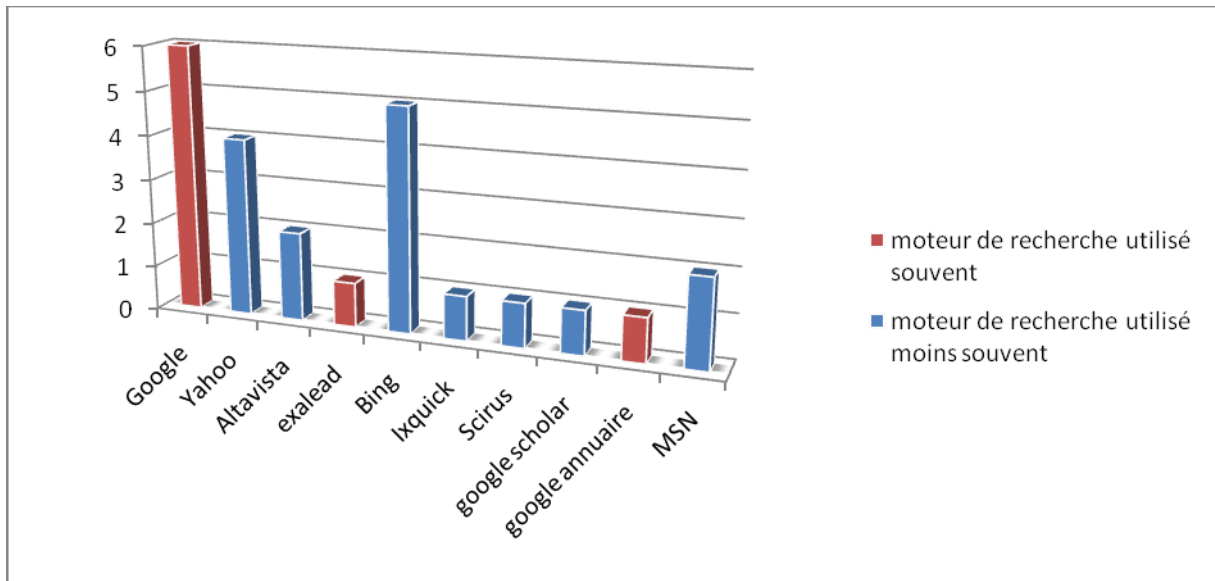


Figure 6.37 : Les moteurs de recherche utilisés par les participants de la deuxième série d'expérimentation

Nous avons remarqué, à partir des réponses des participants au premier questionnaire, que le facteur de confiance est important pour entamer une collaboration avec une autre personne. Nous avons observé également que les participants ont une tendance à collaborer et qu'ils désirent même que les SRI qu'ils utilisent leur permettent de collaborer en temps réel lors de la recherche d'information. Nous avons remarqué, en plus, que les utilisateurs ont besoin d'être aidés dans la clarification de leurs problèmes informationnels afin qu'ils puissent mieux les comprendre. Cette étude montre également que la source principale d'information pour les utilisateurs aujourd'hui est Internet.

Comme dans la première série d'expérimentations, il a été donné aux participants un problème informationnel à résoudre. Ils ont réfléchi dans un premier temps individuellement comment le résoudre puis ils ont recherché ensemble en utilisant le système MECOCIR pour répondre à ce besoin informationnel. Après cette étape, les participants ont rempli la version modifiée du deuxième questionnaire (voir annexe 1.4).

Le deuxième questionnaire a été conçu pour permettre de recueillir des informations sur les apports des fonctionnalités intégrées dans MECOCIR aux activités de recherche collaborative

d'information parmi les participants et pour permettre également de savoir les types de connaissances que les utilisateurs partagent entre eux lors de la collaboration. Les statistiques sur les réponses au deuxième questionnaire sont suivantes :

- i. 5 participants sur 6 indiquent que la phase de définition et de clarification de problème informationnel facilite la recherche d'information ainsi que la collaboration
- ii. 4 participants sur 6 indiquent que les fonctionnalités incluses dans MECOCIR leur permettent de clarifier au mieux leur problème informationnel
- iii. Tous les participants ont utilisé la fonctionnalité de chat et ils indiquent qu'il a amélioré le processus de collaboration
- iv. 4 participants sur 6 indiquent que la fonctionnalité de conscience de groupe, qui leur permet d'avoir connaissances en temps réel sur les activités de leur partenaire, a amélioré le processus de collaboration
- v. Tous les participants sont d'accord qu'il est nécessaire de pouvoir partager des tâches lors d'une collaboration
- vi. 1 participant a souvent utilisé la fonctionnalité de répartition des tâches intégrée dans MECOCIR
- vii. 4 participants sur 6 indiquent que cette fonctionnalité de répartition des tâches (*Split search*) est pertinente à la RCI
- viii. 5 participants sur 6 indiquent que MECOCIR leur permet **de partager** leurs connaissances du sujet de recherche avec leurs partenaires lors de la RCI
- ix. 4 participants sur 6 indiquent que MECOCIR leur permet **de partager** leurs connaissances des sources d'information avec leurs partenaires lors de la RCI
- x. 3 participants sur 6 indiquent que MECOCIR leur permet **de partager** leurs connaissances sur le fonctionnement des SRI utilisés lors de la RCI
- xi. 5 participants sur 6 indiquent que MECOCIR leur permet **d'acquérir** des connaissances sur le sujet de recherche de leurs partenaires
- xii. 3 participants sur 6 indiquent que MECOCIR leur permet **d'acquérir** des connaissances des sources d'informations à utiliser de leurs partenaires lors de la RCI
- xiii. 4 participants sur 6 indiquent que le fait de collaborer lors de recherche d'information rend la recherche plus rapide.

### 6.5.3 Bilan des deux séries d'expérimentation

Les deux séries d'expérimentation que nous avons effectuées nous ont permis de valider le modèle de pyramide de collaboration que nous avons proposé dans la section 5.2 ainsi que le modèle COCIR (cf. 5.3). Nous avons remarqué que les six phases de la pyramide de collaboration (phase de confiance de départ, phase de compréhension partagée du problème, phase de communication, phase de partage et complémentarité de connaissances, phase de conscience de groupe et phase de répartition des tâches) sont très essentielles pour la réussite de la RCI. Notons également que ces phases ne sont pas séquentielles mais peuvent s'effectuer en parallèle. Par exemple, quand une répartition des tâches est effectuée, la conscience de groupe est également assurée pour que chaque membre de groupe puisse savoir qui fait quoi.

Nous avons aussi remarqué que certains facteurs comme la relation interpersonnelle, les cognitions sociales, une mentalité de compétition etc. (cf. 1.5.3 et 1.5.3.1) influent énormément sur la réussite de la RCI. Par exemple, dans le cas de l'un des groupes de la deuxième série d'expérimentations, nous avons remarqué un climat de compétition. Les deux membres de ce groupe sont très compétents dans la recherche d'information mais ce climat de compétition les a empêchés de bien co-construire dans la collaboration. Nous avons remarqué également que les stéréotypes que l'on détient sur un individu influent sur la collaboration avec lui. Dans le cas cité ci-dessus, nous pouvons déduire que le climat de compétition est lié aux stéréotypes et à la relation interpersonnelle entre les deux participants. Par exemple, l'un des deux membres du groupe désire que le système lui permette de visualiser l'écart entre ses recherches en termes de sources consultées et celles de son partenaire. En discutant avec cet utilisateur après l'expérimentation pour savoir pourquoi ce désir, il nous a expliqué qu'il voulait savoir à quel niveau ses contributions sont complémentaires à celles de son partenaire et aussi pour éviter la redondance. Nous avons également discuté avec son partenaire à cause de sa réponse à l'une des questions du deuxième questionnaire : « Est-ce que le fait de collaborer avec quelqu'un pour chercher des informations rend la recherche plus rapide ? ». Sa réponse était la suivante : « Je pense que ça dépend des gens avec qui on travaille. N'ayant pas atteint la phase de validation des résultats, je peux difficilement juger ». Cette personne nous a expliqué que les facteurs tels que le biais à la collaboration, la fierté, la culture, le fait de ne pas vouloir montrer ses faiblesses, course à la meilleure performance (qui est celui qui trouve en premier) et le fait de faire valoir ses compétences et connaissances peuvent empêcher la réussite de la collaboration. Nous avons déduit évidemment que cette personne

parle de son partenaire notant également que ces deux participants se connaissent depuis plus de sept ans.

Les deux séries d'expérimentation nous ont permis de beaucoup apprendre sur les comportements informationnels des utilisateurs et sur les fonctionnalités qui facilitent la RCI. Nous constatons également qu'il sera nécessaire d'effectuer d'autres séries d'expérimentation à une échelle plus grande avec plus de participants pour pouvoir généraliser nos constats. Nous continuons à améliorer le système MECOCIR pour le rendre plus robuste en termes de performance et de stabilité.

### 6.6 Comparaison du MECOCIR avec certains SRCI existants

Nous avons comparé les fonctionnalités fournies par MECOCIR avec d'autres SRCI existants qui permettent également une collaboration explicite autour d'un problème informationnel. Cette comparaison est résumée dans le tableau 6.1.

Fonctionnalités/SRCI	ARIADNE	CIRE	METIORE	SEARCHTOGETHER	MECOCIR
Clarification du problème informationnel	Non	Non	Non	Non	Oui
Partage du processus de recherche	Oui	Oui	Oui	Oui	Oui
Partage de résultat de recherche	Oui	Oui	Oui	Oui	Oui
Partage d'écran	Non	Non	Oui	Oui	Oui
communication entre les acteurs	Non	Non	Oui	Oui	Oui
Répartition des tâches	Non	Non	Non	Oui	Oui
Filtrage collaboratif	Oui	Oui	Non	Non	Oui
Recherche d'information sur l'internet	Non	Oui	Non	Oui	Oui
Collaboration synchrone	Oui	Non	Oui	Oui	Oui

Tableau 6.1 : Tableau comparatif des SRCI

## **6.7 Conclusion du sixième chapitre**

Nous avons présenté dans ce chapitre l'implémentation de nos modèles dans le système MECOCIR. La conception et le développement du système ont été détaillés ainsi que les modules fonctionnels comportant le système. Nous avons également présenté le mode de fonctionnement du système accompagné par une description de deux séries d'expérimentation effectuées.



# Conclusion générale

---

Notre travail de thèse a porté sur la gestion de la recherche collaborative d'information dans le cadre du processus d'intelligence économique. Notons que l'identification du problème décisionnel, sa traduction en problème informationnel et la collecte d'information font partie des étapes déterminantes pour la réussite du processus d'intelligence économique. Ces étapes nécessitent la collaboration entre les acteurs impliqués dans le processus. Nous avons ainsi cherché à proposer des modèles et une architecture d'un système qui permettent de faciliter et de gérer les activités de collaboration autour d'un problème informationnel.

Nous avons, tout d'abord, étudié les évolutions qui ont vu le jour dans le domaine de recherche d'information tout en s'intéressant aux deux axes de ce domaine à savoir : le processus de RI et le système de RI. Nous avons pu constater que les utilisateurs, de manière naturelle, ont une tendance à demander de l'aide à leurs entourages lors de leurs démarches de RI et que ce comportement n'est pas pris en compte dans les SRI classiques. Ceci nous a amené à étudier le comportement informationnel de l'utilisateur et les SRI existants ainsi que les SRCI existants. Bien que plusieurs auteurs aient tenté d'éclairer le concept de la RCI et de développer des systèmes de RCI, nous remarquons qu'ils n'abordent ce sujet que du point de vue d'un groupe d'utilisateurs collaborant sur un problème informationnel partagé. Ces auteurs n'ont pas étudié le problème de compréhension du problème par le groupe. *En réalité, un problème partagé n'implique pas forcément une compréhension partagée du problème.* Le niveau de compréhension d'un problème informationnel détermine grandement la qualité et la rapidité de la résolution quelle que soit la nature de l'intervention qui peut être individuelle ou collective.

Suite à ces observations, nous sommes partis d'une problématique qui se traduit en les questions suivantes :

- Comment passer d'un problème informationnel partagé à une compréhension partagée du problème ? Sachant que le point de départ dans une collaboration devrait toujours passer par la question : pourquoi collabore-t-on et pour atteindre quel objectif ?
- Comment gérer les différents types de connaissances mises en œuvre dans la recherche collaborative d'information ?

- Comment assurer la conscience d'appartenance à un groupe (*awareness*) et la répartition des tâches (*division of labour*) dans une activité de RCI ?
- Comment garantir ou assurer la sécurité et la protection du patrimoine dans un système de RCI ?

Nous avons ainsi formulé une hypothèse *qu'une bonne gestion des connaissances collectives impliquées dans la recherche d'information devrait faciliter dans un premier temps la recherche collaborative d'information et dans un deuxième temps, permettre d'améliorer le résultat global du processus d'intelligence économique dans le cadre de la résolution d'un problème décisionnel.*

Afin de répondre à ces questions, nous avons d'abord développé un cadre conceptuel permettant d'identifier et d'analyser les aspects que nous avons identifié comme important pour la gestion et la facilitation de la RCI à travers un environnement numérique. Dans notre cadre, la gestion de la RCI fait intervenir quatre aspects principaux à savoir :

- la communication ;
- le mode de collaboration ;
- la coordination des interactions entre les utilisateurs ;
- la gestion des connaissances impliquées dans la collaboration.

En effet, nous considérons le cas où l'on communique pour collaborer, impliquant une coordination des interactions afin de pouvoir gérer les connaissances impliquées dans la collaboration. Nous appelons ce cadre 3CM (*Communication, Collaboration, Coordination, Management*) selon sa traduction de l'anglais. Chaque élément de ce cadre est composé de sous-éléments qui le caractérisent. Par exemple, pour la communication, on considère le « pourquoi », le « quoi », le « comment », le « qui » et le « quand » de la communication. Pour l'élément « collaboration » du cadre, nous adoptons deux modes de collaboration : le mode observation et le mode interaction. En ce qui concerne l'élément « coordination » du cadre, nous adoptons trois types de dépendance qui résument toutes les dépendances pouvant se manifester entre les activités des collaborateurs. Ces trois dépendances sont : la dépendance de flux (*flow dependency*), la dépendance de partage (*sharing dependency*) et la dépendance



d'ajustement (*fit dependency*). Le dernier élément du cadre « la gestion de connaissances » nécessite la modélisation de l'utilisateur, l'acquisition et l'exploitation des connaissances.

L'**originalité** de notre travail réside, tout d'abord, dans le modèle de pyramide de collaboration que nous avons élaboré. Ce modèle montre les différentes phases et composants nécessaires pour la réussite d'un processus de collaboration surtout dans le cadre de la recherche d'information. Pour que la recherche collaborative d'information soit réussie, il faut un niveau minimum de confiance de la part des collaborateurs. Celle-ci est le socle de la collaboration. La base d'une collaboration repose sur la confiance réciproque des partenaires. La deuxième phase dans la collaboration est l'élaboration et la clarification du problème informationnel autour duquel les collaborateurs travaillent. Il ne suffit pas seulement de partager un problème, mais il est important de faire passer les collaborateurs d'un problème partagé à une compréhension partagée du problème. C'est une phase de concertation et de dialogue autour du problème informationnel afin que les collaborateurs puissent arriver à un consensus concernant leurs représentations du problème. La communication qui est le troisième élément du modèle est très importante, car elle dépeint la phase d'échange des informations à l'intérieur du groupe et vers l'extérieur du groupe. Le partage de connaissances entre les utilisateurs et la complémentarité de leurs connaissances permettent la mise en synergie des compétences. C'est le quatrième élément de la pyramide de collaboration. Quand des individus travaillent ensemble, ils cherchent toujours à avoir un retour sur les activités des uns des autres, ceci afin d'assurer une conscience de groupe. Cet aspect très important dans une collaboration constitue le cinquième élément de la pyramide. Pour réussir une collaboration, il faut un bon mécanisme de répartition des tâches qui permette aux collaborateurs de changer de rôle quand ceci est nécessaire. La répartition des tâches est le dernier élément de la pyramide.

Au cours de ce travail de thèse, il nous est apparu évident que la communication, qu'elle soit interpersonnelle, communication homme-machine ou communication interpersonnelle par machine interposée, est l'un des éléments les plus importants dans la RCI surtout quand il s'agit d'une collaboration synchrone. Nous avons donc modélisé le processus de communication dans le cadre de la RCI. Ce modèle que nous nommons COCIR (*Communication model for Collaborative Information Retrieval*) est une modélisation du contexte collaboratif pour le partage de connaissances et il représente pour nous la **deuxième originalité** de notre travail. Il est composé de quatre éléments : l'expéditeur, l'objet

informationnel, le contexte d'échange et le destinataire. Les éléments du modèle COCIR et leurs attributs nous ont permis de répondre à certaines des questions composant notre problématique. Par exemple, pour passer d'un problème partagé à une compréhension partagée de celui-ci, l'ensemble des attributs des sous-éléments de l'élément « contexte » du modèle COCIR permet aux collaborateurs d'effectuer cette phase dans leurs activités de RCI.

Les attributs de ce modèle permettent également de contextualiser chaque échange dans la collaboration afin de gérer les différents types de connaissances exprimées et ceci permet de faciliter le partage de connaissances entre les collaborateurs. La contextualisation de chaque échange permet aussi d'assurer la sécurité et la protection de patrimoine informationnel dans le SRCI que nous avons développé. Seuls ceux qui participent à la résolution d'un problème ont accès aux informations et connaissances associées au problème. Il est même possible, quand il s'agit de la résolution d'un problème décisionnel sensible, de limiter l'accès des participants aux connaissances exprimées sur les problèmes informationnels (découlant du problème décisionnel) les concernant sans avoir accès à l'ensemble des connaissances sur le problème décisionnel. Dans un cadre correspondant à celui-ci, c'est l'initiateur d'un problème décisionnel qui définit les droits d'accès.

Pour valider nos propositions, nous avons développé un SRCI qui est une implémentation du système et des modèles pour gérer les activités de RCI. Ce système que nous appelons MECOCIR est fondé sur le cadre 3CM, les éléments de la pyramide de collaboration, ainsi que le modèle COCIR. MECOCIR représente pour nous la **troisième originalité** de notre travail. Les deux séries d'expérimentation que nous avons effectuées avec MECOCIR ont prouvé que les six phases du modèle de pyramide de collaboration sont très pertinentes pour la réussite de la RCI.

### **Perspectives**

Comme nous l'avons évoqué dans le chapitre 5, nous ne prétendons pas être exhaustifs par rapport à tous les attributs possibles pour représenter le processus d'échange entre des collaborateurs. Nous continuerons à réfléchir sur d'autres attributs possibles pour représenter le contexte collaboratif du processus de recherche d'information afin d'améliorer notre modèle.

Nous avons remarqué que certains composants logiciels (le module de filtrage collaboratif et l'analyseur des indicateurs) de notre système sont à améliorer. Le calcul de similarité et l'analyse effectuée par ces modules sont statistiques. Dans le cadre de la RCI, nous nous sommes plus particulièrement intéressés au calcul de similarité sémantique entre les activités des utilisateurs afin de pouvoir leur recommander des collaborateurs potentiels. Nous travaillons actuellement sur cet aspect sémantique qui constitue l'une de nos perspectives à court terme.

Le bilan des deux séries d'expérimentation effectuées a montré le besoin d'effectuer d'autres séries d'expérimentation plus profondes avec plus de participants. Nous avons remarqué également qu'il serait nécessaire dans les prochaines expérimentations de permettre aux participants d'utiliser le système MECOCIR pour une plus longue durée (ex. deux semaines) afin d'étudier au mieux leurs comportements informationnels et leurs interactions autour du système. Compte tenu de notre domaine d'application : l'intelligence économique, nous avons également l'intention de tester MECOCIR auprès des veilleurs professionnels pour vérifier l'apport de notre approche à leurs démarches informationnelles.

Dans un futur proche, nous comptons développer et implémenter d'autres mécanismes de répartition des tâches dans MECOCIR. Nous comptons également développer une autre version de MECOCIR qui pourra être installée comme *plug-in* dans les navigateurs web les plus répandus (ex. Mozilla, Internet explorer et Safari). Nous pensons que ceci facilitera l'utilisation du système car les utilisateurs auront moins de choses à apprendre sur l'utilisation du système.

Nous avons remarqué le besoin d'étudier de manière plus profonde et empirique, en utilisant le système MECOCIR, d'autres facteurs sociocognitifs qui influent sur les activités de RCI dans un groupe d'utilisateurs et comment ces facteurs peuvent être pris en considération dans la conception et le développement des SRCI. Ceci représente pour nous une perspective à long terme et nous comptons réaliser cela en collaboration avec d'autres chercheurs en psychologie cognitive et en sociologie.

Pour terminer, nous aimerions souligner que nos expériences au cours de ce travail nous a fait comprendre que la thèse n'est que le début d'une carrière de recherche. Ayant bien pris conscience de cela, nous acceptons qu'il puisse y avoir quelques imperfections dans ce travail

et nous comptons chercher toujours à améliorer notre approche et à avancer dans notre méthodologie de recherche.



# Annexe 1 : Exemples des questionnaires utilisés lors d'expérimentation

---

## 1.1 Premier questionnaire pour la première série d'expérimentations

Merci de bien répondre à ces questions avant d'utiliser le système MECOCIR

1. Quand vous recherchez des informations sur un problème :

a) Utilisez-vous Internet ?

Non       Plutôt non       Plutôt oui       Oui

b) Quel(s) moteur(s) de recherche utilisez-vous souvent sur Internet ?

c) Quel(s) moteur(s) de recherche utilisez-vous moins souvent ?

d) Demandez-vous des conseils à votre entourage (proches, collègues) par rapport à votre problème de recherche d'information avant de vous lancer dans la recherche ?

Non       Plutôt non       Plutôt oui       Oui

e) Allez-vous à la bibliothèque pour rechercher des informations ?

Non       Plutôt non       Plutôt oui       Oui

f) Partagez-vous votre problème avec vos collègues avant de rechercher sur Internet ?

Non       Plutôt non       Plutôt oui       Oui

g) Partagez-vous votre problème avec vos collègues au cours de la recherche d'information ?

Non       Plutôt non       Plutôt oui       Oui

h) Partagez-vous votre problème avec vos collègues seulement quand vous ne trouvez point de solution ?

Non       Plutôt non       Plutôt oui       Oui

i) Avez-vous souvent peur de partager votre problème de recherche d'information avec une autre personne ?

Non       Plutôt non       Plutôt oui       Oui

j) Quand vous partagez votre problème de recherche d'information avec une autre personne méfiez-vous parfois de la personne ?

Non       Plutôt non       Plutôt oui       Oui

k) Partagez-vous les résultats de votre recherche avec d'autres gens ayant des problèmes similaires ?

Non       Plutôt non       Plutôt oui       Oui

l) Avez-vous du mal à juger tout(e) seul(e) la pertinence d'une information ou d'un document retrouvé vis-à-vis de votre problème ou sujet de recherche ?

Non       Plutôt non       Plutôt oui       Oui

m) Désirez-vous que les systèmes de recherche d'information que vous utilisez, intègrent des moyens de partage avec d'autres utilisateurs en temps réel ?

Non       Plutôt non       Plutôt oui       Oui

2. Anticipez-vous comment vous allez rechercher des informations sur un sujet ci-dessous :  
« Que peut-il se passer dans l'organisme de quelqu'un expliquant qu'il ait une odeur épouvantable ? »

3. Simulez avec votre partenaire comment vous allez résoudre ensemble ce problème

Merci d'avoir bien voulu prendre le temps de répondre à ce questionnaire

## 1.2 Deuxième questionnaire pour la première série d'expérimentations

### *Questionnaire d'évaluation de l'approche de recherche collaborative d'information et du système MECOCIR*

Vous venez d'utiliser le système MECOCIR qui est un système de recherche collaborative d'information. Merci de bien répondre aux questions suivantes sur votre expérience.

1. Pensez-vous que la phase de clarification de votre sujet de recherche d'information facilite votre recherche ?  
 Non       Plutôt non       Plutôt oui       Oui
2. Pensez-vous que les fonctionnalités incluses dans ce système vous permettent de mieux clarifier votre problème ?  
 Non       Plutôt non       Plutôt oui       Oui
3. Utilisez-vous souvent le chat lors de la collaboration avec votre partenaire ?  
 Pas du tout       Rarement       Souvent       Très souvent
4. Pensez-vous que ceci améliore le processus de collaboration ?  
 Non       Plutôt non       Plutôt oui       Oui
5. Pensez-vous que le fait de savoir en temps réel ce que fait votre partenaire, améliore le processus de collaboration ?  
 Pas du tout       Un peu       Largement       Très largement
6. Pensez-vous que la division du travail lors d'une collaboration est nécessaire ?  
 Non       Plutôt non       Plutôt oui       Oui
7. Utilisez-vous la fonctionnalité « split search » lors de la collaboration avec votre partenaire ?  
 Pas du tout       Rarement       Souvent       Très souvent
8. Trouvez-vous cette fonctionnalité pertinente à la recherche collaborative d'information ?  
 Non       Plutôt non       Plutôt oui       Oui
9. Est-ce que ce système vous permet de partager vos connaissances avec votre partenaire ?  
 Pas du tout       Rarement       Souvent       Très souvent
10. Est-ce que votre système vous permet d'acquérir des connaissances de votre partenaire ?  
 Non       Plutôt non       Plutôt oui       Oui
11. Est-ce que la fonctionnalité fournie dans l'onglet « Indicator » vous aide à juger la pertinence d'un document à la résolution de votre problème ?  
 Non       Plutôt non       Plutôt oui       Oui



12. Qu'auriez-vous aimé faire que le système ne vous permet pas de faire?

13. Quelle fonctionnalité pensez-vous qu'il manque dans ce système, mais pourrait améliorer la recherche collaborative d'information si elle est intégrée ?

14. Est-ce que le fait de collaborer avec quelqu'un en répondant à un besoin informationnel rend la résolution du problème plus rapide ?

Non

Plutôt non

Plutôt oui

Oui

Merci d'avoir bien voulu prendre le temps de répondre à ce questionnaire

### 1.3 Premier questionnaire pour la deuxième série d'expérimentations

Merci de bien répondre à ces questions avant d'utiliser le système MECOCIR

1. Quand vous recherchez des informations sur un problème :

a) Utilisez-vous la bibliothèque pour rechercher des informations ?

Non       Plutôt non       Plutôt oui       Oui

i. A la bibliothèque, utilisez-vous l'outil informatique pour rechercher des informations ?

Non       Plutôt non       Plutôt oui       Oui

ii. Posez-vous des questions aux bibliothécaires à propos de votre problème de recherche d'information ?

Non       Plutôt non       Plutôt oui       Oui

iii. Recherchez-vous directement dans les rayons de livres à la bibliothèque ?

Non       Plutôt non       Plutôt oui       Oui

b) Utilisez-vous Internet pour rechercher des informations ?

Non       Plutôt non       Plutôt oui       Oui

i. Quel(s) moteur(s) de recherche utilisez-vous souvent sur l'internet ?

ii. Quel(s) moteur(s) de recherche utilisez-vous moins souvent ?

c) Demandez-vous des conseils ou de l'aide à votre entourage (proches, collègues) par rapport à votre problème de recherche d'information avant de vous lancer dans la recherche ?

Non       Plutôt non       Plutôt oui       Oui

i. Demandez-vous de l'aide pour comprendre le sujet de recherche ?

Non       Plutôt non       Plutôt oui       Oui

ii. Demandez-vous de l'aide pour savoir quelle source d'information (internet, moteur de recherche, base de données spécialisée etc.) à utiliser ?

Non       Plutôt non       Plutôt oui       Oui

iii. Demandez-vous de l'aide sur le fonctionnement d'un système de recherche d'information à utiliser ?

Non       Plutôt non       Plutôt oui       Oui

iv. Demandez-vous de l'aide sur comment procéder dans la recherche d'information (comment formuler des requêtes) ?

Non       Plutôt non       Plutôt oui       Oui

d) Demandez-vous de l'aide à vos collègues au cours de la recherche d'information ?

Non       Plutôt non       Plutôt oui       Oui

i. Demandez-vous de l'aide pour comprendre le sujet de recherche ?

Non       Plutôt non       Plutôt oui       Oui

ii. Demandez-vous de l'aide pour savoir quelle source d'information à utiliser ?

Non       Plutôt non       Plutôt oui       Oui

iii. Demandez-vous de l'aide sur le fonctionnement d'un système d'information à utiliser ?

Non       Plutôt non       Plutôt oui       Oui

iv. Demandez-vous de l'aide sur comment procéder dans la recherche d'information (comment formuler des requêtes) ?

Non       Plutôt non       Plutôt oui       Oui

e) Demandez-vous de l'aide à vos collègues seulement quand vous ne trouvez pas de solution ?

Non       Plutôt non       Plutôt oui       Oui

i. Demandez-vous de l'aide pour comprendre le sujet de recherche ?

Non       Plutôt non       Plutôt oui       Oui

ii. Demandez-vous de l'aide pour savoir quelle source d'information à utiliser ?

Non       Plutôt non       Plutôt oui       Oui

iii. Demandez-vous de l'aide sur le fonctionnement d'un système d'information à utiliser ?

Non       Plutôt non       Plutôt oui       Oui

iv. Demandez-vous de l'aide sur comment procéder dans la recherche d'information (comment formuler des requêtes) ?

Non       Plutôt non       Plutôt oui       Oui

f) Lorsque vous demandez de l'aide, expliquez-vous votre objectif de recherche à qui vous vous adressez ?

Non       Plutôt non       Plutôt oui       Oui

g) Méfiez-vous du fait de demander de l'aide à une autre personne par rapport à votre problème de recherche d'information?

Non       Plutôt non       Plutôt oui       Oui

Si oui, pourquoi ? :

-----

h) Partagez-vous les résultats de votre recherche avec d'autres gens ayant des problèmes similaires ?

Non       Plutôt non       Plutôt oui       Oui

i) Avez-vous du mal à juger seul(e) la pertinence d'une information ou d'un document retrouvé, vis-à-vis de votre problème ou sujet de recherche ?

Non       Plutôt non       Plutôt oui       Oui

j) Désirez-vous que les systèmes de recherche d'information que vous utilisez, intègrent des moyens permettant de partager de connaissances et d'informations avec d'autres utilisateurs en temps réel ?

Non       Plutôt non       Plutôt oui       Oui

2. Réfléchissez sur comment vous allez rechercher des informations sur un sujet ci-dessous :  
« Que peut-il se passer dans l'organisme de quelqu'un expliquant qu'il ait une odeur épouvantable ? » Ecrivez les démarches que vous allez suivre pour la recherche.

3. Simulez avec votre partenaire comment vous allez résoudre ensemble ce problème

Merci d'avoir bien voulu prendre le temps de répondre à ce questionnaire

## 1.4 Deuxième questionnaire pour la deuxième série d'expérimentations

### *Questionnaire d'évaluation de l'approche de recherche collaborative d'information et du système MECOCIR*

Vous venez d'utiliser le système MECOCIR qui est un système de recherche collaborative d'information. Merci de bien répondre aux questions suivantes sur votre expérience.

1. Pensez-vous que la phase de clarification de votre sujet de recherche d'information facilite votre recherche ?  
 Non       Plutôt non       Plutôt oui       Oui
2. Pensez-vous que les fonctionnalités incluses dans ce système vous permettent de clarifier au mieux votre problème ?  
 Non       Plutôt non       Plutôt oui       Oui
3. Utilisez-vous souvent le chat lors de la collaboration avec votre partenaire ?  
 Pas du tout       Rarement       Souvent       Très souvent
4. Pensez-vous que ceci améliore le processus de collaboration ?  
 Non       Plutôt non       Plutôt oui       Oui
5. Pensez-vous que le fait de savoir en temps réel ce que fait votre partenaire améliore le processus de collaboration ?  
 Pas du tout       Un peu       Largement       Très largement
6. Pensez-vous que la répartition des tâches lors d'une collaboration est nécessaire ?  
 Non       Plutôt non       Plutôt oui       Oui
7. Avez-vous utilisé la fonctionnalité « split search » lors de la collaboration avec votre partenaire ?  
 Pas du tout       Rarement       Souvent       Très souvent
8. Trouvez-vous la fonctionnalité « split search » pertinente à la recherche collaborative d'information ?  
 Non       Plutôt non       Plutôt oui       Oui
9. Est-ce que ce système vous permet de partager vos connaissances sur le sujet de recherche avec votre partenaire ?  
 Pas du tout       Rarement       Souvent       Très souvent
10. Est-ce que ce système vous permet de partager vos connaissances sur les sources d'information avec votre partenaire ?  
 Pas du tout       Rarement       Souvent       Très souvent

11. Est-ce que ce système vous permet de partager vos connaissances sur le fonctionnement des systèmes de recherche d'information utilisés avec votre partenaire ?
- Pas du tout     Rarement     Souvent     Très souvent
12. Est-ce que le système vous permet d'acquérir des connaissances de votre partenaire sur le sujet de recherche ?
- Pas du tout     Rarement     Souvent     Très souvent
13. Est-ce que ce système vous permet d'acquérir des connaissances de sources d'information de votre partenaire ?
- Pas du tout     Rarement     Souvent     Très souvent
14. Est-ce que ce système vous permet d'acquérir des connaissances de votre partenaire sur le fonctionnement des systèmes de recherche d'information utilisés ?
- Pas du tout     Rarement     Souvent     Très souvent
15. Est-ce que la fonctionnalité fournie dans l'onglet de « Indicator » vous aide à juger la pertinence d'un document à la résolution de votre problème ?
- Non     Plutôt non     Plutôt oui     Oui
16. Qu'auriez-vous aimé faire que le système ne vous permet pas de faire par rapport aux activités de recherche d'information et de la collaboration ?
17. Est-ce que le fait de collaborer avec quelqu'un pour chercher des informations rend la recherche plus rapide ?
- Non     Plutôt non     Plutôt oui     Oui

Merci d'avoir bien voulu prendre le temps de répondre à ce questionnaire



## Résumé

La maîtrise de l'information interne et externe de l'organisation est déterminante pour la réussite du processus d'intelligence économique (IE) visant à la résolution d'un problème décisionnel. La collecte et le traitement de l'information parmi les acteurs impliqués dans le processus d'IE nécessitent un processus de collaboration. Cette thèse porte sur la gestion du processus de collaboration dans la recherche d'information (RI) afin de faciliter la résolution d'un problème décisionnel. Nous avons développé deux modèles et un système de recherche collaborative d'information (RCI) pour faciliter la gestion des activités collectives ainsi que la collaboration synchrone et explicite entre des collaborateurs lors de la RI. Le premier modèle est la pyramide de collaboration composée de six phases nécessaires pour la réussite d'une RCI. Ces phases sont: (1) la phase de confiance de départ, (2) la phase de compréhension partagée du problème à résoudre, (3) la phase de communication, (4) la phase de partage de connaissances, (5) la phase de conscience de groupe et (6) la phase de répartition des tâches. Le deuxième modèle est un modèle de communication pour la RCI. Ce modèle que nous nommons COCIR (*Communication model for Collaborative Information Retrieval*) est une modélisation du contexte collaboratif pour le partage de connaissances lors de la RI. Les attributs de ce modèle permettent de contextualiser chaque échange dans la collaboration afin de gérer les différents types de connaissances exprimées et ceci pour faciliter le partage de connaissances entre les collaborateurs. Ces deux modèles ont été implémentés dans le système MECOCIR que nous avons développé pour valider nos propositions.

**Mots-clés :** Recherche collaborative d'information, intelligence économique, pyramide de collaboration, partage de connaissance, conscience de groupe, collaboration synchrone, modèle de communication, répartition des tâches, système de recherche d'information, COCIR, MECOCIR

## Abstract

The management of an organisation's internal and external information is a determinant factor for the success of economic intelligence (EI) process with the goal of resolving decision problem. Information collection and analysis among EI actors necessitate a collaboration process. This dissertation centres on managing collaborative information retrieval (CIR) within the context of EI in order to facilitate decision problem resolution. We proposed two models and developed a collaborative information retrieval system (CIRS) to manage collective activities and to facilitate synchronous and explicit collaboration among users during information retrieval (IR) process. The first model is a collaboration pyramid made up of six important phases necessary for the success of collaborative information retrieval (CIR). These phases include: (1) initial trust phase, (2) consensus building phase, (3) communication phase, (4) knowledge sharing phase, (5) group awareness phase and (6) division of labour phase. The second model is a communication model for collaborative information retrieval (COCIR). This model is a representation of the collaborative context for knowledge sharing in IR. The attributes of this model facilitate the contextualisation of interactions during collaboration in order to manage expressed knowledge and consequently allow knowledge sharing among collaborators. The two models were implemented in a CIRS called MECOCIR which we developed to validate our propositions.

**Key words :** Collaborative information retrieval, economic intelligence, collaboration pyramid, knowledge sharing, awareness, synchronous collaboration, communication model, division of labour, information retrieval system, COCIR, MECOCIR.