



**HAL**  
open science

# L'annotation pour la recherche d'information dans le contexte d'intelligence économique

Charles A. Robert

► **To cite this version:**

Charles A. Robert. L'annotation pour la recherche d'information dans le contexte d'intelligence économique. domain\_stic.docu. Université Nancy II, 2007. Français. NNT: . tel-00131856

**HAL Id: tel-00131856**

**<https://theses.hal.science/tel-00131856>**

Submitted on 19 Feb 2007

**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.

# **L'annotation pour la recherche d'information dans le contexte d'intelligence économique**

## **THESE**

Présentée et soutenue publiquement le 16 février 2007  
pour l'obtention du

Doctorat de l'Université Nancy 2  
(Spécialité Sciences de l'information et de la communication)

par

**Charles Abiodun ROBERT**

President de Jury : Mr. Mohamed HASSOUN : Professeur à l'ENSSIB (Lyon)

Composition du jury

Rapporteurs :

Mr. Philippe DUMAS : Professeur à l'Université de Toulon

Mr. Laïd BOUSIDI : Professeur à l'Université Jean Moulin, Lyon 3

Examineurs :

Mr. Mohamed HASSOUN : Professeur à l'ENSSIB (Lyon)

Mr. Amos DAVID : Professeur de Sciences à l'Université de Nancy 2  
(Directeur de thèse).

Mme. Odile THIERY : Professeur à l'Université de Nancy 2

Mr. Louis-Philippe LAPREVOTE : Professeur à l'Université de Nancy 2.

**Dédié à :**

*La présence de Makapangyarihang  
La mémoire de Victoria Abike Robert  
L'avenir de Guillaume, P<sup>n</sup> D<sup>n</sup> et John Robert*

## Remerciements

Mes remerciements vont en tout premier lieu au Professeur Amos DAVID, mon directeur de thèse, qui m'a accueilli au sein de son équipe, pour la confiance qu'il m'a accordée et les conseils qu'il m'a apportés durant ces années de recherche.

Je remercie les rapporteurs, Professeur Philippe DUMAS et Professeur Laïd BOUZIDI de m'avoir fait l'honneur de rapporter cette thèse, ainsi que pour l'intérêt qu'ils ont manifesté à l'égard de mon travail et des remarques enrichissantes qu'ils m'ont formulées.

Je remercie Professeur Odile THIERY pour les conseils, les corrections et les remarques avisés qu'elle a su me prodiguer.

Mes remerciements vont aussi au Professeur Louis-Philippe LAPREVOTE et Professeur Mohamed HASSOUN pour la disponibilité dont ils ont su faire preuve.

Je remercie le Dr. Adenike OSOFISAN pour le soutien administratif auprès de University of Ibadan, Nigéria et l'intérêt qu'il a manifesté pour mes travaux et l'aide qu'il m'a accordée.

Je tiens à remercier Hanene MAGHREBI pour sa disponibilité dans les moments difficiles. Mes remerciements aussi à Marie KNIPPER, Mr. Nobert MARCINIAK, Mr. José PEREIRA et Mme Myriam BEYAERT pour leur soutien dans la relecture du mémoire.

Je tiens également à remercier ma famille et mes amis pour leur aide et leur soutien dans les nombreux moments de doute qui jalonnent inévitablement un travail de longue haleine comme celui-ci.

Merci enfin tout particulièrement à mes enfants pour leur soutien infatigable.

## Tables de matières

1. Les pincipaux concepts de notre étude .....	19
1.1. Comment définir l'information ? .....	19
1.1.1. L'information en tant que <i>processus</i> .....	22
1.1.2. L'information en tant qu' <i>objet</i> .....	22
1.1.3. Typologie des informations .....	23
1.1.3.1. L'information BLANCHE .....	24
1.1.3.2. L'information GRISE .....	26
1.1.3.3. L'information NOIRE .....	27
1.1.4. Le document .....	28
1.1.4.1. Les définitions spécifiques d'un document .....	28
1.1.4.2. Définition d'un document <i>selon sa création</i> .....	29
1.1.4.3. Définition d'un document <i>selon son utilisation</i> .....	30
1.1.5. L'annotation .....	31
1.2. L'intelligence et sciences de l'information et de la communication (SIC)....	35
1.2.1.1. L'intelligence concurrentielle (competitive/ business intelligence) .....	38
1.2.1.2. L'espionnage .....	40
1.2.2. L'intelligence économique .....	41
1.2.2.1. L'IE en France .....	43
1.2.2.2. Les acteurs dans le processus de d'IE .....	44
1.2.2.2.1. Le décideur .....	45
1.2.2.2.2. Le veilleur .....	50
1.2.3. Le processus de l'IE .....	52
1.2.3.1. Définition du problème décisionnel .....	54
1.2.3.2. Transformation du problème décisionnel en problème de recherche d'information .....	55
1.2.3.3. Collecte d'informations pertinentes .....	57
1.2.3.4. Validation des informations .....	58
1.2.3.5. Traitement des informations .....	58
1.2.3.6. Interprétation des indicateurs .....	60
1.2.3.7. Décision .....	60
1.2.4. L'intelligence militaire C <sup>4</sup> ISTAR et l'IE .....	62
1.2.5. Processus décisionnels .....	64
1.2.6. Les systèmes informatisés liés aux processus de décision .....	65
1.2.6.1. Les systèmes d'aide à la décision (SIAD) .....	67
1.2.6.2. Système d'Information Exécutifs (SIE) .....	68
1.2.6.3. Les systèmes d'aide à la décision de groupe .....	69
1.2.6.4. Le traitement analytique en ligne (OLAP) .....	70
1.2.7. L'intelligence et l'annotation .....	71
1.2.8. Les problématiques liées à la recherche et au traitement de l'information	72

Chapitre 2.....	73
2. État de l'art sur l'annotation .....	73
2.1. L'annotation dans le domaine de SIC.....	73
2.1.1. Les sciences de l'information et de la communication.....	73
2.1.2. La communication .....	75
2.1.3. La communication et la recherche de l'information .....	76
2.1.4. La recherche d'information .....	78
2.1.5. L'annotation comme un moyen de communication .....	79
2.1.5.1. L'annotation comme un moyen de communication d'un groupe.....	80
2.1.5.2. L'annotation et la cycle communicationnelles .....	81
2.1.5.2.1. La création d'une annotation .....	81
2.1.5.2.2. Les annotations continues.....	82
2.1.5.3. Le cycle informationnel et l'agrégation d'expérience .....	84
2.1.5.3.1. Le partage des expériences à partir d'un seul document .....	85
2.1.5.3.2. Le partage des expériences à partir de plusieurs documents .....	91
2.1.5.4. La désinformation, la propagande et l'annotation .....	92
2.1.5.5. Les outils d'annotation et les outils de communication.....	93
2.1.6. Le concept et les processus d'annotations .....	94
2.1.7. L'annotation dans le processus de recherche de l'information.....	95
2.2. Création et stockage d'annotation.....	97
2.2.1. XML : la base de spécifications des annotations.....	97
2.2.2. Resource Description Framework (RDF) .....	100
2.2.3. Les projets sur l'annotations .....	100
2.2.3.1. ANOTEA .....	102
2.2.3.2. Microsoft Common Annotation Framework (CAF).....	104
2.2.3.3. Extensible MultiModal Annotation Language (EMMA) .....	105
2.2.3.4. Linguistic Annotation Framework (LAF).....	107
2.2.4. Les modèles d'annotation existants .....	109
2.2.4.1. Modèle fondé sur l'organisation du contenu .....	109
2.2.4.1.1. Modèle sémantique.....	109
2.2.4.1.2. Modèle ontologique .....	110
2.2.4.1.3. Les problèmes liés aux modèles organisationnelles .....	112
2.2.4.1.4. Quelques langages d'organisation de contenus.....	112
2.2.4.2. Modèles fondés sur les modes de création d'annotation .....	117
2.2.4.2.1. Les annotations automatiques.....	117
2.2.4.2.2. Annotation semi-automatique.....	119
2.2.4.2.3. Annotation manuelle.....	121
2.2.4.3. Annotation dans les technologies du Web.....	121
2.2.4.3.1. Les dispositions du Web .....	122
2.2.4.3.2. Présentation d'information .....	123
2.2.4.3.3. Stockage d'information .....	125
2.2.4.3.4. Partage d'information .....	127
2.2.4.4. Problèmes liés aux outils d'annotation sur le Web.....	128
2.2.5. Les outils d'annotations spécifiques et leurs spécificités .....	131
2.2.5.1. Grants.....	131
2.2.5.2. One-Thousand Words (OTW) .....	132
2.2.5.3. AMAYA .....	133

2.2.5.4.	L'annotation GeoMedia.....	135
2.2.5.5.	CoNote.....	136
2.2.6.	L'intégration des annotations dans les documents.....	137
2.2.6.1.	Les emplacements des annotations dans un document.....	137
2.2.6.2.	Les annotations non-intégrées sur les documents.....	137
2.3.	Utilisation et intégration d'annotation dans le processus d'IE.....	139
2.3.1.	Un regard sur annotation par son utilisation.....	139
2.3.1.1.	Annotation pour la classification.....	139
2.3.1.2.	Annotation pour la structuration.....	140
2.3.1.3.	Le public dans l'utilisation.....	140
2.3.1.3.1.	Annotation pour un groupe d'utilisateur.....	140
2.3.1.3.2.	Annotation pour un usage personnel.....	141
2.3.2.	Annotation et indexation.....	141
2.3.2.1.	Annotation hors du contexte.....	142
2.3.2.2.	Langage d'annotation.....	142
2.3.2.2.1.	Règle graphique.....	142
2.3.2.2.2.	Règle orale.....	143
2.3.2.2.3.	Règle écrite : explicite ou codée.....	143
2.3.3.	Fonctionnement d'annotation sur un document.....	143
2.3.4.	Granularité d'annotation.....	144
2.3.4.1.	Les valeurs d'annotation et leurs granularités.....	145
2.3.4.2.	Granularité du documents et type des annotations.....	146
2.4.	Conclusion.....	146
Chapitre 3.....		147
3.	Proposition pour la modélisation et la réalisation d'un système d'annotation pour l'IE.....	147
3.1.	Schéma général d'un système d'annotation.....	147
3.2.	Notre choix des paramètres d'une annotation: Le modèle AMIE.....	149
3.3.	Notre choix des types de document possible pour une annotation.....	152
3.4.	La modélisation du système d'annotation.....	155
3.4.1.	Caractéristiques fonctionnelles du système.....	155
3.4.2.	Schéma Entité-Association et relationnel pour les données du système.....	156
3.5.	Schéma fonctionnel du système I-AMIE.....	159
3.5.1.	Le développement du système avec la technologie Internet.....	163
3.5.2.	L'implantation du système I-AMIE au LORIA.....	165
3.5.2.1.	Les composants logiciels du système I-AMIE.....	167
3.5.2.2.	Les interfaces.....	171
3.5.3.	Consultations des annotations existantes.....	178
3.5.3.1.	Affichage des annotations.....	178
3.5.3.2.	Accès et interrogation de base d'annotation.....	179
3.5.3.2.1.	Explorer.....	179

3.5.3.2.2.	Interroger .....	180
3.5.3.2.3.	Analyser .....	185
3.5.4.	Diverse .....	186
3.5.4.1.	Alimentation de la base de document .....	186
3.5.4.2.	Filtration des documents affichés .....	187
3.5.4.3.	Recherche Plus .....	187
3.6.	Domaines d’application du système I-AMIE .....	188
3.6.1.	Application des annotations dans un contexte d’IE .....	188
3.6.2.	Application dans le domaine bibliographique .....	191
3.6.3.	Application dans le domaine d'administration juridique .....	191
3.6.4.	Gestion de développement « Open Source » .....	192
3.6.5.	Application dans les travaux de ressources ouvertes .....	193
3.6.5.1.	WIKI .....	193
3.6.5.2.	DELICIOUS .....	194
3.6.5.3.	Flickr .....	195
3.6.6.	Application globale .....	197
3.7.	Le test et l’analyse du système I-AMIE .....	197
Chapitre 4.	.....	199
4.	Conclusions générale .....	199
4.1.	Conclusion .....	199
4.2.	Perspectives .....	202

## Table des Figures

Figure 1.1 : L'intelligence implique plusieurs processus .....	37
Figure 1.2 : Architecture d'un système d'intelligence économique .....	41
Figure 1.3 : Les étapes de recherche d'information.....	66
Figure 1.4 : Interface d'un system de SAD (LEXSYS 2.3).....	68
Figure 2.1 : Un chercheur d'information communique pour réduire une barrière ...	77
Figure 2.2 : Un extrait du traité Européen de 1951 .....	81
Figure 2.3 : La création d'annotation .....	82
Figure 2.4 : L'annotation continue .....	84
Figure 2.5 : Partage d'annotation en mode modéré.....	86
Figure 2.6 : Partage d'annotation en symbiotique .....	87
Figure 2.7 : Partage d'annotation en propagation.....	88
Figure 2.8 : Partage en mode de transition .....	89
Figure 2.9 : Partage en mode mixte .....	90
Figure 2.10 : Le mise en correspondance des annotateurs et des documents.....	91
Figure 2.11 : L'annotation par rapport au SRI et IE.....	95
Figure 2.12 : L'architecture generale de cadre d'ANNOTEA.....	102
Figure 2.13 : Le cadre d'annotation EMMA .....	106
Figure 2.14 : Un exemple de dépôt sémantique .....	110
Figure 2.15 : La table périodique est un exemple d'organisation ontologique.....	111
Figure 2.16 : Indexation dans Topic maps.....	116
Figure 2.17 : Un exmple d'annotation basée sur le format (GOA) .....	124
Figure 2.18 : Un exemple d'annotation partagée dans le système d'informations LEXSYS .....	128
Figure 2.19 : Une fenêtre de création d'annotation par GrAnt .....	131
Figure 2.20 : Copie d'écran d'annotation d'une image sur OTW .....	133
Figure 2.21 : Interface of AMAYA 9.3 sur <a href="http://www.loria.fr/news">www.loria.fr/news</a> sans les cadres.....	135
Figure 2.22 : Copie d'ecran de CoNote.....	136
Figure 3.1 : Annotation sur un document primaire.....	153
Figure 3.2 : Annotation sur un document secondaire .....	154
Figure 3.3 : Annotation sur des annotations .....	155
Figure 3.4: Schéma Entité-Association de notre système d'annotation.....	157
Figure 3.5 : Graphe de relation de la collection de relations.....	159
Figure 3.6 : Représentation du système I-AMIE .....	160
Figure 3.7 : Présentation de l'architecture 3-tiers.....	164
Figure 3.8: Les interactions entre les composants du système I-AIME .....	166
Figure 3.9 : Les composants logiciels de I-AMIE .....	168

Figure 3 .10 : Utilisation de système I-AMIE pour la recherche d'information .....	170
Figure 3.11: Schéma fonctionnel d'I-AMIE.....	172
Figure 3.12 : Ecran de login .....	173
Figure 3.13: Présentation de système d'annotation .....	173
Figure 3.14 : Ecran de création d'un nouvel utilisateur .....	174
Figure 3.15 : Interface de création d'annotation.....	175
Figure 3.16 : Premier écran dans la création d'annotation.....	176
Figure 3.17 : Zone de création d'annotation .....	177
Figure 3.18 : Ecran de liste des annotations dans la base.....	178
Figure 3.19 : Interface pour explorer le contenu de la base.....	179
Figure 3.20 : Interface pour sélection une variable a exploré.....	180
Figure 3.21 : Interface de recherche des livres de bibliotheque d'université Nancy 2 .....	180
Figure 3.22 : Un essai comparable aux autres systèmes de recherche d'information .....	181
Figure 3.23 : Interface de spécification d'une interrogation.....	182
Figure 3.24 : Interface d'interrogation pour choisir les variables a croiser.....	183
Figure 3.25 : Exemple de résultat d'une interrogation utilisant « OU ».....	184
Figure 3.26 : Interface pour saisir les textes a « analyser » dans la base.....	185
Figure 3.27 : Interface pour alimenter la base de document.....	186
Figure 3.28 : Interface pour spécifier les paramètres de filtration.....	187
Figure 3.29 : Recherche d'information avec annotation.....	188
Figure 4.1 : La mise en correspondance des sujets avec les utilisateurs et les annotations.....	205

## Liste des Tableaux

Tableau 1.1 : Les définitions de l'information blanche.....	25
Tableau 1.2 : Les définitions de l'information grise.....	26
Tableau 1.3 : Les définitions de l'information noire.....	27
Tableau 1.4 : Analogies autour des actions du décideur et veilleur (Kislin et al, 2003) .....	51
Tableau 1.5 : Table de processus de l'IE selon équipe SITE .....	54
Tableau 1.6 : Table de comparaison entre l'IE et l'intelligence militaire .....	64
Tableau 2.1 : La définition de la communication (Alex Mucchielli, 1998) .....	76
Tableau 2.2 : Table de comparaisons des outils communicationnels sur internet.....	94
Tableau 2.3: Les propriétés d'une annotation en format RDF pour ANNOTEA....	103
Tableau 2.4 : Les couleurs des annotations en GENOTATOR .....	119
Tableau 2.5 : La table des systèmes d'annotation spécifiques .....	125
Tableau 2.6 : Tableau des fichiers dans le stockage des informations .....	126
Tableau 3.1 : Tableau comparatif des outils existants .....	149
Tableau 3.2 : Signification des colonnes de l'écran 3.18 .....	179
Tableau 3.3 : Les autres opérateurs qui ne sont pas utilisés .....	184
Tableau 3.4 : Synthèse des exploitation des annotations dans le contexte d'IE .....	190





## **Introduction générale**

On considère que le **siècle de l'information** est l'ère qui suit l'âge industriel. C'est la période à partir de laquelle le mouvement de l'information est devenu plus rapide que le mouvement physique des matériels physiques. Plus spécifiquement, il s'agit des années 80 ou 90 et au-delà. On pourrait dire qu'il n'a réellement commencé que pendant la dernière moitié du XIX<sup>e</sup> siècle, par l'invention du téléphone et de la télégraphie. Il est souvent désigné comme **la période post-industrielle**.

Cette période est dominée par la convergence de plusieurs dispositifs de traitement de l'information et de méthodes de transformations diverses de cette dernière. L'accès à l'information, et sa gestion, devient un problème en raison des dispositifs et des méthodes hétérogènes ou non compatibles, utilisés pour la traiter. L'hétérogénéité de l'information est souvent due à la **diversité des équipements** utilisés : télévision par satellite, téléphone, téléphones mobiles, raccordements à grande vitesse, Voix sur Protocole Internet (VOIP), micro-ordinateurs, radio, Internet sur les médias. Ou encore à la **diversité des moyens** : rayons laser, ondes radio, électricité, électromagnétisme et infrarouges.

Ce siècle de l'information est caractérisé par l'explosion de l'information. La disponibilité de l'information n'est plus réellement en cause; Est-ce que le vrai problème est lié à comment (re)trouver des informations **pertinentes et appropriées** ?

L'identification des informations pertinentes est donc l'un des problèmes majeurs auquel doit faire face aujourd'hui le chercheur d'information. Comment peut-on identifier l'information comme pertinente parmi la masse d'informations disponibles ? C'est la question importante.

Les informations pertinentes sont souvent trouvées dans un état contestable, qui demandera un traitement ultérieur. (Dubois et al, 1997) ont identifié huit situations

différentes d'informations incertaines. Elles se discernent sous les qualificatifs suivants :

- ambiguë : quand l'information fournie ne donne pas un éclaircissement de son contenu,
- brute : - une partie de l'information a besoin d'un traitement approfondi, affiné-,
- biaisée : - les informations sont systématiquement polarisées pour donner une orientation particulière,
- incomplète : quand il manque à l'information une partie nécessaire pour caractériser correctement une situation spécifique,
- imprécise : son contenu ne correspond pas au standard de précision attendu,
- incertaine : l'information est compromise, en raison de la source,
- incohérente : l'information contredit une ou plusieurs autres informations,
- redondante : quand l'information est livrée sous plusieurs formes

Toutes les informations ne sont donc pas forcément utiles et, en particulier dans le domaine communicationnel, on se doit de ne retenir que celles du type "**pertinentes et sans contestation**".

Afin d'accéder aux informations pertinentes sous une forme acceptable, nous avons alors besoin de les traiter.

Une des manières de traiter l'information pertinente est l'application du **concepts et d'outils d'annotations**. L'objectif de ces outils et concepts d'annotation est d'augmenter non seulement la possibilité d'accès à l'information pertinente mais également d'aider à la réduction des difformités.

Dans le cadre de notre étude, nous nous sommes posé les questions suivantes :

- a. La recherche de l'information, préalable à une prise de décision, peut-elle être renforcée par des **outils d'annotation** ?

Des informations pertinentes restent souvent hors de portée de l'utilisateur quand les méthodes d'organisation de l'information ne sont pas bonnes ou que

les algorithmes de recherche sont inadaptés. Mais dans certains cas, cette difficulté peut être liée à un manque d'adéquation entre les mots employés dans la base d'information et les mots utilisés par l'utilisateur pour les retrouver.

Par exemple, dans une base bibliographique, des champs comme « auteur, date de publication, titre » sont utilisés pour décrire les informations que l'on va trouver dans la base. Comment trouver des informations ne sont pas spécifiées par ces attributs ? Par exemple, comment trouver le pays d'origine ou les intérêts personnels de l'auteur d'une publication ? Pouvons-nous étendre la recherche en dehors de ces critères ?

Est-ce que l'intégration d'une base d'**annotations** pourra produire un meilleur résultat en terme de pertinence et en terme de prise en compte des spécificités des utilisateurs ?

Ainsi le classement des informations bibliographiques dans une bibliothèque n'est pas fondé sur les terminologies utilisées par des experts dans leurs domaines spécifiques mais il est élaboré sur la terminologie propre aux bibliothécaires.

Ne serait-il pas intéressant d'utiliser les commentaires d'un expert dans un domaine donné pour classer les publications spécifiques à ce domaine ? Le guidage et les propositions de cet expert, plutôt que les références bibliographiques formelles, ne donneraient-ils pas une meilleure appréhension de l'intérêt de l'ouvrage ?

- b. Peut-on guider, par un modèle approprié, le processus d'**annotation** qui peut soutenir la recherche d'information ?

Les annotations *libres*, -ou les modèles d'annotation existants-, qui ajoutent certes une information, ne sont pas toujours appropriées. Nous supputons que l'utilisation d'annotations prédéterminées par un modèle peut, elle, favoriser une recherche d'information plus efficiente. Quel serait ce modèle ?

- c. Comment les informations annotées peuvent-elles être collectées pour aider à la prise de décision ?

Six méthodes sont utilisées en général pour la collecte de l'information, selon « Harvard Family Research Project » (Harvard, 2004).

- i. Les enquêtes et questionnaires,
- ii. Les entretiens et la constitution de groupes spécifiques,
- iii. Les observations,
- iv. Les revues de documents,
- v. Les tests et évaluations,
- vi. Les revues de sources secondaires.

Y-a-t-il une méthode pour **collecter des annotations utiles** ?

Pouvons-nous combiner certaines de ces méthodes ? Ou sélectionner entre ces propositions ? Est-il possible d'élaborer nous-mêmes une méthode plus efficace dans la collecte d'annotations ?

- d. Comment l'information annotée peut-elle être utilisée pour la prise de décision ? Utiliserons-nous l'information annotée telle quelle, ou une partie des annotations doit-elle être davantage traitée ?

Que devons-nous faire pour que les annotations restent exploitables ?

Y a-t-il des paramètres particuliers qui puissent les valider pour la prise de décision ?

Nous pensons en effet que l'annotation devrait contribuer à la transformation de l'information collectée en des informations à valeur ajoutée qui seront plus adaptées pour la prise de décision ce qui constitue notre **hypothèse pour la thèse**.

Ce travail a pour objectif de proposer **un modèle d'annotation** dans la recherche d'informations, utiles à une prise de décision.

Le but et l'intérêt de l'annotation sont bien l'interprétation de l'information et la transmission d'une connaissance complémentaire. Elle dévoile donc une préoccupation, ou un problème, et parfois résout ce problème. Pour chaque annotation communiquée, il y a mise en évidence du jugement de son créateur.

Dans le cadre de ce travail nous orienterons notre réflexion sur les *annotations d'informations* vers le domaine de l'Intelligence Economique (Concept que nous définirons en son temps).

Des concepts directement connexes à cet objectif doivent être pris en compte, et définis : l'information, les documents, l'annotation, le décideur, le veilleur, l'intelligence et l'Intelligence Economique, et enfin la communication. C'est pourquoi le chapitre 1 présente les principaux concepts de notre étude.

Dans le chapitre 2, nous présentons les travaux qui ont été effectués dans le domaine d'annotation. Nous allons présenter les caractéristiques et les spécificités de l'annotation par rapport les sciences de l'information et de la communication.

Le chapitre 3 est consacré à la conception d'un modèle d'annotation AMIE dans le processus d'intelligence économique. Nous présentons également comment un système I-AMIE fondé sur ce modèle a été développé. Nous montrons l'application de ce modèle dans l'IE en exposant quelques cas particulier.

Dans le chapitre 4 nous présenterons la conclusion et les perspectives liées à ce travail.



# Chapitre 1

*Dans le domaine professionnel, nos méthodes de transmission et d'étude des résultats de la recherche datent de plusieurs générations et ne conviennent plus du tout de nos jours<sup>1</sup>.*

*Vannevar Bush - As We May Think*

*The Atlantic Monthly, July 1945*

## 1. Les principaux concepts de notre étude

### 1.1. Comment définir l'information ?

En raison de la nature dynamique de l'information, il subsiste nombre de problèmes concernant sa caractérisation, sa définition, sa transmission, sa représentation et son formatage. Différentes définitions peuvent être données pour le terme « information », selon leur contexte d'utilisation.

Shannon (Shannon, 1948, page 2) définit l'information comme : *« la mesure de corrélation entre deux objets aléatoires (variable, fonctions, événements, etc.) »*.

Cette information est transmise par un canal de communication à l'aide d'un signal.

Cette définition permet de faire apparaître les éléments importants de la notion d'information qui sont :

- l'information elle-même,
- le signal transmetteur,
- l'émetteur et le récepteur.

*« L'information est la signification humaine attribuée aux données par les moyens des chartes spécifiées employés dans leur représentation »* (National, 1996)

L'information est *« les faits et connaissances déduits des données »*<sup>2</sup> (Dicofr.com)

---

<sup>1</sup> *Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose*

<sup>2</sup> <http://www.dicofr.com/cgi-bin/n.pl/dicofr/definition/20010101002600>

L'information est « *l'élément de connaissance susceptible d'être représenté à l'aide de conventions pour être conservé, traité ou communiqué*<sup>3</sup> » (Futura-Sciences)

Dans l'œuvre de Robert (Reix, 1969, page 4),

« *L'information apparaît comme le moyen et l'objet de la communication. Tout ce qui peut se représenter, s'écrire, se dire pour être communiqué entre hommes ou entre machines constitue de l'information* »

Remarquons que pour nous, « l'information » est un moyen mais également un objet. Le grand Larousse de la langue Française (1996) définit l'information par rapport à cinq domaines :

- **Action** : *L'information est l'action d'informer, de se mettre au courant d'événement*
- **Etat** : *L'information est une nouvelle, un renseignement que l'on communique ou que l'on obtient*
- **Connaissance** : *L'information est un ensemble de connaissances acquises sur quelqu'un ou sur quelque chose*
- **Contenu** : *L'information est le contenu proprement dit des messages transmis*
- **Contenant** : *L'information est un signal par lequel un système donne connaissance de sa position à un autre.*

Selon ODLIS (Online Dictionary of Library and Information Science)<sup>4</sup>, les informations sont : *les données présentées en forme aisément compréhensible, auxquelles une signification a été attribuée dans un contexte d'utilisation. Dans un sens plus dynamique, l'information est le message transmis par l'utilisation d'un moyen de communication ou d'expression. Un message spécifique peut être informatif ou pas ; il dépend en partie de la perception subjective de la personne qui va le recevoir.*

Selon Le Coadic (Le Coadic, 1994), *l'information est une connaissance inscrite*

---

<sup>3</sup> [http://www.futura-sciences.com/comprendre/g/definition-information\\_552.php](http://www.futura-sciences.com/comprendre/g/definition-information_552.php)

<sup>4</sup> [http://lu.com/odlis/odlis\\_i.cfm](http://lu.com/odlis/odlis_i.cfm) (08/04/2003)

(enregistrée) sous forme écrite (imprimée ou numérisée), orale ou audiovisuelle. L'information comporte un élément de sens. C'est une signification transmise à un être conscient par le moyen d'un message inscrit sur un support spatio-temporel : imprimé, signal électrique, onde sonore, etc. Cette inscription est faite grâce à un système de signe (le langage), le signe étant un élément du langage qui associe un signifiant à un signifié : signe alphabétique, signe de ponctuation etc. Le but de l'information reste l'appréhension de sens ou d'êtres dans leur signification, c'est-à-dire le reste de la connaissance ; la transmission du support, de la structure en étant le moyen.

Selon le dictionnaire wikipedia, *L'information au sens commun est le moyen pour un individu de connaître son environnement.*

*On qualifie l'information comme toute donnée pertinente que le système nerveux central est capable d'interpréter pour se construire une représentation du monde et pour interagir correctement avec lui. L'information, dans ce sens, est basée sur des stimuli sensoriels véhiculés par les nerfs, qui aboutissent à différentes formes de perception.*

Au regard de ces définitions, le mot « information » signifie donc deux choses : une action et un objet. Il a pour racines le verbe « *informare* » signifiant ainsi une action, « l'acte d'informer », et le substantif “*informatio*”, « la connaissance communiquée »

Selon le travail de Watzlawick, (Watzlawick 1999, page10) « *une information est le résultat final de la perception et de la transmission* » ; il y a des canaux de communication impliqués dans la transmission de l'information. « *Les canaux de communication peuvent être oraux, écrits, par images, gestuels, parlés, cinématographiques, par toucher, etc...* ». Ainsi, dans un processus de communication trois éléments sont impliqués : l'information, la communication et les canaux de communication.

### 1.1.1. L'information en tant que *processus*

Les définitions ci-dessus impliquent que l'information **doit être créée** par les hommes à la base d'**une donnée**, re-vérifiable, pour être comprise explicitement (et pas simplement perçue) par le public à travers des sens simples.

La compréhension doit être obtenue sans forcément l'aide d'un outil spécial. Sans la participation active des hommes, nous ne pouvons pas avoir de l'information. L'homme doit **traiter** les données existantes et **attribuer une signification spécifique** avant qu'elles deviennent *information*.

L'information doit être communiquée. Nous dirons que l'information est l'acte de l'homme donnant une signification spécifique aux données, par ses activités de transformation et la communication de cette interprétation. C'est seulement quand cette interprétation spécifique est communiquée et comprise par le récepteur que nous pouvons dire qu'il y a information.

L'acte de traitement (transformation) de données est forcément impliqué dans la production d'information.

Le traitement implique un ou plusieurs outils. L'outil peut être un équipement, comme l'ordinateur, ou nos sens de perception, comme les mains, la bouche. Le plus important c'est qu'il y ait un genre de reformulation, de remise en ordre, de changement ou de regroupement des éléments initiaux (les données). La reformulation est faite pour atteindre un objectif préconçu. L'objectif peut être énoncé ou non.

### 1.1.2. L'information en tant qu'*objet*

Trois entités différentes se trouvent entre la source d'un message et sa destination : **l'émetteur**, **le canal** de la transmission, et **le récepteur**. L'information est le message transmis à travers un canal, sur un sujet spécifique, destiné à un récepteur.

Le récepteur prévu peut recevoir un message différent de ce qu'il est censé recevoir. La différence entre message transmis et message reçu peut être due à plusieurs facteurs tels que :

- (a) le bruit dans le canal de transmission
- (b) l'erreur dans l'émetteur ou/et le récepteur.

Nous ne pouvons pas dire qu'il y a transmission d'information tant que cette dernière ne sera pas reçue et comprise par le récepteur. Si le récepteur ne peut pas décoder le message envoyé, il n'y a évidemment pas transmission d'information.

Les modalités impliquées dans la transmission et le codage d'un message sont hors de notre étude, de même que la problématique de la transmission.

Nous admettons que l'émetteur et le récepteur d'un message sont d'accord sur la modalité de transmission de l'information (cela en termes de codage de l'information, et non pas en termes de canal de transmission.)

### 1.1.3. Typologie des informations

Différentes typologies ont été utilisées pour classer les informations, fondées sur les utilisations et les attentions spécifiques. Nous reconnaissons que notre énumération des types d'informations n'est pas exhaustive. Plusieurs critères sont utilisés pour les identifier.

Quelques typologies sont fondées sur les sources d'informations ; certaines considèrent **le médium** utilisé pour communiquer l'information ; celle-ci peut être notamment communiquée par l'utilisation du papier, de signaux électriques, du magnétisme, de l'onde radio et du rayon laser, d'autres considèrent **le moyen de traitement** de l'information. L'information peut être traitée automatiquement, par le canal d'appareils électroniques (ordinateur, data logger, etc...), ou bien de façon manuelle.

Une autre typologie distingue le **public visé** (cible privée ou cible publique). Elle peut se fonder sur la **date de production** (information actuelle ou ancienne). Nous pouvons même avoir une typologie fondée sur **l'utilisation** de l'information : tactique ou stratégique, offensive ou défensive, générique ou spécifique, etc. L'utilisation peut évidemment prendre des aspects différents selon le contexte.

Pour nous la typologie la plus signifiante est celle qui s'appuie sur le **contrôle d'accès à l'information**. Nous nous sommes intéressés à cette typologie car l'accès à l'information est l'une des étapes les plus importantes dans la recherche d'informations avant une prise de décision.

La prise de décision repose donc sur l'accès à l'information. C'est pourquoi nous retenons la classification des informations en information **blanche**, information **grise**, information **noire**, classement métaphorique qui va être éclairci ci-dessous.

### **1.1.3.1. L'information BLANCHE**

Plusieurs définitions ont été données sur les informations blanches. Parmi les définitions, nous en présentons quelques-unes dans le tableau 1.1 ci-dessous.

Nous considérerons que les informations blanches sont les informations dont l'accès et le stockage sont légaux.

<b>Définition</b>	<b>Référence</b>
<i>Acquise légalement, facilement, gratuitement ou à faible coût, ayant peu de portée stratégique</i>	<a href="http://www.zeknowledge.com/veille_information_savoir.htm">http://www.zeknowledge.com/veille_information_savoir.htm</a> (23/04/2006)
<i>Aisément et licitement accessible</i>	<a href="https://www.escp-eap.net/~gtilab/new/materials/ci_3_les_diffrent_types_dinformation.ppt">https://www.escp-eap.net/~gtilab/new/materials/ci_3_les_diffrent_types_dinformation.ppt</a> (23/04/2006)
<i>C'est l'information directement et librement accessible</i>	<a href="http://www.surfandbiz.com/web/veille/">http://www.surfandbiz.com/web/veille/</a> (23/04/2006)
<i>Information accessible facilement et de manière licite, et disponible pour tous</i>	<a href="http://www.abcpresse.com/content/view/31/32/">http://www.abcpresse.com/content/view/31/32/</a> (23/04/2006)
<i>L'information publique et accessible librement, sans sécurisation (et souvent sans sécurité),</i>	<a href="http://www.yves-simony.net/article.php?id_article=85">http://www.yves-simony.net/article.php?id_article=85</a> (23/04/2006)
<i>Ne fait pas l'objet de publicité, mais on peut la trouver de manière indirecte ou détournée - Information sensible d'accès légal</i>	<a href="http://www.decisionnel.net/veille/index.htm">http://www.decisionnel.net/veille/index.htm</a> (23/04/2006)
<i>Quand elle va dans le sens de ce que "tout le monde dit" : elle est évidente et sert à conforter notre opinion globale.</i>	<a href="http://www.mmt-fr.org/article109.html">http://www.mmt-fr.org/article109.html</a> (23/04/2006)
<i>Information aisément et licitement accessible.</i>	<a href="http://www.doubleveille.net/terminologie_veille.htm">http://www.doubleveille.net/terminologie_veille.htm</a> (23/04/2006)
<i>Ce que chacun peut voir et recueillir. Information ouverte (90 % du total). Information publique</i>	<a href="http://www.01net.com/article/220682.html">http://www.01net.com/article/220682.html</a> (23/04/2006)

Tableau 1.1 : Les définitions de l'information blanche

### 1.1.3.2. L'information GRISE

Les différentes définitions de l'information grise sont présentées dans le *tableau*

1.2 :

<b>Définition</b>	<b>Référence</b>
<i>L'information acquise de façon «limite», cible privilégiée des veilleurs ayant une portée stratégique élevée et en dernier lieu d'information</i>	<a href="http://www.zeknowledge.com/veille_informations_savoir.htm">http://www.zeknowledge.com/veille_informations_savoir.htm</a> (23/04/2006)
<i>Licitement accessible, mais caractérisée par des difficultés dans la connaissance de son existence ou de son accès</i>	<a href="https://www.escp-eap.net/~gtilab/new/materials/ci_3_les_diffrent_stypes_dinformation.ppt">https://www.escp-eap.net/~gtilab/new/materials/ci_3_les_diffrent_stypes_dinformation.ppt</a> (23/04/2006)
<i>C'est l'information directement et librement accessible</i>	<a href="http://www.surfandbiz.com/web/veille/">http://www.surfandbiz.com/web/veille/</a> (23/04/2006)
<i>Il s'agit de l'information que l'on peut acquérir de manière indirecte ou détournée, mais d'accès légal</i>	<a href="http://www.surfandbiz.com/web/veille/">http://www.surfandbiz.com/web/veille/</a> (23/04/2006)
<i>L'information accessible de manière licite mais avec certaines difficultés dans sa recherche</i>	<a href="http://www.abcpresse.com/content/view/31/32/">http://www.abcpresse.com/content/view/31/32/</a> (23/04/2006)
<i>L'information peu référencée, sans publicité ni autopromotion,</i>	<a href="http://www.yves-simony.net/article.php3?id_article=85">http://www.yves-simony.net/article.php3?id_article=85</a> (23/04/2006)
<i>Publique et accessible, ne fait l'objet d'aucune sécurisation particulière</i>	<a href="http://www.decisionnel.net/veille/index.htm">http://www.decisionnel.net/veille/index.htm</a> (23/04/2006)
<i>Quand elle apporte des nuances, soumettant à votre sagacité des données moins évidentes et un regard particulier.</i>	<a href="http://www.mmt-fr.org/article109.html">http://www.mmt-fr.org/article109.html</a>
<i>Information licitement accessible, mais caractérisée par des difficultés dans la connaissance de son existence ou de son accès.</i>	<a href="http://www.doubleveille.net/terminologie_veille.htm">http://www.doubleveille.net/terminologie_veille.htm</a> (23/04/2006)
<ul style="list-style-type: none"> <li>- Ce qui résulte de l'« accouchement de l'esprit ».</li> <li>- Par le traitement de l'information (confidences).</li> <li>- Par l'habileté de l'accoucheur.</li> <li>- Ce qui peut être immoral sans être illégal. Information sensible (% croissant).</li> </ul> <i>De l'ouvert au fermé.</i>	<a href="http://www.01net.com/article/220682.html">http://www.01net.com/article/220682.html</a> (23/04/2006)

Tableau 1.2 : Les définitions de l'information grise

### 1.1.3.3. L'information NOIRE

Les différentes définitions de l'information noire sont présentées dans le *tableau*

1.3 :

<b>Définition</b>	<b>Référence</b>
<i>Information acquise illégalement, information de grande valeur...qui peut avoir une très forte portée stratégique</i>	<a href="http://www.zeknowledge.com/veille_informati on_savoir.htm">http://www.zeknowledge.com/veille_informati on_savoir.htm</a> (23/04/2006)
<i>A diffusion restreinte et son accès ou son usage est explicitement protégé</i>	<a href="https://www.escp-eap.net/~gtilab/new/materials/ci_3_les_diffrent _stypes_dinformation.ppt">https://www.escp-eap.net/~gtilab/new/materials/ci_3_les_diffrent _stypes_dinformation.ppt</a> (23/04/2006)
<i>C'est celle obtenue de manière illégale par l'espionnage industriel.</i>	<a href="http://www.surfandbiz.com/web/veille/">http://www.surfandbiz.com/web/veille/</a> (23/04/2006)
<i>Information dont la diffusion est limitée, avec un accès et un usage restreints voire interdits</i>	<a href="http://www.abcpresse.com/content/view/31/32/">http://www.abcpresse.com/content/view/31/32/</a> (23/04/2006)
<i>Information faisant preuve d'un haut niveau de sécurisation (nécessité de mots de passe, par exemple) et réservée à certaines catégories de personnes bien précises</i>	<a href="http://www.yves-simony.net/article.php?id_article=85">http://www.yves-simony.net/article.php?id_article=85</a> (23/04/2006)
<i>Fait l'objet d'une haute sécurisation - Relève de l'espionnage industriel</i>	<a href="http://www.decisionnel.net/veille/index.htm">http://www.decisionnel.net/veille/index.htm</a> (23/04/2006)
<i>Est l'information qui dérange, elle est connue des seuls initiés et n'a souvent que peu de validité tellement elle est soumise à des biais de sélection. Mais elle est aussi parfois annonciatrice des vérités de demain : c'est donc une information à très forte valeur stratégique.</i>	<a href="http://www.mmt-fr.org/article109.html">http://www.mmt-fr.org/article109.html</a> (23/04/2006)
<i>Information à diffusion restreinte et dont l'accès ou l'usage est explicitement protégé.</i>	<a href="http://www.doubleveille.net/terminologie_veille.htm">http://www.doubleveille.net/terminologie_veille.htm</a> (23/04/2006)
<i>Ce qui est protégé par la loi ou le contrat et réservé aux personnes autorisées. Information fermée (10 % du total). Information secrète.</i>	<a href="http://www.01net.com/article/220682.html">http://www.01net.com/article/220682.html</a> (23/04/2006)

Tableau 1.3 : Les définitions de l'information noire.

Après avoir parcouru les concepts qui encadrent la notion d'information il nous faut éclairer plus précisément le support essentiel de ladite information, c'est-à-dire le document.

#### **1.1.4. Le document**

C'est là l'un des mots-clés de notre étude, qui nous rapproche du thème proposé, l'annotation du document dans le processus d'intelligence économique. L'information est contenue dans un document. Plusieurs définitions du mot "document" ont été données ; certaines d'entre elles dans la perspective d'une discipline particulière. Par exemple un documentaliste, reconnaît comme documents : « *tous les objets auxquels les techniques de la documentation pourraient être appliquées* » (Buckland, 1998).

Selon la définition établie par une agence de la Ligue des Nations, l'Institut International pour la Coopération Intellectuelle, en collaboration avec l'Union Française des Organismes de Documentation, un document, c'est : « *Toute base de connaissances, fixée matériellement, susceptible d'être utilisée pour consultation, étude ou preuve. Exemples : manuscrits, imprimés, représentations graphiques ou figurées, objets de collections, etc.* » (Bringay et al 2003).

##### **1.1.4.1. Les définitions spécifiques d'un document**

Selon l'encyclopédie en ligne (wikipédia), un **document** est défini comme étant *"le support physique d'une information. Le support physique est un matériel qui peut tenir l'information. Il est considéré comme physique dans la mesure où il peut être perçu par un ou plusieurs de nos sens."*

A partir de cette définition, une conception ou une notion ne peut être définie comme un document, parce que le monde extérieur ne peut pas saisir une conception

ou une notion. Une notion est tout d'abord contenue, et elle reste, dans le cerveau de celui qui l'a imaginée. En revanche, une notion extériorisée peut devenir un document à la condition d'être transmise. En d'autres termes, c'est la transmission d'une *notion* à travers un moyen physique qui lui donne sa qualité de *document*.

Nous appliquerons cette transformation (de *notion* en *document*) dans le cadre de notre travail :

Un créateur de document, - *l'auteur*-, conçoit une notion ou une conception dans son cerveau. Cette dernière devient document quand elle est transmise à un public potentiel, à travers un support physique.

En bref, nous pouvons définir un document comme ***une idée ou un concept exprimés sur un support***.

Il est clair qu'un document est fortement ancré dans le processus de communication. Le document contient l'information- *informatio*-, et l'information -*informare*- est l'acte de transmission d'un ou de plusieurs messages entre l'émetteur et le récepteur (cf. 1.3.1)

#### **1.1.4.2. Définition d'un document selon sa création**

Dans sa représentation la plus générale, « *un document est une trace de l'activité humaine* » (Prie, 1999, page 23).

C'est cette considération que nous retenons dans notre travail. D'un certain point de vue c'est l'effort intellectuel humain pour représenter des faits, des connaissances et des savoir-faire.

De ce point de vue, les traces de l'activité humaine peuvent inclure des sources diverses, comme des matériaux archéologiques, des édifices, des œuvres cinématographiques, des manuscrits, des gravures, des monuments d'art, etc.

L'**annotation** est, indubitablement, la trace d'une activité humaine. Elle témoigne du temps, de la présence et de l'intérêt d'un annotateur, dans un document (l'annotation), pour un document donné (l'information).

Tous les documents possèdent les caractéristiques suivantes :

- Ils se trouvent dans un conteneur ou transmis par un médium,
- Ils contiennent certaines informations,
- Les informations dans un document sont codées selon un langage spécifique (écrits, graphiques, signes ou sons),
- Le langage selon lequel les informations sont codées possède toutes les caractéristiques d'un langage humain (grammaire, syntaxe, sémantique etc.),
- La durée de vie d'un document est variable en fonction des caractéristiques du son conteneur ou du médium de transmission.

#### **1.1.4.3. Définition d'un document *selon son utilisation***

Etymologiquement, un document est « *ce qui sert à enseigner* »<sup>5</sup>.

Le sens originel n'a rien à voir avec le format ni avec le support physique, mais avec **sa fonction**. "*Le lien entre document et connaissances existe d'emblée de façon claire*" (Kolmayer et Pevrelong, 1999).

Un document « *est un objet construit afin de provoquer la transformation dans l'état d'esprit des utilisateurs à qui le document est communiqué* » (Nicholas et al, 1996, page 3).

L'effet prévu du document sur ses publics potentiels prend une grande importance. Si le document « **provoque une transformation de l'état de ses utilisateurs** », alors notre étude est d'une grande importance, car elle est fondamentalement fondée sur la transmission de la connaissance par le moyen de **l'annotation**. Nous nous introduisons donc le concept d'annotation dans la section suivante.

---

<sup>5</sup> (*document*, d'origine latin est *docere* c.-à-d. enseigner)

### 1.1.5. L'annotation

Nous nous inspirons de certaines définitions pour proposer notre propre définition.

#### 1.1.5.1. Définition générale du concept d'annotation

Différentes définitions ont été données au terme d'*annotation*. Ces définitions sont des points de vue spécifiques, qui ne sont pas tout à fait appropriées à notre étude.

*“Une annotation est une information graphique ou textuelle attachée à un document et souvent placée dans ce document”* (Desmontils et al, 2003, page 3).

Cette définition est certes valable, mais la relation entre l'objet de source et l'annotation n'est pas claire. Or, pour nous, l'association entre document source et annotation est impérative. S'il n'y a aucune association ou lien entre les deux, nous ne pouvons pas parler d'*annotation*. L'association doit être énoncée directement ou indirectement.

*“Une annotation est une explication qui accompagne un texte”* dit le Petit Robert, dictionnaire de la langue française.

Dans cette définition le mot *accompagne* est important, bien que formel, tant il est évident que toute annotation, de fait, n'existe que se rapportant à un « texte »-source.

En revanche, comme on ne peut reconnaître au mot *texte* qu'un seul sens, cette définition exclut les annotations appliquées aux objets graphiques ou aux sonorisations.

Les annotations sont « *Tous les objets associés à d'autres objets par une relation* »

(W3 Workgroup on Collaboration, Knowledge Representation and Automatability)<sup>6</sup>  
Pour nous, tous les objets associés à d'autres objets par une relation n'ont pas forcément une connotation d'*annotation*.

Et d'abord, comment définir un objet ?

Dans le domaine de la science de l'information et de la communication, un **objet** est « *la matière et la désignation caractérisant le contenu d'un article* ». Dans ce cas, « *le paragraphe* » d'un article est associé au titre principal, mais le paragraphe d'un article n'est pas forcément une annotation de son titre.

Selon le Grand Dictionnaire<sup>7</sup>, une annotation est « *un bref commentaire, une explication d'un document ou de son contenu, ou même une très brève description, habituellement ajouté(e) en note après la référence bibliographique du document* ». Cette définition est évidemment intéressante, mais nous paraît un peu réductrice, quant à sa localisation ou son expression.

« *Les annotations sont les commentaires, les remarques, les suppressions, les signets externes etc... qui peuvent être attachés à distance à n'importe quel document du Web ou à une section de document du Web* »<sup>8</sup>

Cette définition est encore plus réductrice puisque ne concernant que les documents du Web

---

<sup>6</sup> An annotation can be loosely defined as "any object that is associated with another object by some relationship" <http://www.w3.org/Collaboration/> (04/10/2005)

<sup>7</sup> [http://www.granddictionnaire.com/btml/fra/r\\_motclef/index800\\_1.asp](http://www.granddictionnaire.com/btml/fra/r_motclef/index800_1.asp) (28-Mar-06)

<sup>8</sup> Annotations are external comments, remarks, deletions, bookmarks etc. that can be attached remotely to any Web document or a selected part of it. Annotations are external comments, remarks, deletions, bookmarks etc. that can be attached remotely to any Web document or a selected part of it ». (<http://annotest.w3.org/annotations>).

### **1.1.5.2. Définition du concept d'annotation dans le cadre de ce travail**

Nous considérons l'annotation comme une action et un objet.

L'annotation comme une action :

- L'annotation est une interprétation d'un objet exprimé sur un document. L'interprétation peut être faite par le producteur de l'objet ou par une autre personne.
- C'est l'expression d'une connaissance ou compréhension spécifique d'un document (document peut être un concept exprimé) sur le même document.

L'annotation comme un objet :

- L'annotation est un objet écrit, oral ou graphique, habituellement attaché à un document, fait pour influencer la perception de ses publics potentiels sur un sujet particulier.

Le sujet de départ est l'annotation, qui est prédéterminée par le contenu d'un document d'origine. La perception et l'interprétation individuelle attribuée au sujet d'un document sont subjectives. Cette subjectivité variant parfois avec le temps; l'annotation peut être évolutive. La signification des annotations faites sur un document peut changer avec le temps.

L'évolution d'un sujet d'annotation de son sujet d'origine est une fonction d'interprétation subséquente. Le sujet de notre étude porte sur cette évolution.

Les annotations sur les document écrits renvoient à diverses entités : un ensemble de documents, un seul document, un passage, une phrase, un terme, un mot, une image. Les annotations décrivent (récapitulent le contenu important) et évaluent (analysent) la ressource sur la base des critères standards.

Une annotation n'est pas un résumé, ni un sommaire : les résumés et les sommaires, habituellement décrivent ou récapitulent le contenu, mais ne l'évaluent pas. Des

annotations sont destinées à évaluer des livres, des sites Web, des articles, des documents de gouvernement, des vidéos ou d'autres types de document.

Les annotations peuvent prendre plusieurs formes, comme :

- des icônes (pour décrire des avis; en utilisant des étoiles, des points d'interrogation, d'exclamation...),
- des symboles de liens (pour décrire des associations, des relations entre mots...),
- des notes textuelles en marge, en bas de page ou en fin de document, repérées dans le texte par des icônes (numéros, étoiles...),
- des mises en formes typographiques (surligner, italique...),
- des redécoupages de texte à l'aide d'accolades, de numérotation de passages....,
- des images, des sons,
- des concepts et leurs attributs (annotations sémantiques)...

Cette énumération laisse entendre que l'annotation, sur un document, peut s'exprimer avec un large éventail de possibilités.

Nous identifierons deux types d'annotations, cette dichotomie étant fondée sur l'origine de l'annotation :

(a). *Les annotations volontaires* (ou d'origine intérieure). Elles sont les annotations produites d'instinct par la personne-même qui a créé le texte. Ce sont des remarques, ou des corrections personnelles sur des documents et en un temps que l'auteur a choisis. Ces annotations ne présentent aucun caractère de contrainte, ni de crainte.

(b). *Les annotations requises* (ou d'origine extérieure). La créateur d'annotation est prié de s'engager dans le processus d'annotation; L'annotation est une réponse à une question. Le plus souvent dans les organisations, les annotateurs se trouvent alors être des témoins légaux. Ce groupe semble être le plus important, en nombre, dans la dichotomie présentée ici.

Afin de mieux apprécier l'impact d'annotation dans le processus d'IE, nous étudions de près ce dernier dans la section suivante.

## **1.2. L'intelligence et sciences de l'information et de la communication (SIC)**

Nous pouvons distinguer plusieurs aspects d'intelligence par affiliation aux sciences de l'information et de la communication. La propriété de « *discernement et d'évaluation* » est le terme fondamental pour toutes les dispositions d'intelligence et leurs applications (intelligence économique, intelligence d'affaires, intelligence militaire, espionnage etc...).

On parle d'intelligence lorsqu'il y a un problème à résoudre. Qu'est-ce donc l'intelligence ?

L'intelligence est définie comme « *la capacité d'acquérir et appliquer la connaissance* »<sup>9</sup>.

Selon le dictionnaire en ligne wikipedia : *L'intelligence est la capacité de résoudre des problèmes. C'est l'ensemble des fonctions mentales ayant pour objet la connaissance conceptuelle et rationnelle (Par opposition à la sensation et à l'intuition). Aptitude à comprendre et à s'adapter facilement à des situations nouvelles.*

Les extraits suivants sont tirés de « *Terminologie de neuropsychologie et de neurologie du comportement* » disponible en ligne sur wikipédia.com

*L'intelligence est la capacité à découvrir un contexte nouveau, à le comprendre et à réagir à cette nouvelle situation de façon adaptée.* (Berube, 1991, page 2)

*L'intelligence, c'est ce qui permet d'entendre une musique là où d'autres n'entendent qu'un bruit.*

---

<sup>9</sup> <http://www.thefreedictionary.com>

*L'intelligence, modelée par le patrimoine génétique et par l'environnement culturel et affectif, dépendrait d'un équilibre subtil entre un mode de traitement rapide des informations et un mode lent d'analyse de ces données.*

*Faculté de connaître, de comprendre et de s'adapter, faculté d'abstraction et d'anticipation ou encore intuition. Pour connaître, pour comprendre et pour s'adapter, en un mot pour faire preuve d'intelligence, il faut que les perceptions et les sensations que reçoit le système nerveux central aient un sens, c'est-à-dire que les différents stimuli sensoriels soient associés, classés par les structures cérébrales nécessaires au traitement de l'information et mémorisés. Les étapes de ce traitement, par exemple la rapidité de perception d'un stimulus, son stockage, son évaluation par rapport aux informations déjà enregistrées, son analyse et la réaction qu'il déclenche, sont des facteurs de l'intelligence.*

Le terme "Intelligence" vient du latin « *intellegere* », dont le préfixe *inter* (entre), et le radical *legere* (lier) suggèrent essentiellement l'aptitude à relier des éléments qui, sans elle, resteraient séparés.

Ainsi en est-il de la lecture, en tant qu'aptitude à saisir le sens qui unit les mots, les phrases, les paragraphes etc., au-delà du simple déchiffrement discontinu, des termes inscrits sur le papier.

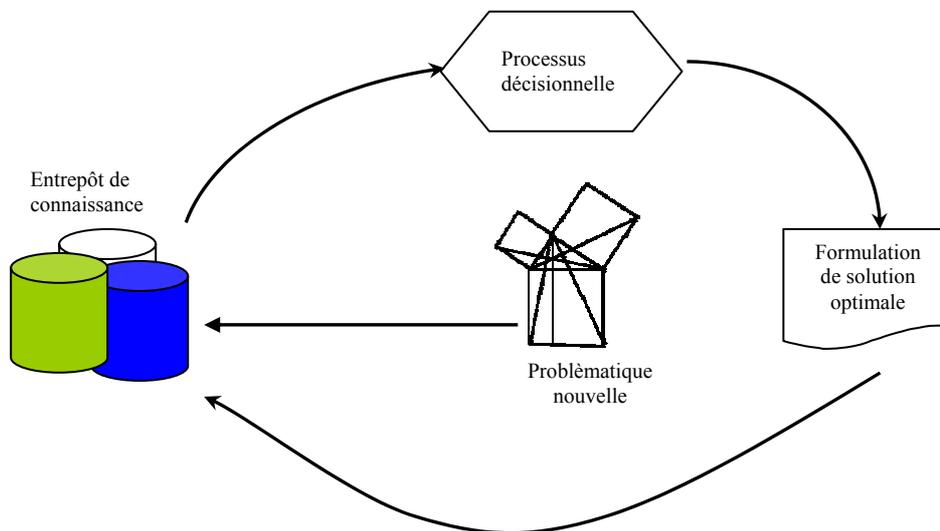
C'est donc la capacité à saisir (et savoir utiliser) des liens entre des éléments disparates. C'est savoir lire au sens le plus large (Lire les signes écrits par l'homme, mais aussi les signes inscrits dans la nature).

L'intelligence est liée à la mémoire du passé pour produire une solution vérifiable à un problème du moment présent, puis elle en magasine la solution trouvée, en vue d'un usage futur.

Les étapes suivantes sont invariablement présentes dans un processus d'intelligence :

- i. chargement de la mémoire,
- ii. recherche dans la mémoire,

- iii. faire un choix, fondé sur une certaine logique, entre les solutions qui ont été recherchées et
- iv. renvoi de la nouvelle solution choisie, pour être archivée à son tour.



*Figure 1.1 : L'intelligence implique plusieurs processus*

Le processus d'intelligence n'est pas un processus intrinsèquement continu. C'est un processus spontané, qui est lancé en présence d'un problème non résolu, dans les conditions environnementales existantes et avec les ressources du moment présent.

Le problème demande le rappel des acquis qui ont été stockés lors d'expériences précédentes.

Il se finit également en stockant la nouvelle expérience acquise dans le même entrepôt de connaissance.

Ce système d'intelligence appréhende un seul problème à la fois, non pas deux problèmes simultanément, même s'il est capable de se référer à des bases de données hétérogènes pour résoudre le problème en question.

L'ouvrage « *The Art of War* » (dans la langue chinoise: 孫子兵法; Hanyu Pinyin : Sūn Zǐ Bīng Fǎ ; littéralement "*La stratégie militaire de Sun Tzu*") (Lionel, 1910) est un traité militaire chinois écrit durant le VI<sup>ème</sup> siècle avant JC par Sun Tzu. Il a été considéré comme le premier travail définitif sur les stratégies et les tactiques intelligentes.

Ce traité est une des bases des plus célèbres pour apprendre la stratégie et il a eu une influence majeure sur la planification militaire et les travaux sur l'intelligence.

Considérons donc des domaines importants dans lesquels l'intelligence est mise en œuvre et qui ont un rapport étroit avec notre préoccupation, l'information.

Pour mieux discerner, un peu plus loin, les données de l'intelligence économique, il semble utile de lui opposer d'autres usages de l'information, l'intelligence concurrentielle et l'espionnage

### **1.2.1.1. L'intelligence concurrentielle (competitive/ business intelligence)**

L'intelligence concurrentielle a été définie comme « *processus formalisé, pourtant sans interruption d'évolution, par lequel l'équipe de gestion évalue l'évolution de son industrie et ses capacités, et le comportement de ses concurrents, et de ses potentiels concurrents, pour aider à maintenir ou à développer un avantage concurrentiel* » (Prescott et Gibbons 1993).

Le projet d'intelligence concurrentielle essaie de fournir à un organisme intéressé des informations précises sur ses concurrents, ainsi qu'un plan pour l'usage, à son avantage, de cette information (Selon McGonagle et Vella, 1990).

Ici, l'idée centrale est que l'information est employée intelligemment pour l'analyse du marché, afin de devancer les concurrents sur un marché commun.

L'un des précurseurs dans ce domaine définit l'intelligence compétitive comme suit :

L'intelligence compétitive - ou concurrentielle- est le processus qui permet d'augmenter sa compétitivité sur le marché par une meilleure compréhension des concurrents et des conditions de la concurrence. (SCIP, 2002) <sup>10</sup>.

Dirk (Vriens, 2004, pg 15) considère que l'intelligence compétitive est plus qu'un processus. Il l'a examinée dans quatre perspectives :

- Comme **Produit** : les informations sont pertinentes pour des objectifs stratégiques.
- Comme **Processus** : Le cycle d'intelligence a été identifié dans quatre étapes, la direction, la collection, l'analyse et la diffusion.
- Le **But** : L'intelligence compétitive est activée pour livrer un outil efficient dans l'objectif d'une décision stratégique.
- La **Différence** : elle diffère de l'espionnage, de l'intelligence d'affaires et d'autres genres d'intelligence et d'étude de marché.

Prior en (Vernon, 2005, Page 3) définit l'intelligence d'affaires (business intelligence) ainsi :

*L'intelligence d'affaires est n'importe quelle combinaison de **données**, d'**information**, et de **connaissance** au sujet de l'environnement d'affaires dans lequel une entreprise réalise ses affaires (au moment nécessaire), elle confèrera un avantage concurrentiel significatif qui leur permettra de prendre des décisions intelligibles.*

Selon lui, les opérations d'intelligence d'affaires permettront une organisation pour, entre autres :

- Prévoir et gérer les risques
- Chercher les opportunités et les nouveaux marchés ;

---

<sup>10</sup> *Competitive intelligence is the process of enhancing market place competitiveness through greater understanding of a firm's competitors and competitive environment*

- Agir avant des concurrents,
- Innover,
- Exploiter les faiblesses des concurrents,
- Améliorer la planification et la prise de décision.

La traduction d'intelligence d'affaires est souvent le traitement et l'interprétation d'information. Par exemple, la définition suivante a été donnée en <sup>11</sup>:

*« L'intelligence d'affaires » (Business Intelligence) ou « informatique décisionnelle » englobe les solutions informatiques apportant une aide à la décision avec, en bout de chaîne, rapports et tableaux de bord de suivis, à la fois analytiques et prospectifs.*

Le but est de consolider les informations disponibles au sein des bases de données de l'entreprise.

### **1.2.1.2. L'espionnage**

L'espionnage est la pratique qui consiste à surveiller les autres (états ou autres entités) afin d'obtenir des informations pour son propre avantage, économique, politique ou militaire. La plupart des nations surveillent habituellement leurs ennemis et leurs alliés, bien qu'elles le nient. L'espionnage fait pour des buts commerciaux est nommé *espionnage industriel*. Le *dictionnaire Wikipédia en ligne* définit l'espionnage comme :

*« La récolte de renseignements que le détenteur peut légalement dissimuler et se réserver. L'espionnage est donc illégal et clandestin »* L'espionnage par un citoyen de l'état- cible, son propre Etat - est généralement considéré comme une forme de trahison.

---

<sup>11</sup> [http://solutions.journaldunet.com/dossiers/pratique/business\\_intelligence](http://solutions.journaldunet.com/dossiers/pratique/business_intelligence)

## 1.2.2. L'intelligence économique

Le rapport de Martre (Martre, 1994) définit l'intelligence économique comme étant :  
« les actions coordonnées de recherche, de traitement et de distribution, en vue de son exploitation, de l'information aux acteurs économiques. Ces diverses actions sont menées légalement avec toutes les garanties de protection nécessaires à la protection du patrimoine de l'entreprise... »

Et selon C. Revelli (Revelli, 1998) c'est le processus de collecte, de 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.

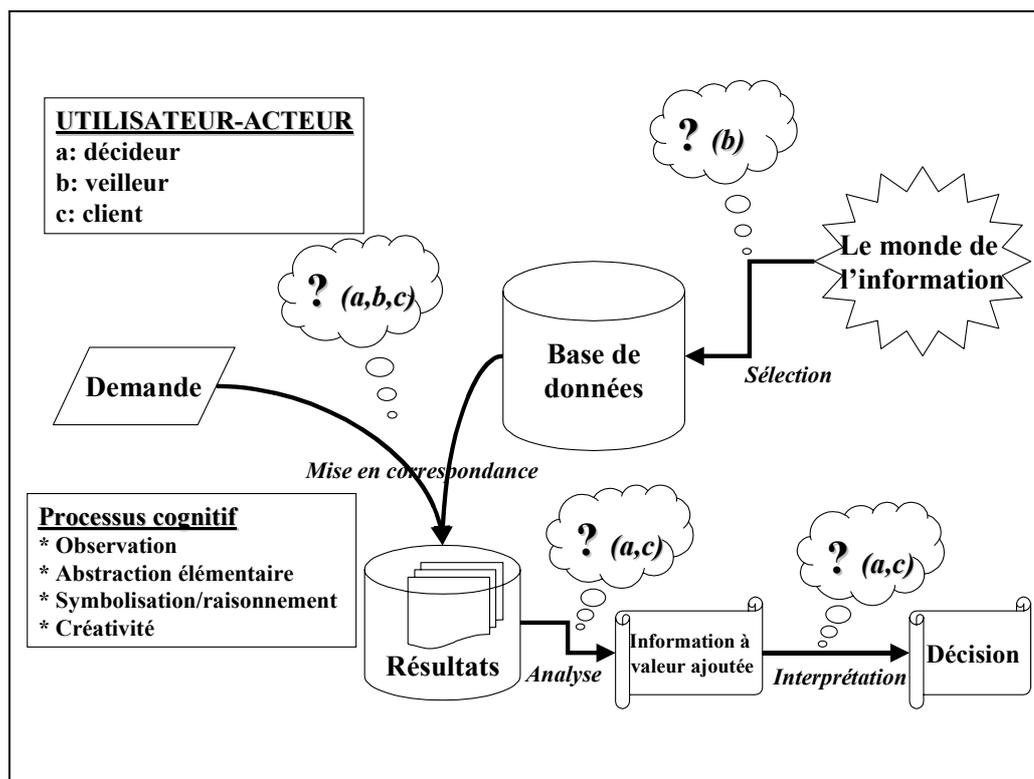


Figure 1.2 : Architecture d'un système d'intelligence économique

L'extrait suivant vient des travaux de (Thiéry et David, 2002). Nous précisons qui sont les acteurs dans les processus dans la section 1.4.2.2.

Le processus d'Intelligence Economique (IE) repose en particulier sur l'utilisation de systèmes d'informations stratégiques (SIS). Les systèmes d'informations (SI) existent depuis fort longtemps. Ils ont subi des évolutions profondes, soit par le but final de l'utilisation du système, soit par le type d'information gérée, soit par la combinaison des deux. C'est ainsi que nous voyons apparaître la déclinaison du terme SIS en « système d'information » « stratégique » (SI-S) et « système » d'« informations stratégiques » (S-IS).

Afin de faciliter le processus de recherche d'information dans les Systèmes d'Information (SI), les techniques de modélisation de l'utilisateur ont été introduites. Ces techniques sont adaptées à la nature des besoins en information des utilisateurs. Certains besoins sont plus stables que d'autres. Dans ce cas, des techniques de filtrage de l'information sont proposées pour recevoir les nouvelles informations. D'autres besoins sont dynamiques. Dans ce cas, des techniques de modélisation, implicites ou explicites, sont proposées. Ces dernières peuvent être globales, quand il s'agit de stéréotypage, ou individuelles.

La *figure 1.2* montre l'importance d'un système d'information (SI) dans le processus d'Intelligence Economique. Suivons les flèches :

- **Sélection** : Elle permet de constituer le Système d'Information de l'entreprise qui peut être (i) la base de données de production (celle qui permet l'exploitation courante de l'organisation), (ii) l'ensemble des informations support d'un système de recherche d'informations (en documentation par exemple) ou (iii) un système d'information stratégique reposant sur un entrepôt de données. Ce SI est constitué à partir de sources de données hétérogènes, à l'aide d'un filtrage de la réalité.
- **Mise en correspondance** : La mise en correspondance permet à tout type d'utilisateur d'accéder aux informations du SI. Deux principales méthodes d'accès à l'information sont actuellement proposées aux utilisateurs : accès par **exploration** et accès par **requête**. L'exploration est basée sur la technique d'hypertexte. Les requêtes sont exprimées à l'aide d'opérateurs booléens. Le résultat de la mise en correspondance est un ensemble d'informations.

- **Analyse** : Afin de donner de la valeur ajoutée aux informations trouvées, des techniques d'analyse sont appliquées au résultat. Par exemple, l'assistante d'un chef de service que nous considérons comme un veilleur pourra établir des tableaux de bord pour son chef de service. Ainsi, les rapports fournis par l'assistante, qui connaît bien les souhaits du patron, seront une bonne base de décision.
- **Interprétation** : Il s'agit là de permettre au décideur, ou en général au client du système, de prendre les bonnes décisions. L'idée est que le décideur n'est pas forcément le client du système. Ce peut être, par exemple, un veilleur. On voit alors tout l'intérêt de capturer des connaissances sur le décideur et de les mettre dans les métas données de l'entrepôt afin de construire une base métier spécifique à un groupe de décideur ou mieux encore à un décideur particulier.

En résumé, notre centre d'intérêt porte sur l'intégration du modèle de l'utilisateur dans la conception d'un S-IS (« système », « d'Information Stratégique ») que nous pouvons considérer comme un système d'intelligence économique (SIE). Le modèle de l'utilisateur aura une application plus large que le filtrage de l'information. Il s'agit de proposer un SIE qui s'adapte aux différents acteurs impliqués dans un processus d'IE.

### **1.2.2.1. L'IE en France**

En 1994, la France a formé une commission de haut niveau sous le bureau du Premier Ministre. La Commission était dirigée par Martre (Martre, 1994). Sa mission était d'étudier comment un nombre de pays étrangers choisis intégraient l'Information Economique qu'ils obtenaient de diverses sources, gouvernementales et non gouvernementales. Parmi les pays étudiés il y avait : le Japon, la Suède, l'Allemagne, les Etats-Unis, et la Grande-Bretagne.

Lorsque la Commission a étudié les systèmes que les différents pays développaient pour intégrer le commerce, l'intelligence et l'information de propriété industrielle, avec l'information de source ouverte, elle a estimé que la France était bien en retard. Elle conclut que la France, bien qu'elle possédât les composants désirés, souffrait

d'un manque d'attention et n'intégrait pas les activités et les sources d'informations qui auraient pu -qui auraient dû- déclencher une action nationale efficace (Joyal, 1996)

L'intelligence économique a vu son champ substantiellement élargi en juin 2003, après la remise au Premier Ministre d'un rapport intitulé «intelligence économique, compétitivité et cohésion sociale » par Bernard Carayon, député (UMP) du Tarn Ce dernier présente les propositions de la Commission comme une politique publique destinée à garantir la cohésion sociale en assurant le développement économique.

On trouvera ci-dessous une partie de la lettre du Premier Ministre, tirée de l'appendice à cette publication :

*« La compétition économique mondiale contraint nos entreprises à des efforts permanents d'innovation de prospection, de qualité et de rentabilité. Pour conduire au mieux ces actions, les acteurs économiques ont besoin d'une information fiable et prospective, et doivent pouvoir se prémunir contre des accès non souhaités à leurs propres données. Le concept d'intelligence économique recouvre ces préoccupations. Apparue en tant que tel dans les pays anglo-saxons et au Japon, il y a plusieurs décennies, force est de constater qu'il n'occupe pas encore en France une place digne des enjeux qu'il doit prendre en compte.*

.....

*Pour ces raisons et parce que la France ne peut se permettre de négliger un outil essentiel à sa performance économique, j'ai décidé de vous confier une mission de réflexion et de proposition sur ce thème. En conséquence vous serez placé, en application des dispositions de l'article I.O. 144 du code électoral, en mission auprès du ministre de l'intérieur, de la sécurité intérieure et des libertés locales à compter du 1<sup>er</sup> janvier 2003 et pour une durée de six mois. »*

### **1.2.2.2. Les acteurs dans le processus de d'IE**

Avant de présenter le processus d'Intelligence Economique il nous faut connaître les acteurs de ce processus. Il existe deux acteurs principaux : le veilleur et le décideur. Nous pouvons identifier d'autres acteurs ; cette identification étant fondée sur leurs fonctions dans la résolution des problèmes décisionnels.

### **1.2.2.2.1. Le décideur**

Dans le travail de (David et al, 2001), les auteurs donne une définition du décideur dans l'environnement interne et externe de l'entreprise comme :

*« celui qui est apte à identifier et à poser le problème à résoudre en termes d'enjeu, de risque ou de menace qui pèse sur l'entreprise ».*

Une autre définition donnée par (Schneider, 1996) repose également sur le rôle joué par le décideur dans son environnement. Ce dernier considère le décideur comme

*«un acteur social. Il doit être décrit et traité comme un système ouvert qui, activement, saisit de l'information, qui la traite et qui agit. Il possède une certaine indépendance cognitive ».*

En se fondant sur ces deux définitions nous considérons que le décideur, est, parmi les autres acteurs, l'observateur le mieux placé dans son environnement interne et externe, et c'est grâce à sa position qu'il est apte à détecter les signaux de l'environnement (Bouaka et al, 2002).

Ci-dessous, nous allons nous inspirer du modèle du décideur présenté par (Bouaka, 2002) et Thiery (Thiery et al, 2002), pour le montrer face à un problème décisionnel ; cette représentation jouera un rôle très important dans le dialogue entre le veilleur et le décideur dans le cadre de décisions stratégiques.

#### **1.2.2.2.1.1. Les paramètres liés au décideur**

Le décideur est celui qui a identifié le problème. Sa version des faits est par conséquent fondamentale. Nous le considérons comme un producteur de connaissances centrées sur un problème décisionnel. Il est capable de traduire le problème en terme d'enjeu comme nous venons de l'expliquer dans la section précédente. Nous considérons que la première dimension clé de l'explicitation d'un problème décisionnel est relative au décideur du fait qu'il est le mieux placé pour identifier l'émergence d'un problème décisionnel. L'intérêt porté à l'intégration du profil de l'utilisateur dans les systèmes d'informations est présenté par (Alloway,

1976). Il considère que l'évolution de « tout système d'information concerne en dernier ressort un individu pourvu d'un profil psychologique donné, confronté à un problème précis dans un contexte organisationnel déterminé, problème pour lequel il (l'individu) a besoin d'éléments de prise de décision, ces éléments étant perçus au travers d'un mode de représentation propre au décideur ». Dès lors, il convient d'accorder une grande importance aux comportements individuels vis-à-vis de l'information.

De ce fait, pour modéliser le décideur dans le contexte d'explicitation d'un problème décisionnel, nous avons retenu après la validation de nos propos les paramètres suivants :

#### **1.2.2.2.1.2. Le style cognitif**

Le style cognitif peut être défini comme la façon propre à chacun de percevoir et de comprendre l'information reçue 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.

D'après (Hayes, 1998), « Le style cognitif a une influence sur la façon dont les individus scrutent leur environnement pour recueillir de l'information, sur la façon dont ils organisent et interprètent cette information et sur la façon dont ils intègrent leurs interprétations dans les modèles mentaux qui guident leurs actions ». Le style cognitif est soit analytique, soit pragmatique.

L'intérêt d'intégrer des informations sur le style cognitif se manifeste surtout au niveau de la présentation, au décideur, des résultats de la recherche d'information. Ces résultats peuvent être sous forme de détails techniques, de données statistiques et de graphiques, rapports, etc. Ces résultats doivent être présentés de manière à ce qu'ils soient compréhensibles par le décideur et surtout ils doivent respecter le style cognitif du décideur.

### **1.2.2.2.1.3. Les traits de personnalité**

La personnalité est définie par (Darmon, 1993) comme « un ensemble de structures cognitives au fil du temps par les individus pour faciliter l'ajustement aux événements, aux individus et aux décisions ».

Pour identifier les traits de personnalité d'un décideur, en tant qu'individu, nous nous sommes intéressés tout d'abord aux travaux faits en psychologie sur l'identification de la personnalité d'une personne. Ensuite nous avons identifié les principaux traits qui caractérisent un décideur à l'aide des travaux de psychologie appliqués dans le domaine de management, en particulier des ressources humaines.

Parmi les instruments les plus utilisés en psychologie, nous pouvons citer l'indicateur de personnalité de Myers et Briggs (Cauvin, 1994). Cet indicateur permet d'analyser les types de personnalité en partant des quatre paires de caractéristiques suivantes :

- la première paire d'indicateurs est déterminée par rapport à l'orientation de l'énergie, il s'agit de : Introverti/Extraverti,
- la deuxième paire d'indicateurs est déterminée par rapport aux modes de réception. Il s'agit de : Sensation/Intuition,
- la troisième paire d'indicateurs est basée sur les critères de décisions. Il s'agit de la Réflexion/sensibilité connue aussi sous le nom Pensée/Sentiment,
- la dernière paire d'indicateurs est identifiée par rapport au style de vie de l'individu. Il s'agit de: Jugement/perception.

Si on arrive à faire une corrélation entre les différents types de personnalité, on peut détecter les potentialités de chaque personnalité et ses limites. L'objectif est de combiner à la fois le mental, l'émotionnel et l'instinctif. Nous considérons que ceci va permettre au veilleur d'adopter une communication correspondant à la personnalité du décideur et d'élaborer, à partir de ces besoins spécifiques, les actions informationnelles les plus adaptées.

#### **1.2.2.2.1.4. Les facteurs liés à l'organisation**

Par la modélisation de l'organisation nous cherchons à identifier l'effet de l'observation faite par le décideur sur l'organisation. En effet, cette modélisation ne prend pas en considération ni la taille ni l'activité de l'organisation. Nous considérons donc les caractéristiques de l'organisation comme étant un déterminant prépondérant du comportement du décideur à l'égard d'un problème décisionnel identifié. En effet, nous mettons l'accent surtout sur toute entité qui a le potentiel d'être source d'enjeu, soit la source ou la cible d'une action de l'entreprise. Ces caractéristiques permettent d'avoir une idée de la complexité, l'incertitude et la stabilité de l'événement observé. En effet, nous considérons que tout événement observé peut être décomposé en trois éléments. Ces éléments sont :

- l'objet de l'environnement,
- le signal émis par cet objet,
- et enfin l'hypothèse que nous pouvons déduire de la détection de ce signal.

Notre propos est que l'observation d'un signal reste dépendante de l'enjeu que porte cette observation, même si cet enjeu est un peu flou au début de l'observation. En effet le décideur ne réagit que lorsqu'il se sent menacé par l'apparition d'un problème ou le risque de perdre une opportunité. Nous supposons 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 à cette observation. De ce fait, la dimension de l'enjeu demeure une dimension fondamentale pour passer du niveau de l'explicitation d'un problème décisionnel au niveau du problème de recherche d'information dont nous parlerons dans la section suivante. Par conséquent, pour réussir cette étape (l'identification de l'enjeu), fondamentale dans le modèle que nous proposons, le décideur et le veilleur s'engagent dans un processus d'adaptation et de rétroaction afin de valider ou modifier les valeurs des paramètres de l'enjeu (objet, signal, hypothèse) identifié par le décideur au début de ce processus. A ce niveau deux cas peuvent se présenter :

- Si les valeurs des paramètres de l'enjeu restent identiques, le problème décisionnel ne change pas de nature, c'est la compréhension du problème qui évolue grâce à l'apparition des informations complémentaires.
- Si une ou plusieurs valeurs des paramètres de l'enjeu sont modifiées, le problème décisionnel change de nature et provoquera la définition d'un nouveau problème décisionnel.

Nous considérons aussi que le fait de comprendre la nature des influences qui nous gouvernent peut nous aider à affiner notre perception de l'enjeu et donc à améliorer notre perception du risque ainsi que notre performance. Nous considérons que la perception de l'enjeu par le décideur influence forcément sa propension à rechercher ou à éviter le risque. Or la notion de perception de l'enjeu n'est pas si simple, car le décideur peut faire appel à des références personnelles. C'est ce qui était montré dans la section concernant la modélisation du décideur. Dès que l'on émet une hypothèse, la plupart d'entre nous avons tendance à retenir les informations qui vont dans notre sens plutôt que celles qui la contredisent. Par conséquent, le processus de recherche d'information doit prendre en considération cette hypothèse émise par le décideur.

#### **1.2.2.2.1.5. Les facteurs liés à l'environnement**

Nous avons remarqué qu'en terme de gestion stratégique, l'environnement a connu une analyse en terme d'opportunités et de menaces. Cette méthode d'analyse revient à (Porter, 1986) qui considère que la stratégie de l'entreprise peut être conçue en se fondant sur la réaction de celle-ci aux forces concurrentielles qui constituent son environnement. En fait, c'est à ce courant que se rattache la notion de veille. Aux États-Unis le courant de « *l'environmental scanning* » a été initié par (Ansoff, 1965), qui a voulu adapter la stratégie à un environnement complexe et agité. Il considère comme possible la surveillance des signaux faibles et la gestion de la surprise. En nous fondant sur le modèle de (Marinet, 1989), lui même inspiré du modèle de Porter, nous retenons ces forces concurrentielles comme des sous environnements susceptibles d'affecter la définition d'un problème décisionnel. Nous

adoptons également le point de vue de (Bourgeois, 1980), qui propose de décomposer l'environnement en deux sous catégories :

- L'environnement immédiat qui affecte l'organisation de façon directe et qui concernent les clients, les fournisseurs et les concurrents.
- L'environnement général qui regroupe l'environnement social, l'environnement économique et l'environnement politique.

L'importance de ce modèle d'explicitation d'un problème décisionnel (MEPD) se manifeste essentiellement dans la traduction du problème décisionnel en termes d'enjeux, tout en tenant en considération le profil du demandeur de l'information et les particularités du contexte. Il permet d'éclairer, d'expliquer et de compléter les connaissances qu'a un décideur à propos d'un problème décisionnel complexe.

Notre modèle est transformé par la suite en un formulaire de description d'un problème décisionnel et une application a été développée afin de faciliter la gestion de ces formulaires (saisie, recherche, analyse). En effet, l'exploitation des informations contenues dans le formulaire contribue d'une part au niveau du premier filtrage, en précisant le contexte du problème et d'autre part au niveau du deuxième filtrage en donnant des informations sur le décideur. En plus, nous considérons que l'analyse de la base des formulaires va nous permettre de repérer certains cas d'utilisation qui pourront être comparés avec ceux contenus dans l'ensemble de traces qui constitue la base d'expériences. Ces derniers peuvent alors être utilisés pour prédire le comportement probable du décideur.

#### **1.2.2.2. Le veilleur**

Le veilleur est le spécialiste qui est chargée de collecter, analyser et diffuser l'information en vue de rendre plus intelligible l'environnement interne et externe de l'entreprise (Kislin et David, 2003, p. 4).

Nous allons dans cette partie, nous inspirer du modèle du veilleur présenté dans (Kislin, 2002). Le décideur, acteur social, fortement ancré dans l'environnement

stratégique, saisit l'information, la traite et agit en vue de résoudre des problèmes décisionnels à la fois interdépendants et intertemporels. Le décideur est auto-organisateur et équilibré, c'est-à-dire qu'il peut atteindre le même objectif avec des moyens différents. Décider signifie agir et la décision est l'aboutissant d'une action qui possède une finalité, une planification et une évaluation.

Les principales fonctions du décideur et du veilleur sont ainsi résumées dans le tableau et suivants :

<b>Les fonctions du décideur.</b>	<b>Les fonctions du veilleur.</b>
[Réagir & Décider]	[(re)Connaître & Veiller]
<ul style="list-style-type: none"> <li>• Identifier les stratégies qui permettraient d'améliorer les performances de l'entreprise.</li> <li>• Opérer parmi ces stratégies des choix en fonction des caractéristiques du secteur d'activité, des concurrents et des partenaires.</li> <li>• Allouer des ressources financières et humaines aux actions spécifiques engendrées par la prise de décision.</li> <li>• Mesurer et assumer les conséquences liées à la prise de décision</li> </ul>	<p>Collecter, analyser et diffuser l'information pour rendre plus intelligible l'environnement :</p> <ul style="list-style-type: none"> <li>• Définir les indicateurs, les objectifs de recherche.</li> <li>• Sélectionner les sources d'informations pertinentes</li> <li>• Suivre les évolutions des flux informationnels</li> <li>• Coordonner la DSI (Diffusion Sélective de l'Information)</li> <li>• Gérer et alimenter le SBC (Système à Base de Connaissance) de l'entreprise.</li> </ul> <p>Métafonction : Traduire le problème décisionnel en problème de recherche d'information</p>
<b>Ces paramètres définissent l'espace du problème décisionnel mesuré en terme d'enjeux.</b>	<b>Ces paramètres définissent l'espace du problème de recherche d'information mesuré en terme de savoirs</b>

Tableau 1.4 : Analogies autour des actions du décideur et veilleur (Kislin et al, 2003)

Si nous mettons en regard les activités du décideur et du veilleur, nous voyons des similitudes autour des actions comme identifier (les signaux pour le décideur, les indicateurs pour le veilleur), sélectionner (les stratégies, les sources), évaluer (les risques, les résultats) et gérer (les coûts, les systèmes d'information). Ces actions se fédèrent autour de deux dynamiques :

- Pour le veilleur de connaître les informations issues de l'environnement mais aussi de connaître les enjeux du décideur,
- Pour le décideur de réagir suite aux signaux détectés mais aussi de réagir suite aux informations restituées par le veilleur.

Deux systèmes agissent comme des filtres à la traduction du problème décisionnel et à la mise en relation des deux environnements.

Le premier est le système de préférences du décideur. Il est défini en fonction de la nature du problème décisionnel, de la qualité des évaluations, des contraintes (Darses, 1994), du degré de liberté, du temps, des croyances ou du style cognitif du décideur. Ces paramètres sont intégrés dans le modèle du décideur sous l'intitulé : (CI) caractéristiques individuelles par Bouaka et David, (Bouaka et al, 2003)

Le second est le système de pertinence du veilleur. La recherche des informations nécessaires à la résolution du problème informationnel doit pouvoir s'évaluer en quantité et en qualité. La qualité du traitement de l'information collectée dépend des outils et méthodes utilisées et s'évalue en terme de pertinence : fiabilité des sources, recoupements et hiérarchisation des informations, respect des délais impartis, lisibilité du document produit par le veilleur. L'information collectée est fiable et pertinente pour le veilleur mais l'est-elle également pour le décideur ? Une analyse rétrospective permet d'évaluer quelles ont été les influences de l'information transmise et d'en mesurer les impacts sur les conséquences.

### **1.2.3. Le processus de l'IE**

Différentes écoles de pensée regardent les processus dans l'intelligence économique suivant des perspectives différentes. Il y a ceux qui les regardent comme composés de deux étapes notamment : la collecte d'information et la prise de décision. Les autres les considèrent en détails. Dans le travail de (Dou et al, 1987), il propose de « *analyser la littérature pour prendre des décisions adaptées* ».

Le travail de Léveillé (Léveillé, 2000, p.50-58) distingue quatre étapes dans la veille technologique. Les quatre étapes ne sont pas tout à fait les étapes de veille mais plutôt des étapes d'intelligence économique. Les étapes proposées par (Léveillé, 2000) sont :

- l'orientation (expression des besoins),
- la Collecte (Identification des sources et organisation de la collecte),
- l'Exploitation (Prétraitement, mémorisation, vérification, analyse et synthèse) et enfin,
- la Diffusion.

Nous dirons que forcément, il y aura un décideur qui détermine l'orientation ou les besoins décisionnels. C'est la définition de problème décisionnel que le veilleur traduit comme un problème de recherche de l'information. Si c'est le cas, la première étape d'orientation appartient au décideur. Les trois étapes restantes doivent être attribuées au veilleur. Dans ce cas, les quatre étapes peuvent être liées aux étapes de processus d'intelligence économique. Typiquement, une activité de veille commence à partir d'une identification de problème décisionnel qui nécessite une solution. La solution est faite par la traduction du problème décisionnel en un problème de recherche d'information pour le problème actuel. Nous dirons que les activités d'intelligence économique sont l'activité de veille et la décision.

Les étapes retenues par l'équipe SITE dans les processus d'intelligence économique sont du nombre de sept. Ces processus ne sont pas indépendants ni exclusifs. Nous croyons qu'ils sont des étapes qui aideront dans la résolution d'un problème décisionnel en contexte de l'intelligence économique. Les sept étapes sont présentées dans (Knauf et David, 2004) et reproduis dans la *Tableau 1.5* .

Phases	Rôle	Compétence	Acteur
1	Définition du problème décisionnel	Il connaît l'environnement dans lequel il travaille – Maîtrise son secteur - Détecte les risques et menaces pour son entreprise ou son service. <b>Connaît les enjeux</b>	Décideur
2	Transformation du problème décisionnel en un problème de recherche d'information	Stratégie méthodologique - Mise en place d'indicateurs / Tache analytique (compréhension de la demande, de l'enjeu et du contexte) / Tache opérationnelle / Tache méthodologique	Veilleur
3	Identification des sources pertinentes	Identifie et évalue les sources d'information et en assure le suivi – Sélectionne les sources les plus adaptées à la demande	Veilleur
4	Recherche des informations pertinentes	Extrait ponctuellement ou périodiquement les informations – Contrôle les informations collectées	Veilleur
5	Traitement - analyse des informations	Traitement intellectuel des informations : analyse du contenu – Mise en forme plus élaborée des informations sous forme de synthèse, bibliographie, panorama de presse...	Veilleur
6	Interprétation	Relative à la description du problème initial	Décideur
7	Décision	Elle doit être basée sur les indicateurs mis en place avec le veilleur	Décideur

Tableau 1.5 : Table de processus de l'IE selon équipe SITE

Nous commenterons chacune de ces étapes, et conclurons avec une note sur la façon dont le processus d'annotation peut s'appliquer à certains d'entre elles.

### 1.2.3.1. Définition du problème décisionnel

La définition d'un problème peut avoir une signification différente selon le contexte. Presque chaque domaine a sa propre « *définition de problème* ». Le concept de base est la question de la transformation du problème donné en problème *scientifique*, problème *social*, problème *psychologique*,... *religieux*, ...*historique*, ...*médical*, etc... que nous pourrions étudier et analyser. La définition du problème peut être considérée comme la tâche principale d'un projet ou de la résolution d'un problème décisionnel.

Dans l'IE, notre vue est tout à fait différente. Nous sommes concernés par la « *définition de problème* » comme un processus spécifique dans les processus d'IE.

Plusieurs travaux liés à la définition du problème décisionnel ont été effectués dans l'équipe de recherche SITE (Goria, 2006) (Bouaka, 2004).

Le travail de Bouaka est lié à l'identification du problème décisionnel en termes d'objet, de signal et d'hypothèse. Goria se fonde sur l'interprétation de problèmes pour assurer la compréhension des questions posées.

La base de la définition du problème est le fait qu'un décideur décide qu'il y a un problème. Ce qui peut être problème pour un décideur peut ne pas être problème pour un autre décideur.

### **1.2.3.2. Transformation du problème décisionnel en problème de recherche d'information**

La traduction (Transformation) d'un problème décisionnel en un problème de recherche d'information exige la coopération d'un veilleur et d'un décideur. Selon (Kislin et al, 2003) la définition d'un problème décisionnel se traduit en termes d'enjeu, et se projette en demande du veilleur dans le problème de recherche d'informations stratégiques, dans le but de résoudre un problème décisionnel. Pour établir cette transformation, le veilleur doit accomplir les tâches suivantes :

- Une tâche analytique, qui correspond à la compréhension de la 'demande-enjeu-contexte' et exige de définir les indicateurs qui lui permettront de traduire l'enjeu du problème décisionnel en paramètres de recherche d'information. Cette tâche nécessite d'avoir les compétences qui permettent de comprendre l'enjeu et l'objectif, et de pouvoir y "brancher" les paramètres associés dans le monde de l'information.
- Une tâche méthodologique, qui résume les critères de compétences de traduction du problème décisionnel en projet de recherche d'informations (PRI), puis en problème(s) informationnel(s) (PI). En se servant des paramètres identifiés, le veilleur doit employer ses compétences en méthodologie de recherche d'information (et éventuellement du domaine d'étude) pour la collecte des informations pertinentes.

Il faut donc aussi que le veilleur ait des moyens par lesquels il pourra identifier l'information demandée.

- Enfin, une tâche opérationnelle, par laquelle le veilleur doit s'adapter au mieux, et à l'enjeu et au décideur, pour déterminer la forme de présentation des informations. On peut rappeler ici qu'il lui faut connaître les caractéristiques individuelles du décideur.

C'est une prérogative du processus d'intelligence économique de pouvoir énoncer sans ambiguïté un problème décisionnel car cela facilite sa transformation en problème de recherche d'information. Ceci entend, bien sûr, qu'aucun des termes utilisés dans la description de problème ne soit ambigu.

Dans cette étape de transformation de problème, plus que le problème d'ambiguïté, le problème de spécificité est pris en compte. Gorla (Gorla, 2006, p.119-122) a proposé une manière de présenter une requête de façon à obtenir l'information pertinente. Pour nous, « Je cherche information sur la voiture Peugeot 207, fabriquée en 2006 » n'est pas la même requête que « Je cherche information sur une voiture Peugeot 207, fabriquée en 2006 ». Nous préférons une requête comme « Je cherche une information sur la voiture Peugeot 207 rouge, fabriquée à Mulhouse en 2006 ». Ce sont les requêtes spécifiques qui peuvent avoir des solutions spécifiques.

Plus que l'ambiguïté et la spécificité se pose le problème de l'*objectif*. L'objectif de la recherche d'information doit être spécifié. Par exemple, une requête comme « J'aimerais trouver une université technique pour mon fils à Paris » est admissible. L'objectif de « recherche sur une université » est "pour mon fils ". Parce que, nous pouvons avoir les universités techniques à Paris "pour les filles."

Quand sont clarifiés les problèmes d'ambiguïté, de spécificité et d'objectif, nous pouvons commencer l'étape de recherche d'information.

### 1.2.3.3. Collecte d'informations pertinentes

Deux méthodes de collecte de l'information ont été suggérées par J. TREURNIET (Treurniet, 2004, pages 2,3) : la collection passive et la collection active.

Nous ne nous attarderons pas sur la façon par laquelle l'information active est collectée. En revanche, les considérations plus générales impliquées dans la collecte active d'informations sont des sujets d'intérêt pour nous. A savoir sont :

**L'éthique** : Quels sont les principes de base que nous devons respecter dans la collecte de l'information ? Bien sûr, les règles impliquant les droits personnels doivent être respectées. Il faut noter que dans la collecte d'informations inter-entreprises ou la collecte d'informations internationales, plusieurs facteurs ne sont pas définis.

**La Source** : La source des informations à collecter concerne en fait la façon dont elle est collectée. L'information issue d'Internet ne sera pas collectée de la même manière que celle qui provient d'une source orale.

**Méthodologies** : Plusieurs méthodologies sont disponibles pour la collecte d'informations. La méthode adoptée par un individu dépend de plusieurs facteurs : son "emplacement" social, les circonstances, le hasard etc. Certaines des méthodes incluent des questionnaires, des observations et des prélèvements (questionnaires, observation, matériel...).

Comme il est dit plus haut, nous ne nous attarderons pas sur ces méthodes.

**Le type d'information à collecter** : Selon le type d'information à collecter la tâche demandera plus ou moins de précautions et la récolte en sera plus ou moins facile, rendant plus aléatoires les réponses à "*comment ? où ? et quand ?*". L'information légale ne sera pas collectée de la même manière que l'information sur des enfants ou l'information à caractère commercial.

**Format** : Le format d'un document est étroitement lié avec le contenu d'information. Du format de la source d'information dépendra sa collecte. Les

informations de type *texte* n'ont évidemment pas les mêmes caractéristiques que les documents graphiques.

**Utilisation** : De l'utilisation prévue d'une information dépendront les paramètres de la collecte. Ceux-ci doivent être déterminés en regard de l'utilisation proposée ou souhaitée : l'information sur le réseau d'autoroutes sert assez rarement à l'interprétation géologique de la même autoroute ; sauf, bien sûr, si des facteurs géologiques étaient inclus dans la collecte.

#### **1.2.3.4. Validation des informations**

La question de la validation de l'information soulève plusieurs questions.

L'origine des informations est le principal sujet de prudence, en particulier dans cette ère -et dans la sphère- de l'Internet. Plusieurs sites Internet sont si dynamiques que l'information change constamment. Quelques autres sont obsolètes. D'autres existent par intermittence, souvent pour des raisons techniques. Comment valider l'information provenant de l'Internet ?

Un autre problème concernant la validation des informations est la question de la cohérence. La cohérence dans la perspective de la source et la cohérence du côté de l'utilisateur. Est-ce que les informations collectées resteront les mêmes après une première collecte ? L'auteur des informations changera-t-il complètement ou modifiera-t-il ses informations ?

Les sources de l'information doivent être validées.

Mais le transport des informations, c'est à dire leur transmission jusqu'au décideur, doit aussi faire l'objet d'une grande rigueur, et contrôlée avec prudence.

#### **1.2.3.5. Traitement des informations**

Nous n'exposerons pas ici comment l'information est traitée. La méthodologie et la théorie de traitement de l'information sont un champ qui est lié avec tant de disciplines. Il implique des métiers comme les statistiques, les mathématiques, la géométrie, le calcul, l'analyse quantitative, l'algèbre vectorielle et l'analyse

complexe. Il implique la chimie, la physique, la géographie, la géologie, l'informatique et les autres sciences discrètes. Il implique également les sciences sociales et celles des arts littéraires. Chaque discipline a sa propre théorie et méthodologie pour le traitement de l'information.

Le traitement de l'information dépend de plusieurs facteurs et il est difficile d'annoncer le traitement d'une information quand le type d'information est inconnu. Par exemple, une information multimédia ne sera pas traitée de la même manière que des textes. Quand le type de l'information est connu, les utilisations de l'information collectée sont un meilleur guide pour la façon dont il sera traité.

Les outils disponibles pour le traitement d'information ont une importance considérable, et d'abord, aujourd'hui, les outils informatiques et leurs logiciels ; mais aussi les dispositifs de stockage, et les outils de présentation : moniteurs, écrans etc...

La connaissance et les expériences du spécialiste dans le domaine de l'information sont un facteur qui influencera le traitement d'information.

La volonté du décideur influencera également le traitement de l'information.

## **Calcul des indicateurs**

Les indicateurs liés à la recherche de l'information sont calculés sur les résultats obtenus. C'est l'utilisation indiquée qui déterminera si le but de la recherche de l'information a été réalisé ou pas.

Dans les processus de l'IE, les buts doivent être énoncés avant la recherche de l'information.

On suppose que les objectifs désignés au début d'une recherche d'informations ne changent pas en cours du processus. Selon le modèle MEPD, un problème décisionnel est énoncé comme un objet, qui émettra un signal ; le problème identifié par le signal est lié à une ou plusieurs hypothèses. Les hypothèses doivent nous

guider dans les spécifications des indicateurs. Une hypothèse peut donner naissance à un ou plusieurs indicateurs.

Prenons un exemple : nous voulons découvrir pourquoi il y a une baisse dans la vente d'un produit. Pour résoudre ce problème, les ventes d'un produit peuvent être considérées comme un objet tandis que la chute des ventes du produit est le signal d'intérêt (le problème). Une des hypothèses que l'on peut déduire est que le signal (chute dans les ventes) est lié aux produits alternatifs disponibles sur le marché.

Les informations seront cherchées, collectées, validées et traitées sur la base des sept étapes de d'équipe SITE.

### **1.2.3.6. Interprétation des indicateurs**

L'interprétation des indicateurs est exclusivement une fonction de ce qui a été indiqué comme les indicateurs, les paramètres standards dans le domaine de la recherche d'information (gestion, chimique etc.) et le but de la recherche de l'information.

Plusieurs facteurs influenceront l'interprétation des indicateurs. Certains de ces facteurs sont liés à la connaissance du veilleur en matière de traitement de l'information, tandis que les autres facteurs dépendront des décideurs. Les outils utilisés pour le traitement de l'information influenceront aussi sur l'interprétation.

### **1.2.3.7. Décision**

La décision peut être vue dans des perspectives différentes selon (Thagard, 2001). Nous pouvons le voir soumise à l'intuition, ou au calcul et à la cohérence.

#### *Une décision appuyée sur l'intuition*

La décision prise est basée sur le sentiment et les convictions personnelles.

Supposons qu'un étudiant ait le choix d'étudier :

- soit l'histoire de l'art, pour lequel il a un intérêt affirmé, mais qui ne lui assure pas un avenir intéressant.
- soit l'informatique, qui peut le mener à une carrière plus lucrative.

Il peut céder à la raison, calculer, ou bien prendre une décision intuitive, qu'il justifiera par ses réactions émotives. Ainsi les décideurs intuitifs choisissent une option fondée sur ce que leurs réactions émotives leur indiquent comme préférable.

Certes une option peut sembler attrayante en raison du manque de considération pour les autres options disponibles, mais la notion d'intuition laisse une grande part à un brouillard intellectuel, probablement causé par - ou à la faveur d'- une information imprécise ou non pertinente.

Il va de soi que, dans une entreprise, une prise de décision fondée sur l'intuition peut avoir quelques sérieux inconvénients. Le raisonnement intuitif est problématique dans des situations de groupe, où les décisions doivent être prises collectivement. Si d'autres ne sont pas d'accord avec le choix d'intuition, on ne peut pas simplement affirmer la décision d'intuition. Un consensus collectif est exigé à travers une approche plus analytique que l'expression d'intuitions.

### *Une décision appuyée sur un calcul*

La décision prise est fondée sur des paramètres vérifiables. Les experts prennent une décision fondée sur une approche plus systématique et plus calculée. Des publications décisionnelles, comme Bazerman (Bazerman, 1994, p. 4), posent que la prise de décision raisonnable devrait inclure les six étapes suivantes :

- Définition du problème, avec les caractéristiques de l'usage universel de la décision.
- Identification des critères, spécification des buts ou des objectifs à accomplir.
- Evaluation des critères, associée à l'importance relative des buts.

- Production des solutions de rechange, identification des lignes qui pourraient accomplir les divers buts.
- Évaluation de chaque alternative, sur chaque critère, évaluant le point auquel chaque action accomplirait chaque but.
- Calculer la décision optimale

### *Une décision cohérente*

Quand nous avons à discerner le sens d'un texte, d'une image, d'un événement, ou d'une expérience avec une personne, nous construisons une interprétation qui est en accord avec les meilleures informations disponibles. La meilleure perception des objets, des événements, des personnes, est celle qui constitue le lien le plus logique avec ce que nous voulons comprendre, accordant deux informations conformes l'une à l'autre et refusant celles qui ne le sont.

Il y a quelques similitudes entre les processus d'IE et quelques autres formes d'intelligence. Par exemple, dans le cas d'espionnage, le travail d'intelligence est effectué au détriment -et contre la volonté- d'un groupe ou d'une cible individuelle. Nous nous proposons de discuter maintenant ce genre d'intelligence bien particulière, en comparant certaines de ses caractéristiques à celles de l'IE.

#### **1.2.4. L'intelligence militaire C<sup>4</sup>ISTAR et l'IE**

Le C<sup>4</sup>ISTAR (Command and Control, Communications, Computing, Intelligence, Surveillance and Reconnaissance : *commandement, contrôle, communications, informatique, renseignement, surveillance et reconnaissance*) comprend la doctrine, les concepts, la connectivité, les systèmes d'information, les capteurs et les outils qui sont nécessaires pour appuyer efficacement le commandement dans l'ensemble des opérations des forces militaires, grâce à la transmission rapide d'informations fiables et pertinentes. Il comprend également les systèmes d'information, la technologie des satellites et des capteurs en plus d'autres outils, compétences et processus venant appuyer la collecte de l'information.

C<sup>4</sup>ISTAR est fondé sur la guerre basée sur un réseau-centrale (**Network-centric warfare : NCW**), ou les opérations sur un réseau-centrale (**Network-centric operations : NCO**)<sup>12</sup>. C'est une nouvelle doctrine militaire initiée par les Etats-Unis, préconisant l'exploitation des opportunités techniques dans le domaine technologie de l'information et de la télécommunication pour améliorer la conscience situationnelle et la vitesse de la prise de décision (Boutelle, 2003).

L'objectif de NCW/NCO est de fournir un accès à une information sans faille à tous les militaires et décideurs, à chaque échelon dans la hiérarchie militaire. Le but est d'assurer une connectivité omniprésente dans toute l'organisation, les centres de commande, pour chaque soldat d'infanterie isolé, chaque véhicule sur le terrain, tous les avions et véhicules navals et bien sûr les vaisseaux spatiaux.

Cela permettrait à tous les éléments de partager l'information, rassemblée pour être combinée dans une image logique et précise du champ de bataille, rendue disponible à toutes les unités. Chaque unité "verrait" la somme de ce que toutes les autres unités « voient », ayant ainsi une conscience considérablement accrue de la situation.

On suppose que le fait d'avoir un accès rapide à l'information, donc la connaissance de la situation, grâce à une image opérationnelle commune, aura comme conséquence une planification stratégique plus rapide et l'exécution de décisions tactiques plus efficaces.

Les huit étapes dans le processus de C<sup>4</sup>ISTAR (Boutelle, 2004) est presque identique aux processus d'IE. Nous les comparons dans le tableau suivant :

---

<sup>12</sup> Note : Les autres formes de guerre incluse asymétrique, usure, conventionnel, fortification, la terre, guérillero, main à la main, invasion, manœuvre, naval, siège, total, fossé, non conventionnel

Processus d'intelligence Militaire C <sup>4</sup> ISTAR	Processus d'intelligence économique
Command	Définition de problème, Décision
Control	Identification des sources pertinentes
Communications	Interprétation des indicateurs
Computers	Traitement des informations
Intelligence	Calcul des indicateurs
Surveillance	Collecte d'informations pertinentes, Validation des informations
Target Acquisition	Présentation des résultats
Reconnaissance	Interprétation des indicateurs

*Tableau 1.6 : Table de comparaison entre l'IE et l'intelligence militaire*

### **1.2.5. Processus décisionnels**

Les décideurs, qui sont habitués à de nombreuses expériences, doivent prendre des décisions et agir dans le monde qui change rapidement. Dans cet environnement turbulent, la capacité d'évaluer la situation actuelle par le point de vue du "bon jugement" avec succès est affaiblie avec le bruit externe (une multitude d'émetteurs d'informations sur des matières multiples) et en changeant des paradigmes de la façon dont nous pensons aux issues sociales, culturelles, d'organisation et économiques, créant le bruit interne dans nos modèles mentaux régnants. Ces bruits biaisent notre perception de ce qui se produit vraiment dans le monde. En plus, en face de ce flux constant, les décideurs sont invités à choisir le chemin d'avenir aussi bien que d'expliquer exactement comment ils projettent d'y arriver. Avant qu'ils prennent leur décision, les décideurs commencent en développant et en évaluant des hypothèses au sujet des scénarios possibles, et puis éliminent de nombreuses possibilités jusqu'à ce qu'un petit ensemble de choix viables reste. Une fois que la décision d'agir est prise, la communication des nouvelles initiatives recommence.

Les résultats de ces initiatives produisent habituellement un certain comportement prévu, mais presque toujours, les actions produisent un comportement inattendu, de sorte que cela change de nouveau les situations présentes.

Les étapes suivantes ont été identifiées comme les étapes impliquées dans le processus décisionnel :

- i. l'identification des problèmes dans le monde changeant,
- ii. la collecte d'information,
- iii. le développement et la réflexion sur des solutions de rechange,
- iv. le choix de la meilleure alternative,
- v. la mise en application de la solution par la communication,
- vi. la surveillance de la solution.

Les étapes sont reliées à ce que nous savons du processus en Intelligence Economique. Trois points sont communs et des étapes vers la prise de décision.

- i. le problème doit être identifié,
- ii. les informations doivent être collectées pour résoudre le problème actuel et
- iii. l'évaluation des informations doit se faire avant et après son application.

### **1.2.6. Les systèmes informatisés liés aux processus de décision**

Un système automatisé peut seulement aider à la prise de décision, il ne peut pas remplacer l'homme dans la prise de décision. Plusieurs efforts ont été faits pour utiliser l'ordinateur dans les prises de décision. L'intelligence artificielle a été conçue au commencement pour exécuter les processus similaires aux raisonnements humains. L'idée est de remplacer l'homme par des machines.

Nous essaierons de discuter certains de ces systèmes. Le but de cette informatisation est d'aider l'homme pour prendre une décision rapide, qui ne compromette pas la qualité de sa prestation.

Dans un système d'aide à la décision, un utilisateur compose une requête mentalement, la requête est transformée dans un format qui est en conformité avec les règles et les attributs dans le système informatisé. La requête est renvoyée à une base de données de connaissance ou à une base de données de faits. Des bases de données de connaissance sont créées, fondées sur l'information disponible dans le domaine, l'expérience et la connaissance des experts en matière de ce domaine d'intérêt. La requête est mise en correspondance avec la base de données pour produire les sélections qui s'accordent avec la requête. Cet aspect compare le contenu de la requête aux attributs et aux valeurs dans une base de données. Dans un système où la connaissance est fondée sur des faits ; la requête est guidée par les options choisies par l'utilisateur. C'est à dire que le système propose les options et l'utilisateur choisit une option qui est la meilleure selon son problème.

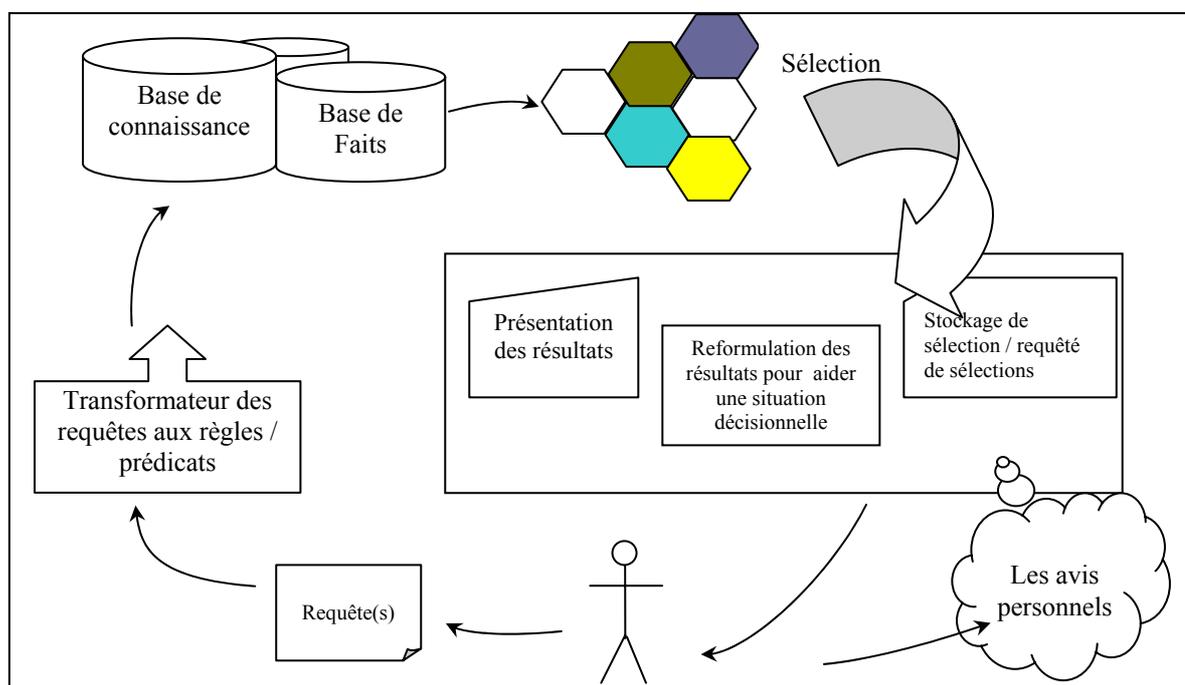


Figure 1.3 : Les étapes de recherche d'information

Quatre méthodes sont généralement utilisées pour développer les systèmes interactifs informatisés d'aide à la décision et les systèmes d'experts. Les

raisonnements à base de cas, les raisonnements à base de règle, les réseaux neuronaux et les réseaux probabilistes.

La requête est reformulée en fonction du résultat. Un des inconvénients principaux des systèmes interactifs informatisés est que la plupart d'entre eux n'ont pas de place pour les commentaires personnels qui peuvent être ajoutés à l'interface de présentation. Ces commentaires et les choix ainsi que la requête utilisée en faisant le choix peuvent être un outil très important pour améliorer la base de données de source.

### **1.2.6.1. Les systèmes d'aide à la décision (SIAD)**

Les systèmes interactifs d'aide à la décision (SIAD en anglais *Decision support system (DSS)*) sont une classe spécifique des systèmes d'information automatisés, qui tiennent des activités des affaires et des organisations pour la prise de décision. Ils sont caractérisés par leur interactivité. Un SAD correctement conçu est un système de logiciels interactifs prévus pour aider des décideurs à compiler les informations utiles à partir des données brutes, des documents, de la connaissance personnelle, et/ou des modèles d'affaires, pour identifier et résoudre des problèmes et pour prendre des décisions.

McNurlin (McNurlin et al, 1989) ont défini les systèmes interactifs d'aide à la décision (SIAD) en tant que « systèmes informatisés qui aident des décideurs à confronter des problèmes mal structurés, par l'interaction directe avec des modèles de données et des modèles d'analyse." Ces systèmes sont habituellement fortement quantitatifs et permettent à des utilisateurs d'explorer les conséquences de différentes décisions basées sur des données différentes. Avec le SAD les modèles divers peuvent être créés et analysés, par exemple sous forme de questions, de « S' il y a ». Les bilans fournissent une méthode simple d'un tel modèle et analyse. Le travail de Farrell (Farrell et al, 1992) montre comment les bases de données, simulation déterministe de modèle, et les systèmes basés sur les règles (expert) peuvent tout

contribuer au SAD. Les applications du système de SAD sont répandues dans le domaine de l'agriculture dans les années 1990.

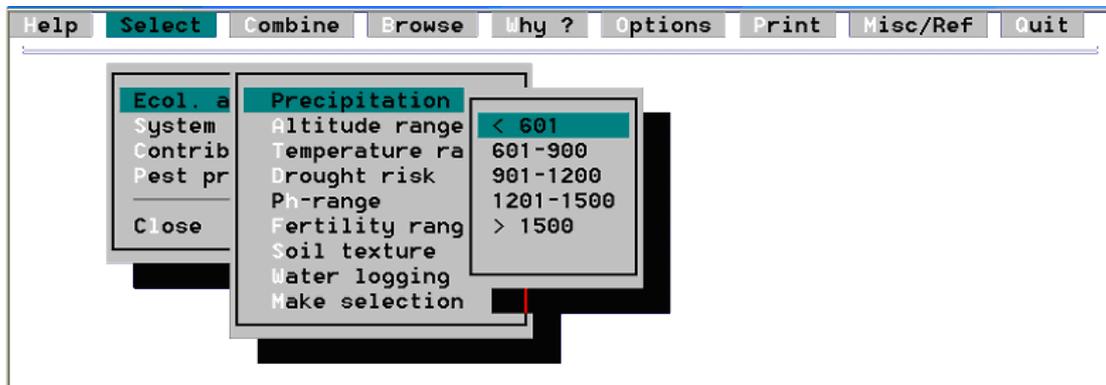


Figure 1.4 : Interface d'un system de SAD (LEXSYS 2.3)

Les SADs ont évolué tôt dans l'ère de l'informatique distribuée. L'histoire de tels systèmes commence vers 1965. Les technologies qui convergent au début du XXI<sup>ème</sup> siècle pour fournir l'appui intégré pour des directeurs qui travaillent dans la solitude, dans les équipes et dans des organisations hiérarchisées, pour contrôler des organismes et pour prendre des décisions plus raisonnables.

### 1.2.6.2. Système d'Information Exécutifs (SIE)

*(Executive information systems EIS)*

Les systèmes d'informations exécutifs sont considérés comme un genre de SAD spécialisés pour satisfaire les besoins de l'information principalement des décideurs.

Ils sont identifiés dans les étapes suivantes par (McNurlin et Sprague 1989) :

- Mesurer la performance de facteurs critiques,
- Descririre de la clé actuelle des problèmes,
- Souligner les choses aux quelles le décideur s'intéressé le plus,
- Détailler les rapports de performance des subalternes.

Les dispositifs typiques d'un bon SIE incluent des agrégats plus larges, des données d'organisations, des rapports « de réel contre des actions prévues », des graphiques clairs, un accès rapide à l'information interne et externe et l'optimisation censée

concerner des facteurs critiques du succès des décideurs. Un des bénéfices du système de SIE sur les aspects du processus décisionnel a été remarqué dans le travail de Leidner (Leidner et al, 1993). Ils ont examiné 46 utilisateurs exécutifs d'SIE. Ils ont constaté que la fréquence et la durée d'utilisation de SIE a augmenté la vitesse d'identification de problème et la prise de décision et l'ampleur de l'analyse dans la prise de décision.

### **1.2.6.3. Les systèmes d'aide à la décision de groupe**

Les systèmes d'aide à la décision de groupe (en anglais *Group decision support systems* -GDSS) aident la prise de décision collective de groupe à travers

- la collaboration et la créativité,
- l'économie de temps,
- les solutions améliorées, et
- la mise à niveau sociale ou la création d'ambiance favorable.

Lewis et Keleman (Lewis et Keleman, 1989) ont remarqué plusieurs avantages de l'utilisation de GDSS. Les avantages comprennent un degré élevé d'engagement pour l'exécution d'un projet et une meilleure gestion des conflits. Les utilisateurs étudiés ont amélioré leurs capacités dans le domaine de résolution des problèmes de planification et observé une grande amélioration dans la productivité, bien plus que dans le fonctionnement ordinaire d'une équipe. Les améliorations sont dues à la possibilité d'exprimer des avis personnels

Un GDSS vise à améliorer le processus de la prise de décision de groupe en enlevant les barrières communes de communication, fournissant des techniques pour structurer l'analyse de décision, et diriger systématiquement le modèle, la synchronisation, ou le contenu de la discussion dans un groupe (Davison, 1995). Le mot groupe implique ici deux personnes ou plus, ainsi qu'un but commun, si ce n'est pas une perspective commune. On a conçu un certain nombre d'outils et de techniques qui peuvent être utilisés pour améliorer le processus de la prise de

décision de groupe. Un certain nombre de résultats peuvent également être mesurés afin de donner une appréciation de ce qui peut se nommer l'efficacité du GDSS.

#### **1.2.6.4. Le traitement analytique en ligne (OLAP)**

Le traitement analytique en ligne (OLAP : On-Line Analytical Processing) est une catégorie de technologie de logiciel qui permet à des analystes, à des décideurs et à des cadres de gagner une perspicacité des données par l'accès rapide, consistant et interactif à une grande variété de vues possibles, d'informations qui ont été transformées à partir des données brutes, pour refléter la vraie dimensionnalité de l'entreprise comme le comprend l'utilisateur. Dans une base d'analyse OLAP, les données sont organisées selon une structure multidimensionnelle. On parle alors de "cube décisionnel".

Ce concept est appliqué à un modèle virtuel de représentation de données appelé cube ou hypercube OLAP. Il existe ensuite plusieurs déclinaisons semblables à des pilotes qui permettent d'adapter le stockage des données sur différents types de base de données pour implémenter le concept OLAP :

- R-OLAP (Relational OLAP)
- D-OLAP (Dynamic ou Desktop OLAP)
- M-OLAP (Multidimensional OLAP)
- H-OLAP (Hybrid OLAP)
- S-OLAP (Spatial OLAP)

Globalement, pour les systèmes d'aide à la décision, il y a un lien formel entre les éléments d'information, ou entre les règles. Dans le système fondé sur l'annotation que nous avons proposé, la base de données d'annotation à créer peut être fondée non seulement sur les liens entre les données mais également sur les expériences et les suggestions du créateur des annotations. Cette situation peut être les bases de données d'annotation ayant un facteur social, puisqu'un utilisateur peut faire des

annotations plusieurs fois sur le même document dans les circonstances différentes. Cette possibilité de créer une situation faisant place à un lien entre les éléments d'annotation (dans la base de données) devient dynamique par rapport aux systèmes interactifs d'aide à la décision, qui, eux, sont statiques, figés. En plus, les entrées peuvent également refléter les changements du temps par rapport à l'environnement et aux perceptions sociales.

Les bases de données fondées sur des règles dans quelques systèmes d'aide à la décision, ou fondées sur l'organisation, comme dans d'autres systèmes analytiques sont généralement créées en premier lieu avec la technologie du jour et avec une très petite prise en compte de l'utilisateur. Les utilisateurs sont autonomes et doivent trouver leur chemin dans les données organisées. Les organisations des faits et des figures dans les systèmes d'aide à la décision sont fondées sur des concepts et non pas sur les situations réelles. Les SAD ne peuvent pas évoluer avec la tendance du temps ni refléter les changements humains.

### **1.2.7. L'intelligence et l'annotation**

L'objet annotation est fortement lié au contenu du document. C'est impossible de faire une annotation de valeur si le contenu d'un document n'est pas saisi. Essentiellement, l'annotation est une conséquence d'une ou plusieurs études intelligemment faites d'un document. Nous nous inspirons des raisons à partir desquelles les annotations sont faites, elles incluent :

- l'interprétation de l'information,
- le résumé d'un document et
- la classification d'un document etc.

Ce sont tous des processus cognitifs qui ont besoin de raisonner sur le contenu du document. Nous disons donc que l'annotation est un effort intellectuel appliqué à une source d'informations.

### **1.2.8. Les problématiques liées à la recherche et au traitement de l'information**

L'information est seulement utile si elle peut aider à la prise de décision. La théorie de la décision ne considère comme information que ce qui est de nature à entraîner ou modifier une décision. Dans le cas contraire, il s'agit d'un simple bruit. On pense souvent que l'information peut être définie comme une donnée réductrice d'incertitude.

Plusieurs problèmes doivent être surmontés pour procurer l'information qui aidera à prendre une décision. Les problèmes suivants sont liés au processus de recherche et au traitement de l'information :

- problèmes d'accès à l'information,
- problèmes de spécificité de l'information recherchée,
- problèmes de traitement de l'information,
- problèmes d'organisation,
- problèmes de formatage et présentation.

Après avoir défini les terminologies de base de notre étude, le chapitre suivant est dédié l'explication de quelques travaux majeurs faits en relation avec l'annotation. Nous discuterons aussi l'importance d'annotation dans les sciences de l'information et de la communication.

# Chapitre 2

*Pour son existence, la dépendance de l'homme aux plantes a été d'une importance capitale depuis que la race humaine existe - Victoria Abike Robert (1965 – 2005)<sup>13</sup>.*

*Pour son existence, l'annotation est indissociable du document de façon permanente  
Charles Abiodun Robert<sup>14</sup>*

## 2. État de l'art sur l'annotation

Ce chapitre sera divisé en deux parties. La première partie discutera l'importance de l'annotation dans les sciences de l'information et de la communication tandis que la deuxième partie considèrera de la création et de stockage d'annotation.

### 2.1. L'annotation dans le domaine de SIC

Nous examinerons la connexion entre l'annotation et les sciences de l'information et de la communication. L'accent sera mis sur les travaux qui ont été effectués à cet égard.

#### 2.1.1. Les sciences de l'information et de la communication

Les sciences de l'information et de la communication sont liées aux sciences de la documentation et à l'informatique. Nous pouvons dire que, bien que les sciences de l'information et de la communication soient proches et associées aux autres disciplines, les liaisons sont d'un intérêt particulier pour notre étude. Avant une présentation des définitions liées aux sciences de l'informations et de la communication, nous présenterons un "périmètre" des six domaines d'études présentés par le Conseil National des Universités (CNU<sup>15</sup> (71)) dans le rapport d'évaluation d'enseignement des sciences de l'information et de la communication du comité national d'évaluation, les sciences de l'information et de la communication (Saget, 1993, p.16). Nous présenterons les « périmètre » :

---

<sup>13</sup> *Man's dependence on plants for his existence has been of paramount importance since human race began*

<sup>14</sup> *Annotation's dependency on document for its existence is inseparable all time*

<sup>15</sup> <http://cnu71.online.fr/12-compe.html> 23/03/2006

- *les études sur les notions d'information et de communication, sur leurs **relations**, sur la nature des phénomènes ainsi désignés ainsi que les **analyses** philosophiques, épistémologiques, méthodologiques, logiques, mathématiques de ces phénomènes ;*
- *l'étude de l'information, de son **contenu**, de ses propriétés et de sa présentation ;*
- *l'étude des **systèmes d'information** et des modèles documentaires, informatiques et autres, qu'ils mettent en oeuvre ;*
- *l'étude des **médias** de la communication sous leurs divers aspects ;*
- *l'étude du **fonctionnement des processus** de communication et des productions et usages de la communication ;*
- *l'étude des **acteurs** de la communication et des agents du traitement et du transfert de l'information, de leurs formations, de leurs professions.*

Si nous devons récapituler les points ci-dessus, nous dirons que la science de l'information et de la communication est l'étude de l'information dans la perspective de (a) sa relation scientifique, (b) le contenu (c) des processus de l'information, (d) les acteurs, et (e) les médias de communication.

En France, les sciences de l'information ont une dénomination officielle, « les sciences de l'information et de la communication (SIC) ». À l'étranger, l'information et la communication sont généralement séparées.

Les sciences de la communication considèrent la communication comme « *toutes les expressions issues d'acteurs sociaux et porteuses d'une intentionnalité analysable du point de vue d'un observateur-lecteur avisé, c'est-à-dire capable d'en comprendre le sens dans un contexte pertinent pour les acteurs concernés.* » (Mucchielli et al, 1998). Dans ces définitions, les mots principaux se répètent comme (a) expression (b) acteur (c) intention analysable (d) contexte pertinent. Nous pouvons appliquer cette définition dans le domaine de l'annotation dans le contexte de la science de l'information et de la communication. Comme une annotation est forcément écrite dans un contexte de document d'origine, l'analyse d'annotation est toujours, elle aussi, dans un contexte de **contenu** de document. Une

annotation est faite par un *acteur*, qui *analyse* le *contenu* d'un document dans un *système d'information*.

Un autre travail d'importance est celui de Paoli (Paoli et al, 2003) qui montre une analyse systématique de l'information dans quelques pays choisis en utilisant un indicateur prédéterminé de manière à évaluer le rôle d'un média d'information (l'Internet).

Dans les sciences de l'information et de la communication, tout objet doit être communiqué. Quelle est la signification du mot « communication ». Plusieurs définitions du terme communication se trouvent dans le travail d'Alex Mucchielli et Jeannine Guivarch (Mucchielli et Guivarch, 1998). Dans les sections précédentes, nous avons défini le mot « information », mais l'information et la communication sont toujours inseparables. Nous définissons le mot « communication » dans le paragraphe suivant.

### **2.1.2. La communication**

Dans la publication de "National Joint Committee for the Communicative needs of Persons with Severe Disabilities (National, 1992, p. 2), la communication se définit comme suit :

*N'importe quel acte pour lequel une personne donne, ou reçoit des autres les informations d'une autre personne sur les besoins, les désirs, les perceptions, la connaissance ou les états affectifs de celle-ci. La communication peut être intentionnelle ou involontaire, elle peut impliquer les signaux conventionnels ou non, peut prendre les formes linguistiques ou non, et peut se produire par le moyen sonore ou autres.*

Malgré le fait que cette définition soit une définition spécifique, nous focalisons sur le mot central dans la définition « l'échange d'information ».

Une autre définition de la communication se présente comme « *le processus de la transmission d'information de l'un émetteur à un récepteur par le moyen de l'utilisation d'un message transmis d'un à l'autre à travers un canal*<sup>16</sup>.

Quatre éléments sont impliqués dans un processus de la communication : l'émetteur, le récepteur, le canal et le message.

Modèle	Définition de la communication	Problématique principale
<b>MODELE POSITIVISTES</b> : le contenu et ses effets		
Emetteur-récepteur	Un transfert d'information	Les propriétés du contenu
De la communication à deux niveaux	Un processus d'influence	L'influence sur les opinions
Marketing	Une « opération » à piloter	L'efficacité de l'opération
<b>MODELE SYSTEMIQUES</b> : le système comme cadre		
Sociométrique	Une relation d'affinité	La structure affective et les places occupées
Transactionnel	Des rituels d'échanges sur plusieurs niveaux	Les motivations profondes et les avantages tirés
Interactionniste et systémique	Une participation à un système d'interactions	Le système des échanges, le sens de l'échange dans le système, les paradoxes
De l'orchestre	Une production collective	L'articulation des jeux individuels
<b>MODELES CONSTRUCTIVES</b> : la genèse du sens		
De l'hypertexte	Un débat latent	Le sens du débat
Situationnel	Une construction collective du sens	Les processus de la genèse

Tableau 2.1 : La définition de la communication (Alex Mucchielli, 1998)

### 2.1.3. La communication et la recherche de l'information

A la suite des explications précédentes, nous croyons que la communication est l'échange de l'information fondé sur des besoins divers. Dans la plupart des situations, la communication peut être lancée parce quelqu'un a besoin de quelque chose. Dans le cas où la personne a besoin d'information, « *l'initiateur* » de la communication est le demandeur d'information. Dans certains cas, la communication peut être lancée en raison d'une déclaration floue qui, elle peut provoquer l'explication, la justification, la clarification ou l'exemplification (besoin d'information provoqué). Dans les deux cas, la participation des deux parties (l'initiateur et le répondant) est systématiquement réalisée. Enfin, la communication est liée au besoin.

<sup>16</sup> <http://www.mind-graph.net/foundations/linguistic/communication.htm> 31/08/06

Dans le cas de la recherche de l'information, le focus est la situation où la communication est provoquée par le chercheur d'information (la personne qui a un besoin). Le chercheur d'information imagine qu'il peut y avoir une réponse à son besoin quelque part dans un système d'information. Il est encouragé à faire une recherche pour son besoin dans ce système d'information. Il est à noter que cette situation s'applique dans les systèmes d'information informatisés ou dans la vie ordinaire. Par exemple, un étudiant recherche le vol le moins cher pour Londres. Son besoin est l'information sur le vol le moins cher pour Londres. Il peut décider d'interroger son ami (initier une communication) à ce sujet (information). On peut voir son ami comme une partie d'un système d'information global. Il peut en même temps lancer une recherche de « *vol moins cher* » sur l'internet.

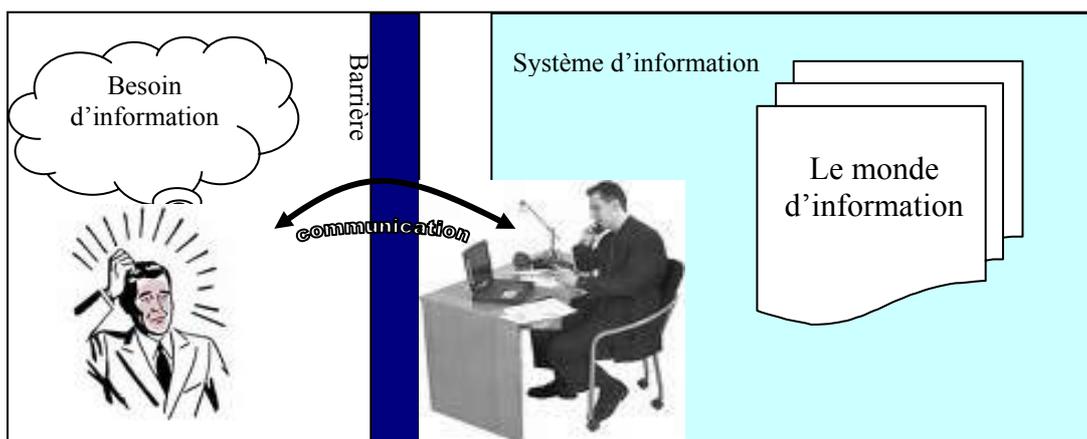


Figure 2.1 : Un chercheur d'information communique pour réduire une barrière

Dans une recherche d'information, le chercheur de l'information est séparé du monde de l'information par une barrière. La barrière peut prendre des formes diverses. Elle peut être une barrière linguistique ou une barrière de localisation (la distance), etc. Le chercheur de l'information suppose qu'il existe un monde d'information où son problème de besoin d'information peut être résolu. Il croit aussi qu'il y a quelqu'un (un expert) ou un outil qui peut réduire ou enlever la barrière entre lui et le monde de l'information. L'expert est expérimenté dans l'organisation du monde de l'information et il sait transformer les besoins du chercheur de l'information en terme du monde de l'information. Le chercheur de l'information lance une communication

avec l'expert qui est un membre du système d'information. L'objectif de la communication, c'est de permettre l'enlèvement ou la réduction de la barrière entre lui et le monde de l'information.

#### **2.1.4. La recherche d'information**

La phrase "recherche d'information" a deux connotations en anglais, « Information research » et « Information Retrieval ». En anglais, la phrase « information research or Information search » est un peu différente de celui de « information retrieval », « Information Research » est plutôt la recherche d'information alors que « information retrieval » est plutôt la récupération d'information dans un stockage informatisé et une mathématique de récupération.

La recherche d'information est un domaine très large. On peut le voir comme une activité mono-processus dans laquelle la recherche d'information pour résoudre un problème est instantanée (l'acte de recherche d'information sur place). On peut le voir comme une activité multi-processus où la recherche d'information prend des étapes différentes. Comme nous l'avons présenté dans les huit étapes de processus d'intelligence économique dévoilées dans les sections précédentes, la recherche d'information dans tous les cas est une activité multi-processus.

Dans le cas où la recherche d'information est vue comme mono-processus (la récupération de l'information), l'attention est portée sur « la requête concrète » de l'information à partir d'une base de l'information. Plusieurs efforts et les recherches scientifiques ont été effectués dans ce domaine. Certains de ces travaux sont dans le domaine du profil d'utilisateur, des travaux bibliographiques, des statistiques d'efficacité du système d'information et des recherches documentaires. Quelques modèles de recherche d'information ont été suggérés. Les recherches d'information (information retrieval) ont été autrefois liées à l'organisation de l'information.

Considérons le travail de Rijsbergen (Rijsbergen, 1979) sur « information retrieval », ou son attention (énumérée ci-dessous) sur la récupération de l'information est distinctement différente de notre conception (SITE-LORIA) de la recherche

d'information. Le « information retrieval » est dévoilée dans les perspectives suivantes :

- *Analyse automatique de texte* – étudie la façon dont le texte d'un document est représenté dans un ordinateur.
- *Classification automatique* - méthodes générales de classification automatique et évaluation de méthodes de la récupération de l'information.
- *Structures de fichier* - structures de fichier selon le point de vue d'un chercheur d'information principalement.
- *Stratégies de recherche* - plan de recherche appliquée aux collections de documents structurées dans les manières distinctes, également l'utilisation de la rétroaction.
- *Recherche probabiliste* - différents modèles formels pour augmenter l'efficacité de récupération en utilisant des informations d'échantillon sur la fréquence d'occurrence.
- *Évaluation* - une vue de la mesure de l'efficacité dans la théorie de l'évaluation.

Les propositions de « information retrieval » sont essentiellement quantitatives. Or la demande en recherche d'information ne se limite pas à des résultats quantitatifs. Néanmoins, il reste à clarifier ce que les résultats et les analyses quantitatives peuvent être.

### **2.1.5. L'annotation comme un moyen de communication**

Comme nous l'avons exprimé dans les définitions de l'annotation du contenu d'un document, une annotation peut être une trace d'usage d'un document. C'est-à-dire, il peut témoigner du passage d'un utilisateur. L'utilisateur a pour intention d'exprimer ses opinions sur un document. Les opinions sont les interprétations et les évaluations du document en question.

Un exemple : Imaginons qu'il y ait eu une annotation faite par Louis-Napoléon Bonaparte sur un document d'objet d'art en Egypte en 1849. A travers l'annotation,

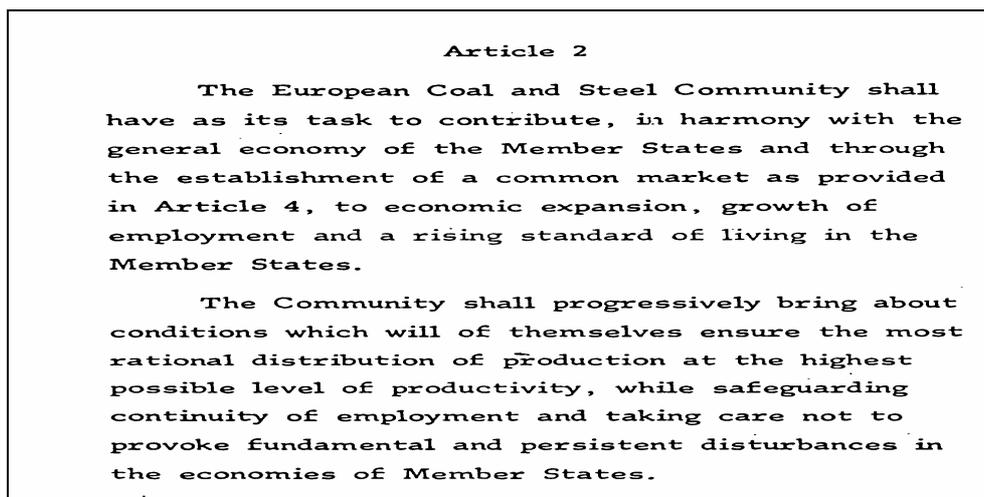
nous pouvons dire aujourd'hui que (a) Louis-Napoléon Bonaparte était directement ou indirectement impliqué dans les objets d'art Egyptiens (b) il y avait une relation forte entre les objets d'arts d'Egypte et l'homme politique Français en 1849. Dans les cas ci-dessus, c'est l'annotation sur le document qui communique ces informations.

L'annotation peut être un outil efficace dans le domaine de la communication. Le but central d'une annotation dans le domaine de communication est que les annotations partagées peuvent fournir une possibilité de communication efficace pour le public. Dans le cas où le travail implique la référence fréquente à un ensemble de documents et d'utilisateurs, tels qu'entre les professeurs et les étudiants dans un cours, l'annotation partagée peut être un outil efficace pour des usages pédagogiques.

### **2.1.5.1. L'annotation comme un moyen de communication d'un groupe**

Nous pensons que l'annotation est un processus continu. L'intérêt sur le document à un moment peut être sur l'implication économique de l'union et à un autre moment, l'intérêt peut être sur la composition de l'union. Le document principal reste le même mais l'orientation change. Dans ce cas, nous pouvons avoir un coordonnateur de l'annotation qui dirige la tendance et l'orientation est menée par l'annotation.

Quand une annotation est présentée, nous poser des questions comme : Pourquoi l'annotation a-t-elle été faite ? Quelle est la personne qui a fait l'annotation ? Quand l'annotation a-t-elle faite ? La réponse n'est pas souvent énoncée dans l'annotation. Comment résolvons-nous ce genre de situation ? Si nous pouvons régulièrement répondre à ces questions, nous pouvons faire une analyse croisée entre les annotateurs, entre le temps et les documents.



*Figure 2.2 : Un extrait du traité Européen de 1951*

Par exemple, si les annotations devaient être faites sur l'extrait de figure 2.2 (un extrait du traité européen de 1951), différents utilisateurs feront les annotations différentes sur cet extrait. Les points saillants qui attireront l'attention des lecteurs (les annotateurs potentiels) de cet extrait peuvent inclure, (a) l'union européenne, (b) le marché unifié (c) la coopération dans le domaine politique et diplomatique (d) souveraineté universel (e) l'emploi (f) la croissance économique. Il y a des utilisateurs qui peuvent être intéressés par d'autres éléments qui ne sont pas directement énoncés dans l'extrait comme (g) les personnalités impliquées dans le traité (h) les pays dans le traité et (i) pourquoi et où ce traité a été fait ?

Dans ce cas, il sera nécessaire d'avoir un coordinateur pour diriger la tendance des annotations sur ce document. Si ceci est fait, nous pouvons effectivement employer l'annotation comme un moyen de communication de groupe.

## **2.1.5.2. L'annotation et la cycle communicationnelles**

### **2.1.5.2.1. La création d'une annotation**

L'annotation est portée sur un document. Les écritures d'un auteur donnent naissance à la production d'un document. Un annotateur est principalement un lecteur. Il s'est inspiré de ce qu'il a lu dans un document publié pour faire une annotation. Son

annotation n'est pas fondée sur le monde de l'information autour de l'auteur du document de base, mais sur les sujets bien précis dans le document publié. Le lecteur du document ou l'annotateur doit avoir accès au document concerné. Il crée une ou plusieurs annotations fondées sur sa compréhension du document. La *figure 2.3*, est une illustration d'une création d'annotation

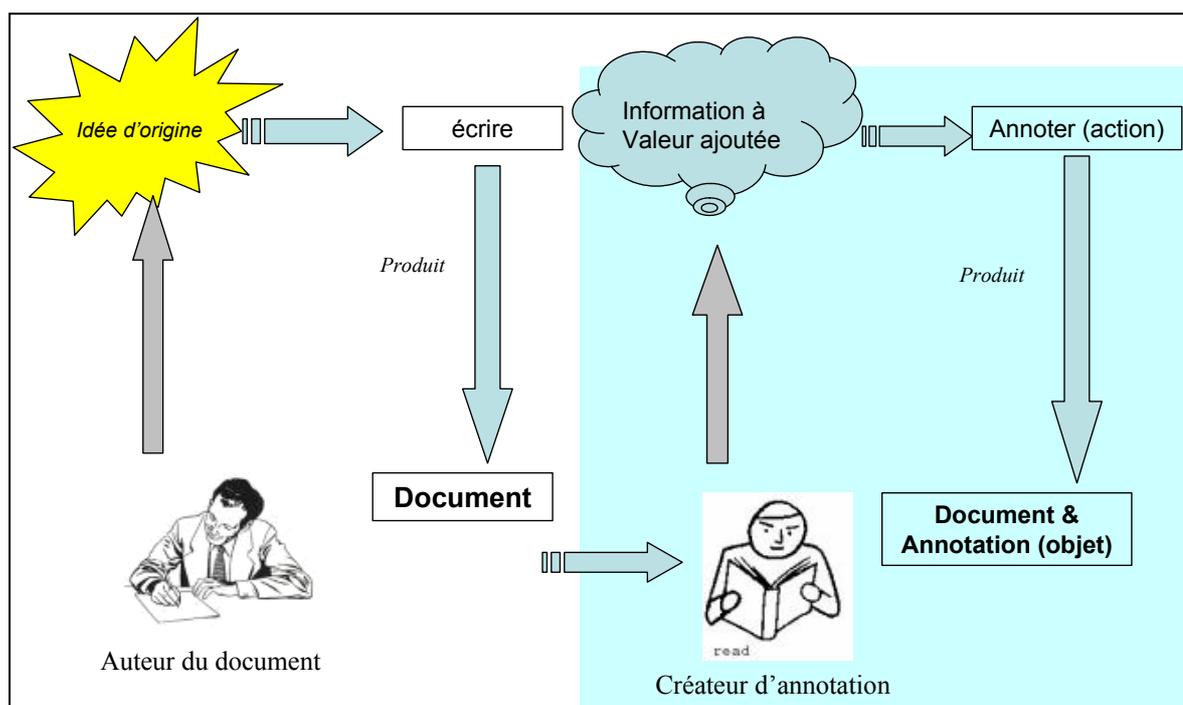


Figure 2.3 : La création d'annotation

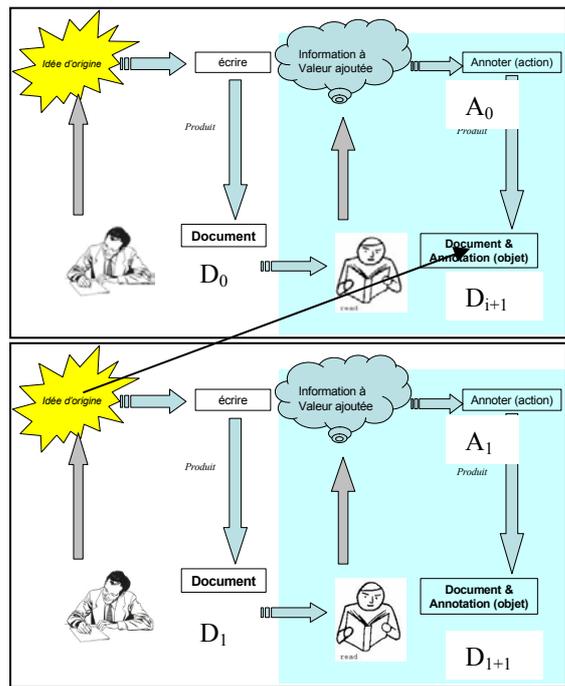
### 2.1.5.2.2. Les annotations continues

Nous pouvons noter que le processus d'annotation n'est pas un processus statique. Une annotation peut être utilisée pour effectuer une série de discussions dans une communication répétitive. Par exemple, une annotation de type « clarification » est une manière de demander d'avantage d'explications dans une communication entre l'auteur d'un document et le lecteur d'un document. L'annotation sur le document d'origine exige une restructuration du document original pour que le contenu du document puisse être mis en évidence. Le document d'origine et l'annotation faite par un lecteur sont renvoyés à l'auteur d'origine par le créateur d'annotation (lecteur). L'auteur du document d'origine devient subséquemment le lecteur de

l'annotation avec le document d'origine. Dans ce cas nous entrons dans ce que nous pouvons appeler « **les annotations continues** ».

Dans un système où le cycle d'annotation est continu, il y a plusieurs repères d'intérêt :

- L'auteur du document d'origine et l'auteur de l'annotation initial peuvent être vus en tant qu'auteurs des annotations successives,
- Le cycle d'annotation peut être infini,
- Le cycle d'annotation peut fournir un forum pour la direction d'une compréhension du document d'origine,
- La compréhension du document peut être demandée par un des participants à l'échange,
- La compréhension du document peut faire référence aux autres documents hors du contexte de la discussion,
- Le cycle des annotations peut être une méthode de communication entre deux ou plusieurs individus,
- Il y a un historique des annotations,
- L'interprétation du document évolue avec le temps.



$D_0$  = Document d'origine  
 $D_1$  = Nouvelle version du document d'origine  
 $A_t$  = Annotation à l'instant t  
 $D_{t-1}$  = Document précédent  
 $D_1 = A_0 + D_0$   
 $D_i = D_{i-1} + A_i$   
 $D_i \neq D_{i+1}$

Figure 2.4 : L'annotation continue

- L'agrégation de la connaissance : puisque plusieurs annotateurs sont impliqués dans le processus d'annotation continue, dans les travaux collaboratif et dans les classifications ou évaluations de la compréhension d'un document, l'agrégation des connaissances des annotateurs est possible.

### 2.1.5.3. Le cycle informationnel et l'agrégation d'expérience

Nous devons souligner que les annotations sont souvent pour être partagées. Le partage peut faire référence au temps ou aux lecteurs. Quand un lecteur crée une annotation pour un usage, l'usage est pour l'avenir, soit pour un usage personnel, soit pour être partagé avec d'autres lecteurs. Nous supposons que nous pouvons avoir les annotations sur les annotations déjà existantes qui pourraient exprimer d'autres concepts que les concepts du document d'origine. Dans ce cas, les nouveaux concepts introduits par l'annotation refléteront les points de vue des annotateurs. Nous nous sommes intéressés au partage de l'annotation à travers un groupe de personnes. Dans ce cas, nous formulons les hypothèses suivantes :

- Nous avons au moins deux participants qui sont en collaboration par le partage d'information,
- Nous avons au moins un cycle d'échange d'annotation entre les participants,
- Il existe une utilisation distribuée et collaborative des annotations,
- Les annotations sont faites pendant une certaine périodes,
- Le moyen d'échange des annotations est de préférence électronique,
- Les annotations reflètent la compréhension des concepts exprimés dans le document.

Dans la section suivante, nous explicitons comment l'annotation peut être partagée par un groupe de personnes.

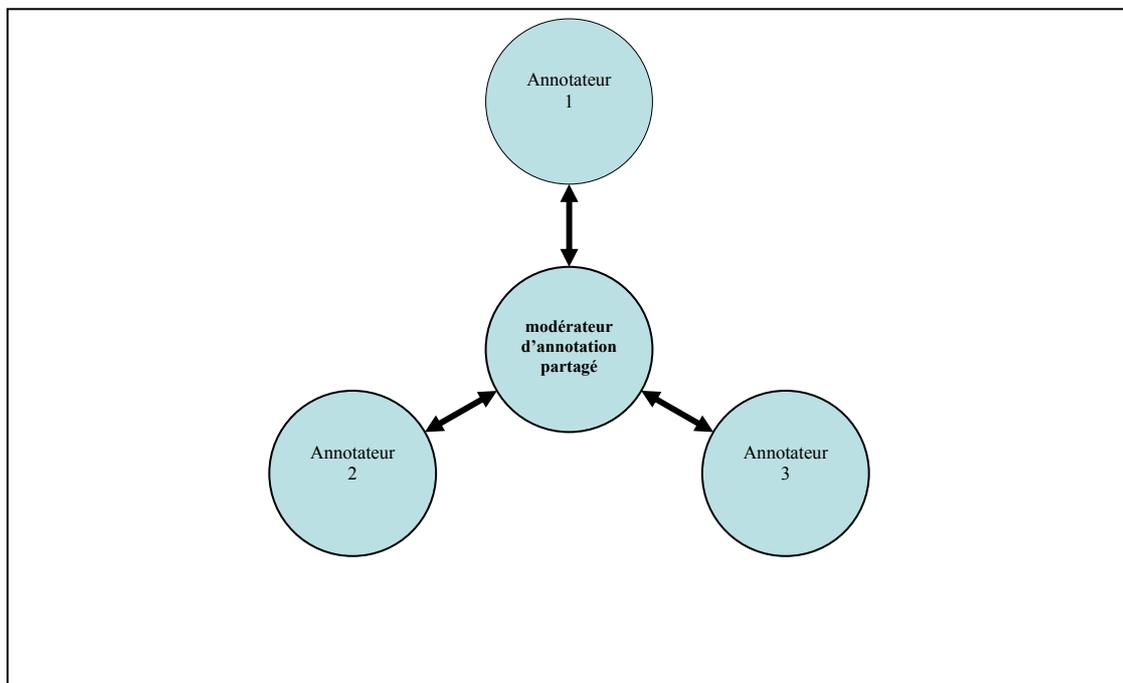
#### **2.1.5.3.1. Le partage des expériences à partir d'un seul document**

Nous considérons d'abord le cas où un seul document annoté est la base d'agrégation de connaissances. Différentes méthodes peuvent être utilisées pour partager des annotations sur un seul document. La méthode utilisée pour partager l'annotation dépend des acteurs impliqués dans le partage et l'objectif de ce partage. Nous discuterons quatre méthodes de partage des annotations ici :

##### ***Partage en mode modéré***

Dans ce cas, une annotation est partagée de manière modérée. Il y a un modérateur ou un responsable qui est chargé de gérer des annotations effectuées. Tous les annotateurs sont reliés à un unique annotateur central, c'est à dire le modérateur ou coordinateur. Un acteur émet ses suppositions à l'annotateur central (coordinateur), ce dernier envoie l'information ajoutée (annotation) avec le document à tous les autres annotateurs l'un après l'autre ou uniquement au destinataire concerné. Le modérateur a plus de contrôle sur les annotations partagées par chaque annotateur que les autres annotateurs. Puisque qu'il y a un modérateur au centre des annotations partagées, les annotateurs autour du modérateur central ne doivent pas partager les annotations entre eux-mêmes.

Cette méthode de partage des annotations peut être une méthode d'importance dans le processus d'intelligence économique où un décideur peut avoir une activité centrale sur les activités des autres. L'utilisation d'annotation partagée peut être efficace pour la gestion des activités des autres. Par exemple, un décideur peut lancer une discussion dans une entreprise. Il peut vouloir mettre en application cette discussion avec l'utilisation des propositions d'annotation. Il décide alors d'envoyer un document (ou le document annoté selon les circonstances), à un agent dans l'entreprise. On s'attend à ce qu'il évalue les propositions faites par celui-ci avant qu'il envoie le document annoté aux autres agents dans l'entreprise.



*Figure 2.5 : Partage d'annotation en mode modéré*

### ***Partage en mode symbiotique***

Des annotations sont partagées fondées sur des expériences diverses des annotateurs dans l'espace collaboratif. Aucun modérateur n'est considéré comme le coordinateur de l'annotation dans ce cas. Chaque annotateur a une contribution unique à faire dans la collaboration.

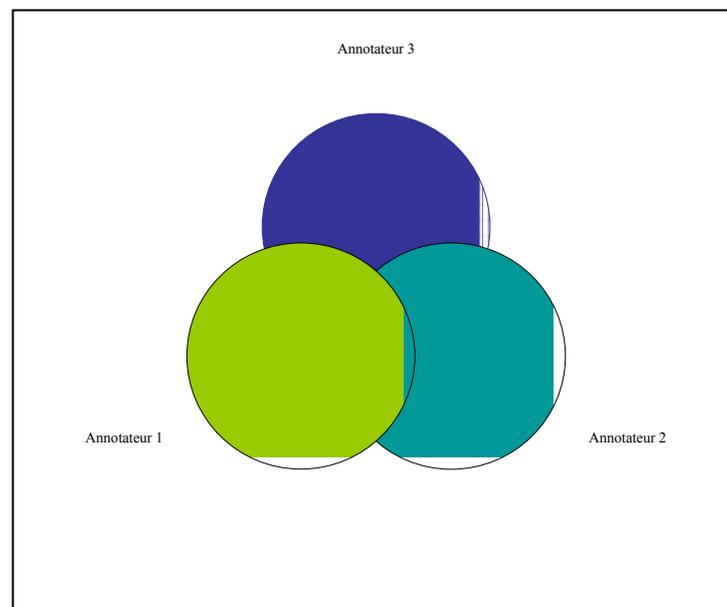
Toujours dans le cas de l'annotation sur un document spécifique, nous supposons que plusieurs annotations sont faites à leur tour par les divers acteurs. Chaque acteur fournit une annotation sur le document (ou le document annoté) en fonction de son expérience unique. C'est possible d'avoir les domaines d'intersection entre les annotateurs.

Dans ce cas d'une annotation partagée, nous considérons deux possibilités pour le partage (a) union des annotateurs (b) intersection des annotateurs.

Dans le cas où le partage d'information est fondé sur ce concept symbiotique, cinq expériences sont importantes dans le partage :

- Toutes les expériences de tous les annotateurs  $A_1 \cup A_2 \cup \dots \cup A_n$
- Toutes les expériences d'un sous-ensemble d'annotateurs  $A_i \cup A_n$ .
- Les expériences communes entre tous les annotateurs  $A_1 \cap A_2 \cap \dots \cap A_n$
- Les expériences communes entre deux annotateurs  $A_1 \cap A_2$
- Les expériences uniques de chaque annotateur  $A_1 \neq A_2 \neq A_3 \neq \dots A_n$

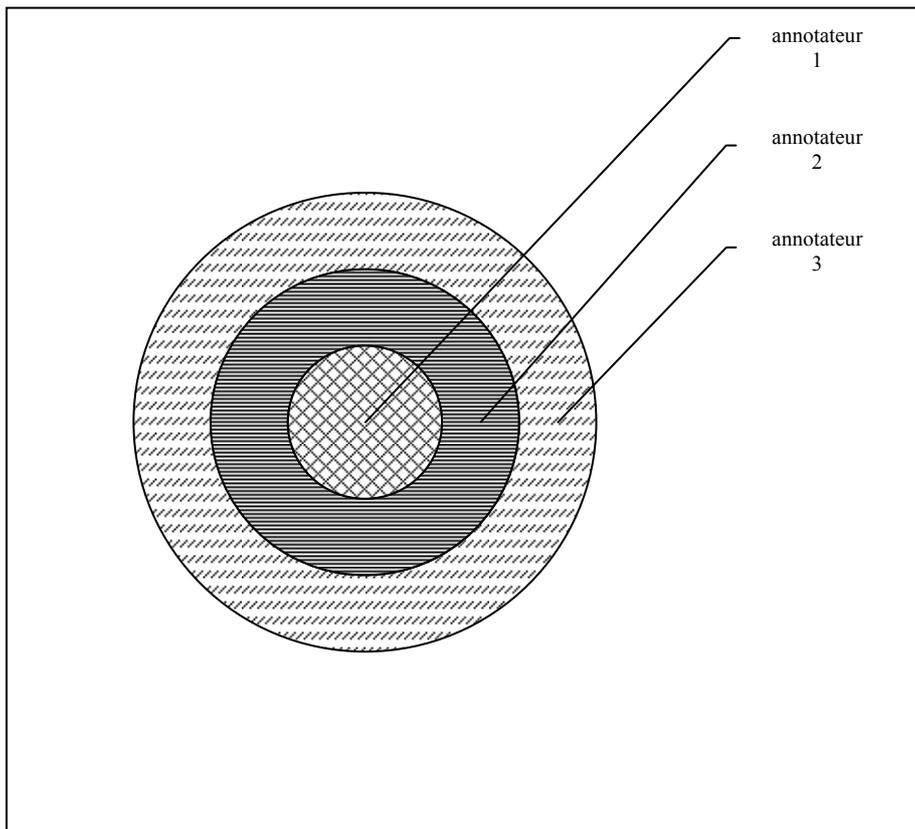
ou  $A_1, A_2 \dots A_n$  sont les annotations par les acteurs 1, 2, ....n



*Figure 2.6 : Partage d'annotation en symbiotique*

### *Partage en mode propagé*

Il y a quelques similitudes entre ce type de partage d'annotation et le type modéré. Le processus d'annotation est lancé par un acteur dans les deux cas. La différence dans le cas propagé est que, l'initiateur de l'annotation n'a pas de contrôle sur les annotations faites. Il n'est pas le modérateur des annotations, bien qu'il soit l'initiateur. On ne s'attend pas dans ce mode, à ce que les annotations retournent à l'initiateur.



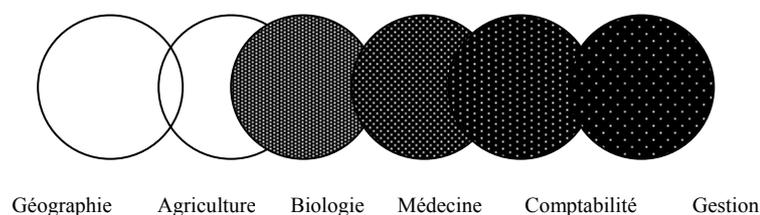
*Figure 2.7 : Partage d'annotation en propagation*

Ce type de partage d'annotation est largement répandu dans les situations où nous avons une hiérarchie entre les annotateurs. Dans la *figure 2.6*, un **annotateur 1**, peut lancer un travail collaboratif par les annotations, soit parce qu'il joue un rôle important dans la hiérarchie soit parce qu'il est le responsable direct du système qui propose l'annotation. Il envoie le document contenant l'annotation à une autre personne sous sa responsabilité au niveau hiérarchique. Successivement

*l'annotateur 2*, envoie le document avec sa propre version d'annotation à *l'annotateur 3*. Ce partage peut être utilisé en tant que moyen pour déléguer la responsabilité.

### ***Partage en mode transition***

Jusqu'à présent, nous avons discuté de trois types de méthodes d'agrégation de compréhension d'un document à travers l'utilisation d'annotation. Nous pouvons voir rapidement que le sujet d'annotation est relativement constant. Les annotations faites sont dépendantes des expériences des annotateurs. Nous avons délibérément négligé le sujet du contenu du document. Ce type de partage des expériences à travers les annotations est fortement lié au contenu du document. Dans la méthode de partage d'annotation de mode transition, une action d'annotation est initiée à partir d'un sujet particulier sans nécessairement considérer les autres sujets qui peuvent être liés. Le fait est que, il n'y a pas de discipline ou de domaine de connaissance que l'on peut isoler complètement des autres. Parce que les disciplines sont liées, l'initiation d'annotation ouvre un raisonnement interdisciplinaire vers un autre domaine de raisonnement.



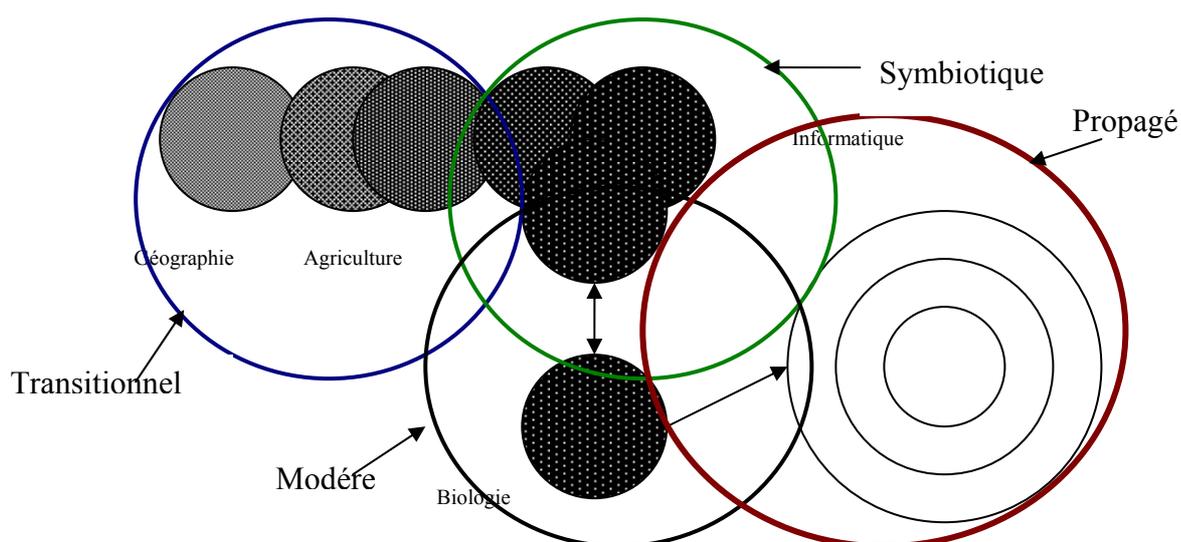
*Figure 2.8 : Partage en mode de transition*

Nous pouvons illustrer ceci grâce à un exemple. Supposons qu'un document dans le domaine de la géographie porte sur le sujet d'une éruption volcanique. Le document peut être sur comment elle affecte la production des produits alimentaires dans une région du monde. Le document peut recevoir une annotation par quelqu'un intéressé par le domaine de l'agriculture. Il peut dire *"les systèmes de production alimentaire peuvent être amélioré grâce au contrôle biologique si l'écologie de la région a été*

*affectée.*" Quelqu'un du domaine biologique peut voir cette annotation qui implique l'agriculteur et peut ajouter par exemple, "*L'anatomie des plantes pour ce contrôle est intéressant.*" Nous remarquerons que le document d'origine était du domaine de la géographie. Sur le document de géographie nous avons des annotations de sujet d'agriculture et sur le sujet d'agriculture, nous avons l'annotation du domaine biologique. Une personne du domaine médical peut ajouter aussi sur les annotations biologiques "*l'application de ces plantes dans le domaine médical est très importante.*" Si l'accumulation des ces annotations est envoyée à un décideur, il peut faire une annotation tel que "*Peut-on en estimer l'importance sur notre entreprise ?*"

### ***Partage en mode mixte***

En réalité, nous n'avons pas exactement un type unique de partage des expériences à travers l'annotation, mais un mélange de tout. Nous devons savoir qu'il est impossible de continuer le partage de l'annotation / agrégation des connaissances, fondé sur le contenu d'un document ou sur les utilisateurs. Le partage peut changer de l'un à l'autre. L'annotation peut commencer sur les expériences des utilisateurs dans le travail collaboratif et après devenir des annotations sur le contenu d'un document qui est référencé par un utilisateur.



*Figure 2.9 : Partage en mode mixte*

### 2.1.5.3.2. Le partage des expériences à partir de plusieurs documents

Le partage de plusieurs annotations par plusieurs personnes sur plusieurs documents adoptera une approche différente par rapport aux annotations sur un seul document. Les questions sont : Qui partage quoi ? Qui peut initier le partage d'annotation ? Les annotations sur des documents sont pour quel objectif ? Quels sont les critères pour choisir des participants dans les annotations ? Quels documents sont intéressants à considérer pour un travail de collaboration ? Quelles sont les situations de collaboration pour lesquels on peut utiliser l'annotation ?

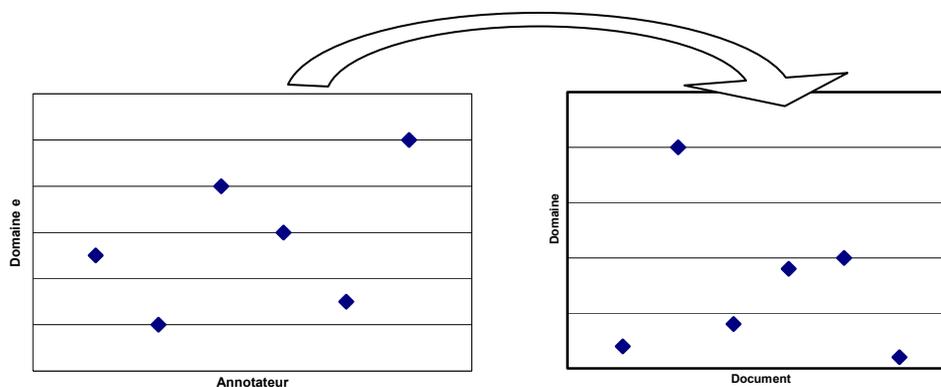


Figure 2.10 : Le mise en correspondance des annotateurs et des documents

Avec des annotations par plusieurs personnes sur plusieurs documents, deux zones de collaboration peuvent être imaginées : (a) zone des annotateurs et (b) zone des documents. Une manière de partager des annotations entre les deux zones est de créer une zone de documents qui classifie ceux-ci fondés sur les disciplines (domaine). La zone des annotateurs peut être classée aussi sur les mêmes paramètres utilisés pour classifier les documents. Dans ce cas une annotation dans la zone des annotateurs est mise en correspondance sur la zone des documents.

Nous avons discuté du partage des expériences par annotation dans la section précédente. Nous devons noter que, quand les annotations sont partagées, l'orientation et l'intention des participants peuvent être différentes. Les individus peuvent utiliser les annotations pour propager leurs propres idéologies aux membres

du groupe de la collaboration. Nous pouvons même avoir des individus qui introduisent des informations dérivées en utilisant des annotations. Nous discuterons de la relation entre l'annotation et la désinformation dans la prochaine section.

#### **2.1.5.4. La désinformation, la propagande et l'annotation**

La désinformation est un acte délibéré, souvent secret, pour la diffusion d'information incorrecte, habituellement avec l'intention d'influencer par déception les actions ou les avis des autres. C'est une technique employée dans les relations étrangères et dans les conflits armés pour tromper les adversaires. La désinformation peut être évaluée à partir des perspectives suivantes (a) autorité (b) exactitude (c) objectivité (d) période (e) couverture.

Puisque l'annotation est un acte d'interprétation de l'information, nous pouvons évaluer une annotation avec ces cinq paramètres s'ils sont conçus pour désinformer le public. Cette étude n'a pas pour objectif l'évaluation de l'annotation. Nous pouvons remarquer que le partage des annotations peut être employé en tant que moyen pour la désinformation.

Aujourd'hui, un des outils le plus importants pour la diffusion des informations est l'Internet. L'Internet est aussi un outil efficace pour diffuser les désinformations. Un exemple remarquable est la guerre d'Irak. Dans ce cas, les groupes concernés propagent leurs interprétations de la situation en utilisant le moyen d'Internet.

La propagande est définie comme « *le fitre qu'on met à la réalité pour qu'elle soit montrable* »<sup>17</sup>. Elle est aussi définie comme « *l'ensemble des mensonges qui vont le mieux permettre de réaliser des buts cachés et inavouables* ».

---

<sup>17</sup> <http://andre.bourgeois.9online.fr/>

La propagande est l'interprétation et la filtration d'information (Ellul, 1967) (Almeida, 1995). Le document d'origine qui contient l'information peut ne pas être modifié. C'est l'interprétation de document source qui tend à changer la compréhension du document source. Si c'est le cas, nous pouvons considérer l'annotation dans la lumière de la propagande, quand l'intention de l'annotateur est de tromper le public potentiel.

Nous entreprendrons de comparer quelques outils d'internet utilisés dans communication avec annotations.

#### **2.1.5.5. Les outils d'annotation et les outils de communication**

Nous n'avons pas de critères standard pour choisir un ensemble d'outils pour comparer les outils communicationnels avec des outils d'annotation. Puisque nous considérons l'annotation en tant que moyen de communication dans un environnement d'échange d'information, nous essayons de le comparer aux autres moyens communs de communication sur l'Internet. Les critères pour la comparaison ne sont pas exhaustifs.

A partir de ce tableau, nous pouvons comparer les fonctionnalités des outils d'annotation avec d'autres moyens de communication sur l'Internet. Nous remarquons que les outils d'annotation ont quelques avantages majeurs sur les autres moyens de communication entre un groupe de personne. Un outil d'annotation fournit la possibilité du filtrage d'informations spécifique dans le stockage d'information. Nous pouvons voir encore que l'annotation ne demande pas beaucoup de ressources informatiques comme la mémoire vivantes. Il est aussi possible de sauvegarder et analyser les informations stockées.

	<b>Outil d'annotation</b>	<b>Outil de « Tchatte » (ex. MSN, yahoo)</b>	<b>Courrier électronique</b>	<b>Blog</b>
<b>Contexte de discussion</b>	Centralisé	Peut changer avec le temps	Peut changer avec le temps	Centralisé
<b>Partage de la connaissance de soit</b>	Efficace	Immédiat	Immédiat	Efficace
<b>Stockage de la connaissance</b>	Excellent	Pas appropriée	Pas approprié	Efficace
<b>Sécurité de la ressource personnelle</b>	Dépend du système	Dépend de l'utilisateur	Excellent	Possible
<b>Réutilisation de connaissances personnelles</b>	Excellent	Pas approprié	Possible	Possible

Tableau 2.2 : Table de comparaisons des outils communicationnels sur internet

### 2.1.6. Le concept et les processus d'annotations

La façon dont nous utiliserons les annotations déterminera l'intérêt, les processus impliqués et leur création. Nous considérons simplement quelques caractéristiques reconnues aux annotations.

- **Cognitif** : L'annotation est un fragment textuel ancré au document, qui révèle une idée d'évaluation. Dans cette dimension, l'annotation permet aux acteurs à distance de proposer des interprétations sur un document et de d'interagir entre eux (Lortal et al, 2005a). Ce genre d'interprétation souligne l'influence de l'annotation sur le lecteur et non pas sur la création de l'annotation.
- **Communication** : La considération de l'annotation du point de vue communicationnel souligne le fait que l'annotation est créée pour un usage communicatif. Lortal et al (Lortal et al, 2005b) définit l'annotation comme l'élément discursif central d'une lecture critique.

- **Directive** : Nous pouvons regarder l'annotation comme un guide sur la façon dont un document ou une section d'un document devrait être utilisé. Par exemple dans les dictionnaires, nous avons souvent des notes qui accompagnent un mot montrant la manière dont celui-ci devrait être prononcé. Dans la musique, nous pouvons avoir des signes sur les notes qui soulignent comment celles-ci doivent être interprétées. Les signes (dits annotations) sont fondés sur les notes originales de la musique. Ainsi, certaines de ces annotations peuvent inclure l' « allégo », « pause » etc.

### 2.1.7. L'annotation dans le processus de recherche de l'information

La recherche d'information peut être comparée à un aveugle qui traverse une autoroute. Il perçoit l'information à travers sa main (bâton de marche) et ses oreilles, puis il décide de traverser la route. L'action de « traverser la route » n'est pas la même chose sur chaque route pour lui. Avant la décision de traverser la route, il analyse les informations reçues. Il compare les informations reçues avec ses expériences. Dans cet exemple, nous avons l'objectif, l'enjeu et les informations. L'aveugle a pour objectif de traverser l'autoroute. Les enjeux incluent le risque d'accident, le risque de perdre sa voie.

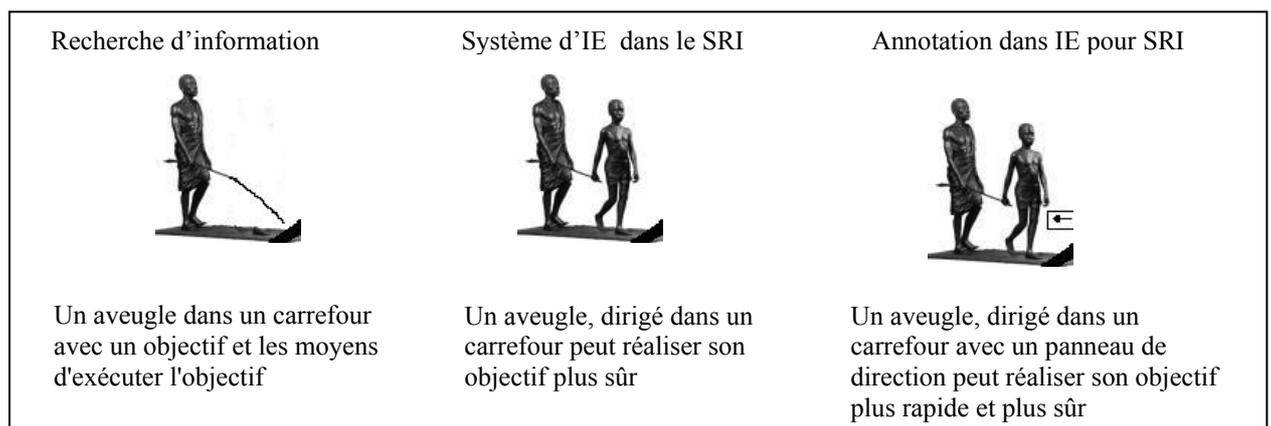


Figure 2.11 : L'annotation par rapport au SRI et IE

Dans la recherche d'information, un décideur a pour objectif de résoudre un problème décisionnel. Plusieurs enjeux sont invoqués comme l'enjeu de ne pas

trouver les informations pertinentes. Ces informations recherchées ont pour objectif de contribuer à la résolution des problèmes décisionnels.

Nous considérons trois cas si un aveugle veut traverser la route : il peut le faire tout seul, il peut le faire avec l'assistance de quelqu'un et sans les panneaux d'informations ou avec l'assistance de quelqu'un et avec les panneaux d'information.

Pour le premier cas, l'aveugle peut accomplir cet objectif avec ses expériences personnelles. Les implications sont le temps et les risques. Bien sûr, il ne peut pas utiliser tous les outils disponibles pour atteindre cet objectif.

Dans le deuxième cas, il peut traverser la route avec l'assistance de quelqu'un. L'assistant travaillera en collaboration avec l'aveugle. Il fera cette collaboration en utilisant sa propre vue, ses expériences dans des situations similaires et son raisonnement spécifique. L'expérience de l'aveugle est moins importante dans cette situation.

Dans le troisième cas, l'assistant utilise les panneaux d'informations en plus de ses expériences personnelles. Il va combiner ses expériences dans la situation de « traverser les routes » avec les expériences dans les autres domaines de sa vie et « comment lire et interpréter les panneaux d'informations ».

Comparons des trois situations avec le processus de la recherche d'informations et l'intelligence économique. Un aveugle peut être comparé à un décideur qui a besoin d'informations pour la prise de décision. L'objectif est de trouver les informations pertinentes pour résoudre un problème décisionnel. L'assistant peut être comparé à un veilleur qui connaît les méthodes et les techniques de recherche d'informations. Le veilleur et le décideur ont des environnements communs à partager. Les expériences du veilleur ainsi que les expériences du décideur peuvent être combinées pour résoudre un problème décisionnel.

Les panneaux de directions sur la voie d'un aveugle sont généralement utiles pour l'assistant de l'aveugle et non pour l'aveugle lui-même. Les panneaux sur les routes sont les indications que l'on peut comparer aux annotations sur les documents. Les panneaux montrent comment la route peut être utilisée ou comment la route a été utilisée. Dans le cas de recherches d'informations, les annotations sont sur les documents afin de montrer comment le document a été utilisé et comment celui-ci peut être utilisé. Les panneaux d'indication ne font pas partie de la route, mais des commentaires sur la route. Ils ont été faits par un utilisateur de la route avec l'intention de guider quiconque passera par celle-ci.

## **2.2. Création et stockage d'annotation**

Plusieurs travaux ont été effectués sur la création et stockage d'annotation avec les points de vue différents. Microsoft a proposé le "Standard Annotation Language"<sup>18</sup> comme point de référence pour des annotations électroniques. Il est difficile d'imaginer la signification du mot « standard » ici. Comment pouvons-nous connaître les critères « standards » pour décrire les besoins de l'homme et les possibilités par les annotations. Si nous devons nous limiter à la création de l'annotation, nous ne pouvons pas être satisfaits de ce « standard » considérant les différentes méthodes d'annotation et de médias différents qui peuvent être impliqués. Ce qui est considéré comme « standard » ne correspond pas à l'implication et l'utilisation de à l'annotation. Elles sont simplement appliquées à la création des annotations. Nous considérons quelques propositions de représentation des annotations dans les sections suivantes.

### **2.2.1. XML : la base de spécifications des annotations**

XML (en anglais Extensible Markup Language) est la norme définie par le consortium W3C<sup>19</sup>. XML est considéré à l'origine comme un langage facilitant la

---

<sup>18</sup> <http://msdn2.microsoft.com/en-us/library/ms235402.aspx>

<sup>19</sup> <http://www.w3c.org>

définition, la validation et le partage de différents formats de documents sur le Web. Descendant direct de la norme de description documentaire SGML (Standard generalized markup language), il suscite aujourd'hui un véritable engouement partout où il est question d'échange et de partage de données. Contrairement à HTML (Hyper Text Markup Language), confiné à la présentation des informations sur un poste Web, **XML** s'efforce de leur donner un sens et de les structurer, comme au sein d'une base de données.

On qualifie **XML** de métalangage : un langage pour la représentation des données portant sur le contenu d'un document et pas uniquement sur son apparence.

Comme leurs acronymes l'indiquent, SGML (Standard Generalized Markup Language), **XML** (eXtensible Markup Language) et HTML (HyperText Markup Language) sont des langages de balisage. **XML** et HTML sont dérivés de SGML. SGML est décrit par la norme ISO (ISO-8859) pour la structuration des documents textuels. Ce langage permet de "baliser" un document de façon très précise. Ce haut degré de précision, de structuration, de portabilité et de pérennité que permet SGML se paie par une certaine complexité qui a jusqu'à présent, limité son usage à quelques applications particulières. HTML a été développé pour répondre aux besoins spécifiques de la mise en page sur le Web d'informations textuelles (au sein desquelles il est possible de référencer divers fichiers, par exemple, des images ou du son). Mais HTML a des limitations : pas de possibilité de représenter des structures d'informations. **XML** a été proposé pour pallier les faiblesses d'HTML sans qu'il y ait besoin d'avoir recours à toute la difficulté de SGML.

Les dialectes XML (WML, XSLT, XML Schema, XHTML, RDF/XML, SOAP, SMIL, MathML, SVG) sont décrits de façon formelle : une structure de données simple est définie avec une DTD<sup>20</sup> (*Document Type Definition*), une structure de

---

<sup>20</sup> Le **Document Type Definition** (DTD), ou **Définition de Type de Document**, est un document permettant de décrire un modèle de document SGML ou XML. Une DTD ne décrit cependant que la structure du document (hiérarchie des champs, paramètres, type des données...) et non, par exemple, les valeurs autorisées des champs ou paramètres, ce en quoi elle se distingue de Schéma XML, Relax NG et Schematron. De plus, la norme DTD fait appel à une syntaxe spécifique distincte de XML. Une DTD n'est donc pas un document XML.

données détaillée est définie avec un XML Schema ou tout autre DSDL (*Document Schema Definition Languages*, c'est-à-dire langage de description de schéma).

L'efficacité de XML dépend fortement des normes strictes qui régissent sa structure. Pour être considéré comme un document XML, un document doit être *bien formé*, c'est-à-dire qu'il doit être conforme aux règles suivantes :

- Dans sa première déclaration, le document doit être identifié comme un document XML. Cette première déclaration est nommée le *prologue*. Il doit contenir des informations sur la version de XML utilisée, le codage des caractères (si nécessaire) et indiquer également si le fichier XML est associé à une DTD ou à un autre langage de définition de document (XML Schema, Relax NG, Schematron, etc.) ou s'il est autonome.
- Un document XML ne doit avoir qu'un seul élément *racine*, tous les autres éléments sont contenus dans cet élément, aussi appelé l'*élément document*.
- Chaque élément XML doit commencer par une balise ouvrante et se terminer par une balise fermante. Un élément vide peut être représenté par une *balise d'élément vide* qui ressemble à `<exemple/>`; cette balise est considérée comme étant une balise ouvrante suivie d'une balise fermante. Ceci est utilisé pour éviter de devoir écrire `<exemple></exemple>` tout en conservant le bon formatage.
- La valeur d'un attribut doit être entre guillemets (simple « ' » ou double « " »).
- Il est à noter que XML est sensible à la casse et typographe. Par exemple, « `<Exemple></Exemple>` » est une paire de balises bien formée alors que « `<Exemple></exemple>` » n'en est pas une.

Quand un document XML est *bien formé* et conforme à la DTD auquel il est associé, ce document est qualifié de *valide*.

Un document XML peut être associé à des feuilles de style XSLT, feuilles qui permettent de générer de nouveaux documents contenant des informations provenant du document XML. Il est ainsi possible de générer des fichiers XML ayant une structure différente du document initial (transformation), mais aussi d'autres

documents : requêtes SQL, pages HTML, etc. Un document XML peut aussi être affiché par certains navigateurs Web (par exemple Internet Explorer, Mozilla Firefox, Netscape etc.) en utilisant une feuille de style CSS.

Les langages de représentation des annotations ne sont pas tout à fait les dialectes de XML mais sont les spécifications de XML pour les annotations.

### **2.2.2. Resource Description Framework (RDF)**

Le “resource description framework” (RDF) est une spécification qui fournit un système d'ontologie léger pour soutenir l'échange de la connaissance sur le Web, fondé sur un modèle de graphique semblable à l'OEM. Le RDF est une langue d'usage universel pour représenter l'information dans le Web. RDF est sur des composant triples (sujet, prédicat, objet), et peut être arrangé en série dans XML, N-TRIPLES, ou graphique. Les sujets et les prédicats sont identifiés en utilisant URIs, tandis que les objets peuvent être URIs ou littérales d'un type de données (telles que les chaînes, le nombre entier, le flottant, la date...). Puisque RDF est destiné à l'échange des données sur le Web, les URIs sont utilisés pour identifier des concepts, des propriétés, et des objets. Le RDF indique simplement l'ensemble minimal d'attributs (rdf:type), et des couches plus élevées telles que RDFS définissent davantage des attributs pour une langue d'ontologie plus complète.

Après avoir présenté succinctement les langages les plus couramment utilisées dans les outils d'annotation, nous présentons dans la section suivante les projets d'annotation existants qui ont donné lieu à des applications concrètes.

### **2.2.3. Les projets sur l'annotations**

Un « **framework** » d'annotation (en anglais, **Application Framework**) est un terme informatique (ou cadriciel). C'est un ensemble de bibliothèques (en anglais libraries) permettant le développement rapide des applications. Il fournit suffisamment de

briques logicielles pour pouvoir produire une application aboutie. Ces composants sont organisés pour être utilisés en interaction les uns avec les autres. Ils sont en principe spécialisés pour un type d'application.

Un cadriciel (terme en usage depuis 1997) est un ensemble de classes abstraites collaborant entre elles pour faciliter la création de tout ou partie d'un système logiciel. Un cadriciel fournit un guide architectural en partitionnant le domaine visé en classes abstraites et en définissant les responsabilités de chacune ainsi que les collaborations entre les classes. Un cadriciel est habituellement implémenté à l'aide d'un langage objet, bien que cela ne soit pas strictement nécessaire. Le déploiement à grande échelle de bibliothèques d'objets exige un cadriciel. Celui-ci fournit un contexte où les composants sont ré-utilisés. (wikipedia.org).

Les « frameworks » d'annotations sont des guides spécifiques qui indiquent comment les annotations doivent être faites. Ils fournissent la langage et la structure pour les annotations que l'on peut faire. Ils précisent également les éléments primordiaux (clés) (ce qui **doit** être inclus), les éléments secondaires (ce qui **peut être** inclus) et ce qui **ne peut pas** être inclus. Chaque « cadre » a ses propres objectifs. La plupart des « cadres » sont écrits avec le langage XML.

Nous avons décrit le langage XML comme la base de plusieurs « framework » et spécifications de stockages et communication dans les travaux d'annotations. Nous présentons dans les sections suivantes quelques « frameworks » d'annotations. Nous allons essayer de voir s'il y a des similitudes et des interactions entre les modèles et les « framework » des annotations. Ainsi nous distinguerons les « frameworks » d'annotation et des modèles d'annotations. En effet, les cadres des annotations donnent une vue générale sur l'ensemble de ce qui peut être adapté d'une manière plus large et dans un domaine spécifique. Un modèle d'annotation est plus spécifique pour atteindre des objectifs particuliers.

### 2.2.3.1. ANOTEA

L'ANNOTEA est un « framework » d'annotation pour le développement des systèmes de création et d'édition des annotations en commun sur des documents du Web. ANNOTEA a été construit sur le HTTP, le RDF, et le XML. Il fournit un protocole interopérable pour les navigateurs. L'objectif principal est pour permettre aux utilisateurs d'attacher des données aux pages Web pour que d'autres utilisateurs puissent, à leur choix, voir les données jointes quand ils passent en revue plus tard les mêmes pages. Le protocole d'ANNOTEA fonctionne sans modifier le document original ; c'est-à-dire, il n'y a aucune condition pour que l'utilisateur accède à la page Web annotée. Le protocole d'ANNOTEA convient pour les données d'annotation qui sont faites pour être visionnées par des internautes pour d'autres programmes d'application, tels que les outils de classification, les moteurs de recherche, et les applications automatiques.

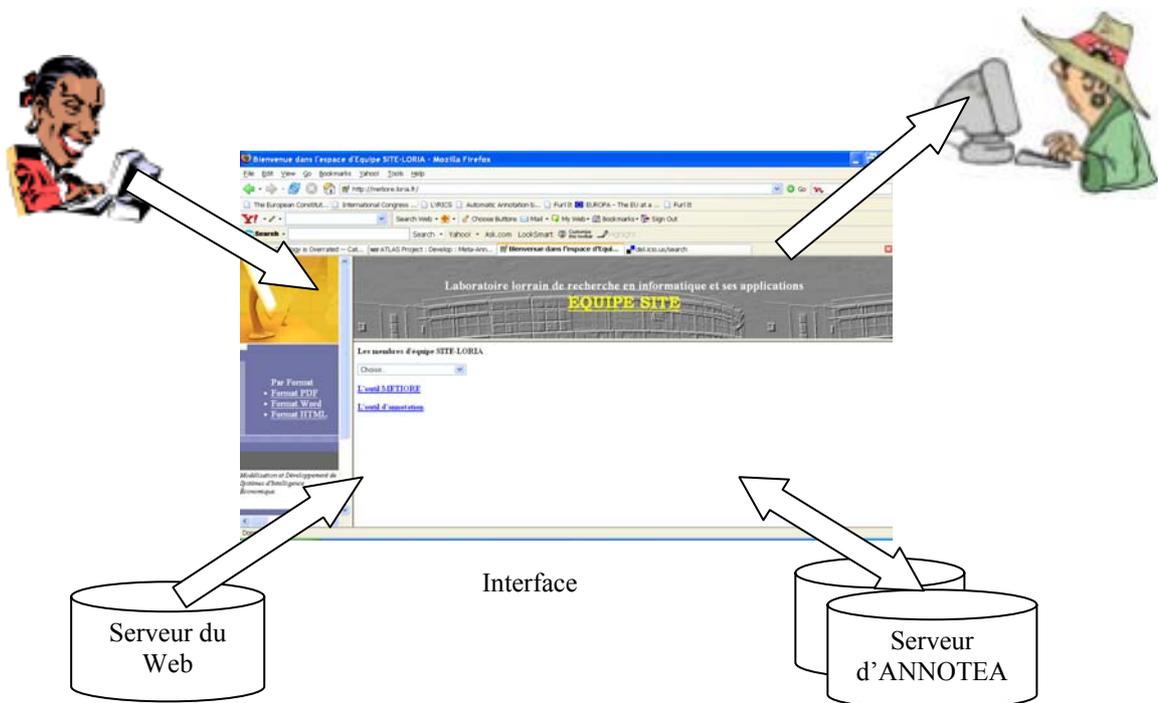


Figure 2.12 : L'architecture generale de cadre d'ANNOTEA

Un des objectifs du « framework » d'ANNOTEa a été de réutiliser autant que possible les technologies existantes de W3C. Les méthodes appliquées la plupart du temps sont la combinaison de RDF, XPointer, XLink, sur HTTP. L'objectif est de permettre l'accès de sorte que chaque utilisateur ayant un accès au serveur d'annotation puisse consulter les annotations liées au document et ajouter ses propres annotations.

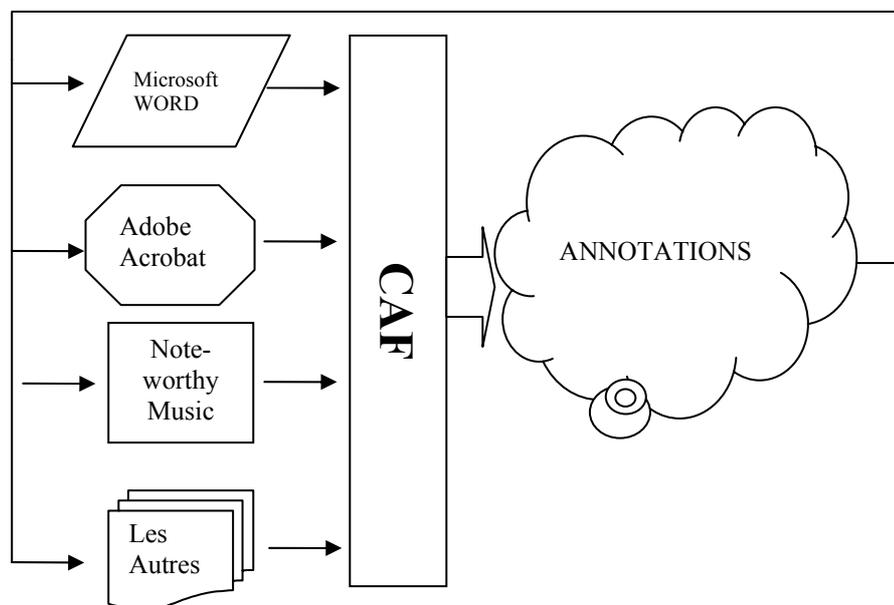
<b>Classe</b>	<b>Description</b>
rdf:type	Le type de l'annotation, qui devrait montrer l'intention de l'annotateur.
annotates	Les ressources auxquelles l'annotation est liée.
body	Le contenu de l'annotation.
context	Le "context" utilisant un XPointer, indique l'endroit exact dans un document où l'annotation est attachée.
dc:creator	L'annotateur d'une annotation
created	La date et le temps de la création de l'annotation.
dc:date	La date et le temps de la modification de l'annotation.
related	Les ressources liées aux informations du document, par exemple, URL, les discussions etc.

*Tableau 2.3: Les propriétés d'une annotation en format RDF pour ANNOTEa*

L'ANNOTEa fournit les types d'annotation suivants : Conseil (advice), Changement (change), Commentaire (comment), Exemple (example), Explication (explanation), Question (question), Voir aussi (see also). L'identification de l'utilisateur est stockée avec la date de création. Un XPointer indique l'endroit où le texte annoté, et de plusieurs autres propriétés. Chaque annotation créée est liée à son propre URI, ainsi les nouvelles annotations peuvent annoter les plus anciennes. En termes d'activité agrégation de contexte, l'approche d'ANNOTEa semble plus adaptée à annoter le contenu existant, le contenu peut être souligné, mais seulement en tant qu'élément du processus d'annotation.

### 2.2.3.2. Microsoft Common Annotation Framework (CAF)

L'objectif du Microsoft Common Annotation Framework (CAF) est de créer un standard généralisé pour annoter les documents sur les pages Web. Le cadre crée un forum commun pour créer et partager des annotations à travers des applications diverses. L'objet est de favoriser l'interopérabilité des annotations entre des applications hétérogènes. C'est-à-dire que l'annotation ne devrait pas dépendre de l'application initiale où l'annotation a été créée mais devrait être "portable" sur d'autres applications. Le « framework » a été fondé sur le modèle logique et se compose d'un schéma compatible avec XLink. Il a pour objectif de soutenir l'annotation sur les médias hétérogènes.



*Figure 2.13 : Common Annotation Framework (CAF) based on Logical model*

Un outil WebAnn est fondé sur le « framework » CAF pour permettre discussion entre les étudiants. Il est un moyen pour la discussion publique d'un groupe. Bien que l'outil WebAnn ait été basé sur le « framework » de CAF, ce « framework » n'est pas connu comme étant une norme pour les annotations des pages web ceci en raison du fait que (a) c'est trop général pour servir de but à la plupart des annotations (b) les communications entre les applications sur le web ne sont pas

statiques (c) le stockage et la structure de l'annotation ne sont pas correctement traités dans CAF

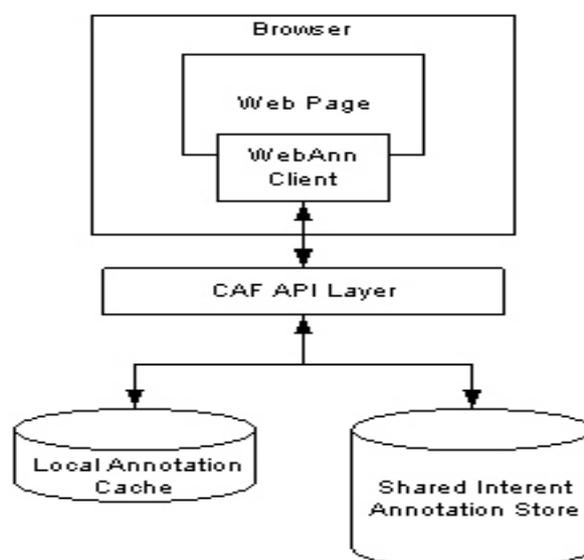


Figure 2.14 : The CAF-based WebAnn tool<sup>21</sup>

### 2.2.3.3. Extensible MultiModal Annotation Language (EMMA)

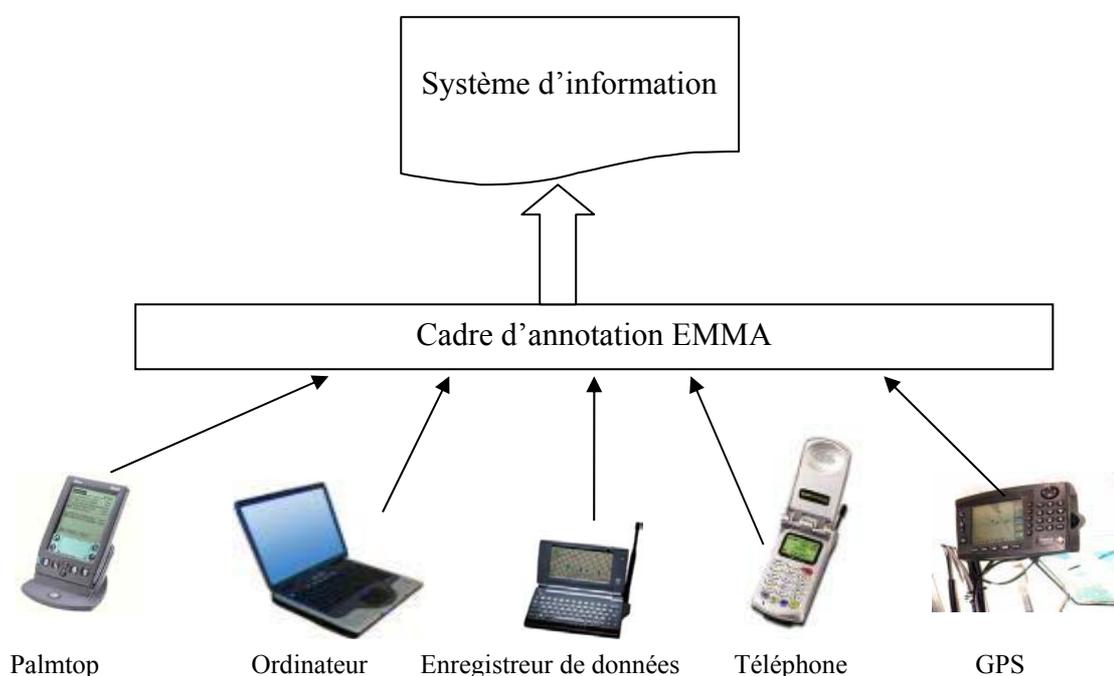
L'EMMA (Extensible MultiModal Annotation language) le langage extensible d'annotation pour les application MultiModal est un langage utilisé pour représenter l'entrée de l'utilisateur dans une application multimodale. On peut le voir comme le « framework » d'interaction Multimodal de W3C. C'est un mécanisme d'échange entre les dispositifs d'entrée d'utilisateur et les possibilités d'une application. Le langage est concentré sur l'annotation des interprétations des informations sur les entrées simples et composées, par rapport à l'information qui pourrait avoir été rassemblée pendant ou au cours d'un dialogue.

L'EMMA a été développé pour permettre l'annotation des informations par les entrées de dispositifs hétérogènes. C'est un rassemblement d'informations multimédia, multimodales, multi plateformes et les sources d'informations ainsi que de systèmes d'information hétérogènes. En fait, c'est une approche pour fournir une

---

<sup>21</sup> Source : David Barger et al, 2001

plateforme commune pour des protocoles et des représentations différentes l'information. Nous avons actuellement des représentations de l'information qui sont particulières aux dispositifs. Certaines de ces représentations incluent WML (Wireless Markup Language) une langage de balisage pour les appareils sans fil, qui a été développé pour la représentation de l'information sur les dispositifs sans fil. Souvent la question est, comment un dispositif sans fil, avec sa représentation de l'information WML, peut être employé pour annoter les informations sur un réseau optique ? Ceci demande un « framework » d'annotation qui favorise les entrées à travers les dispositifs hétérogène.



*Figure 2.13 : Le cadre d'annotation EMMA*

D'autres objectifs d'EMMA sont de permettre des composants tel que (a) reconnaissance de la parole (b) reconnaissance d'écriture (c) moteur de compréhension du langage naturel (d) autres médias d'entrée (par exemple DTMF, pointage, clavier) et (e) Composant d'intégration de Multimodal pour générer un balisage d'EMMA :

#### **2.2.3.4. Linguistic Annotation Framework (LAF)**

Le « framework » d'annotation linguistique est conçu pour l'annotation dans la perspective de l'utiliser pour le traitement du langage naturel (Bird et al, 2000). On croit que les données peuvent être augmentées (annotées) avec l'information linguistique telle que les catégories morphosyntaxiques, la structure syntactique ou les structures de discours, l'information co-référence, etc. ; On suppose également que l'information peut être alignée pour des correspondances (par exemple, traductions parallèles, son articulé et transcription).

L'objectif du « framework » d'annotation linguistique est de pouvoir mettre en place un outil pour identifier « un segment » d'information comme un mot, une phrase, un paragraphe, etc. avec l'utilisation des annotations attachées. La méthodologie impliquée inclut le cadrage syntaxique ou la délimitation de chacun des mots dans un document avec un balisage XML qui identifie le segment comme un mot, une phrase, etc.

Dans les spécifications de cadre d'annotation linguistique (Ide et al, 2003), quelques conditions générales ont été imposées. Certaines de ces conditions incluent :

##### *Le degré d'expressivité*

Le cadre doit fournir des moyens de représenter toutes les variétés d'information linguistique (et probablement aussi d'autres types d'information). Ceci inclut la représentation de la gamme complète de l'information en général à l'information au niveau le plus fin de la granularité.

##### *L'indépendance de médias*

Le cadre doit manipuler tous les types de supports potentiels, y compris le texte, le son, la vidéo, l'image, etc. et devrait, en principe, fournir les mécanismes communs

pour les manipuler tous. Le cadre est fondé sur les normes existantes ou les normes en cours de développement pour représenter le multimédia.

#### *L'adéquation sémantique*

- Les structures de représentation doivent avoir une sémantique formelle, y compris les définitions des opérations logiques.
- Il doit exister une manière de partager des descripteurs et des catégorisations de l'information centralisée.

#### *Possibilité d'incrémenter*

- On est censé soutenir les différentes étapes pour l'interprétation d'entrée et la génération de sortie.
- Prévoir la représentation des résultats partiels /non-spécifié et des ambiguïtés, les rechanges, etc. et leur fusionnement et comparaison.

#### *L'uniformité*

Les représentations doivent utiliser les mêmes "modules de constructions" et les mêmes méthodes pour les combiner

#### *L'adaptation*

Le cadre ne doit pas dicter des représentations qui dépendent d'une simple théorie linguistique.

#### *L'extensibilité*

Le cadre est prévu pour les déclarations et les échanges des extensions à l'enregistrement de catégorie de données centralisé.

#### *La lisibilité humaine*

Les représentations doivent être lisibles, au moins pour la création et l'édition.

#### *L'explicitation*

L'information dans un dispositif d'annotation doit être explicite -- c'est-à-dire, l'interprétation ne devrait pas être faite par le logiciel de traitement.

### *L'uniformité*

Les mécanismes différents ne devraient pas être utilisés pour indiquer le même type d'information.

## **2.2.4. Les modèles d'annotation existants**

Il y a beaucoup de modèles d'annotation aujourd'hui sur le Web et pour les usages personnels. Nous pouvons considérer le processus d'annotation comme un processus de rétroaction. La rétroaction a des importances et des applications diverses non pas seulement dans le marketing mais dans le domaine de développement de produit, la gestion de l'information et dans la recherche et le développement scientifique. En raison de cette application large, l'annotation exige des modèles variables pour répondre aux spécificités des domaines d'application.

### **2.2.4.1. Modèle fondé sur l'organisation du contenu**

Ces modèles fondés sur l'organisation de contenu ne sont pas généralement concernés par les utilisations éventuelles de l'annotation ni la création d'annotation. Ils ne sont pas fondamentalement préoccupés par les méthodes de stockage. L'objectif est l'accès aux annotations stockées.

#### **2.2.4.1.1. Modèle sémantique**

Nous centions dans cette section sur le mot « Web sémantique ». La question est qu'est ce que le « Web sémantique » ? Le Web sémantique est un projet qui prévoit de créer un moyen universel pour l'échange de l'information de sorte qu'un système (ou un logiciel) puisse connaître la signification du contenu des documents du Web. Le Web sémantique a pour objectif de rendre des pages Web compréhensibles par les outils automatiques, de sorte que les logiciels puissent rechercher des sites Web et effectuer des actions d'une manière standard.

Par exemple, comment l'ordinateur peut-il interpréter la phrase « Je cherche un document récents sur les dirigeants des pays développés ? » D'abord, le système doit

comprendre la signification les expressions « document courant », « dirigeants », « pays développés ». Le Web sémantique essaye de voir comment organiser les ressources électroniques de telle sorte que les ordinateurs puissent résoudre les problèmes comme ci-dessous.

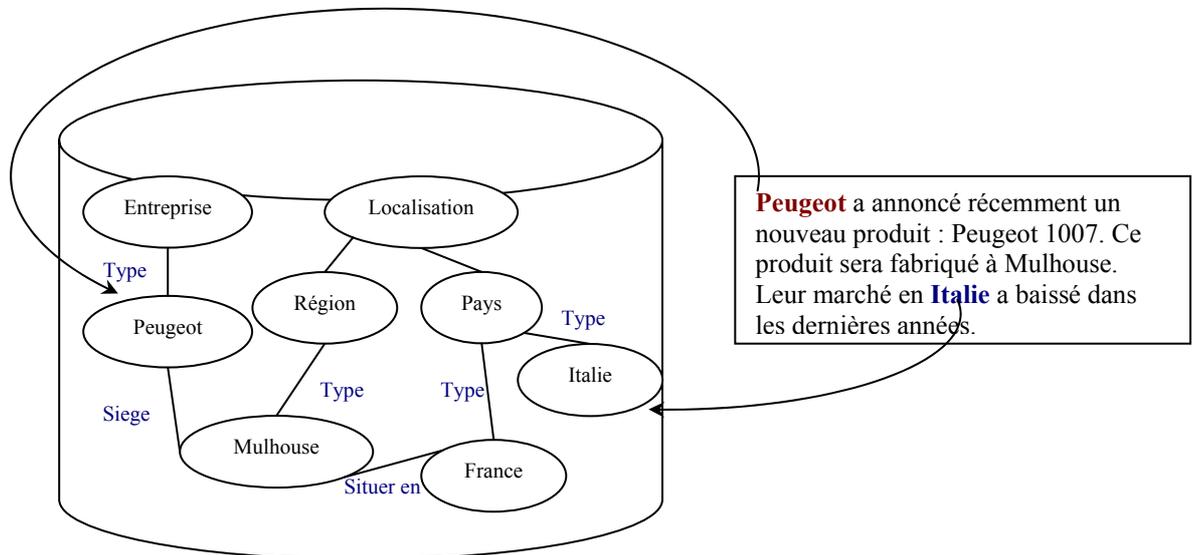


Figure 2.14 : Un exemple de dépôt sémantique

Par exemple, pour la phrase « Peugeot...1007...Italie » dans le schéma (Figure 2.15), le système devrait utiliser l'organisation sémantique pour savoir que « l'Italie est un pays où Peugeot est localisé ».

#### 2.2.4.1.2. Modèle ontologique

L'ontologie est une organisation systématique de toutes les catégories importantes des objets ou des concepts qui existent dans un certain domaine de discours, montrant les relations entre eux<sup>22</sup>. Quand elle est complète, une ontologie est une catégorisation de tous les concepts dans un certain domaine de la connaissance, y compris les objets et toutes les propriétés, relations, et fonctions nécessaires pour définir les objets et pour indiquer leurs actions. Une ontologie simplifiée peut contenir seulement une classification hiérarchique (une taxonomie) montrant le type de relations de subsumption entre les concepts dans

<sup>22</sup> <http://www.answers.com/topic/ontology-computer-science>

le domaine du discours. L'ontologie peut être visualisée comme un graphique abstrait avec des noeuds et des arcs représentant les objets et les relations. Les concepts inclus dans l'ontologie et la commande hiérarchique seront dans une certaine mesure arbitraires, dépendant du but pour lequel l'ontologie est créée. Ceci résulte du fait que les objets sont les variables d'importance pour différents buts, et différentes propriétés. Des objets peuvent être choisis comme critères par lesquels ceux-ci sont classifiés. En outre, différents degrés d'agrégation des concepts peuvent être employés, et les distinctions d'importance pour un but peuvent être sans souci pour un but différent.

Le tableau périodique des éléments chimiques est présenté avec des couleurs distinctes pour chaque groupe principal et transitionnel. Les groupes sont étiquetés en haut : 1A, 2A, 3A, 4A, 5A, 6A, 7A, 8, 1B, 2B, 3B, 4B, 5B, 6B, 7B, 0. Les périodes sont numérotées de 1 à 7 à gauche. Les éléments sont classés par ordre croissant de numéro atomique.

Figure 2.15 : La table périodique est un exemple d'organisation ontologique

La table périodique des éléments est un exemple typique d'une organisation ontologique.<sup>23</sup> Elle organise des éléments par le nombre de protons du noyau. Les éléments sont énumérés par le nombre atomique par ordre croissant (c.-à-d. le nombre des protons au noyau atomique).

Nous voulons souligner que malgré les problèmes inhérents dans les organisations de contenus, il y a des avantages associés. Certains avantages incluent :

- Avec l'ontologie, il est possible de voir les relations entre les éléments
- Il est possible de voir la hiérarchie des objets dans une organisation d'ontologie

<sup>23</sup> <http://sw.derri.ie/svn/aharth/2004/04/query-languages/dip-wp21.html>

- Elle fournit la possibilité de représenter du contenu de réseau sémantique.

Malgré ces avantages, nous pouvons identifier quelques problèmes liés aux modèles fondés sur l'organisation des informations. Nous les discuterons dans la prochaine sections.

#### **2.2.4.1.3. Les problèmes liés aux modèles organisationnelles**

L'organisation de contenu fournit un très bon point de départ pour l'organisation de la connaissance, mais l'organisation fondée sur l'organisation de contenu comporte des problèmes : Problème de stabilité d'organisation dans le temps. Une organisation que l'on assure impeccable aujourd'hui peut s'avérer insatisfaisante demain. Par exemple, si les livres étaient organisés sur les disciplines, comment expérimenterions-nous les évolutions dans les disciplines ?

Nous pouvons noter le problème d'intégration des différences culturelles. Par exemple, ce qui est vu comme des articles religieux dans une certaine société africaine peut être considéré comme des articles d'arts aux Etats-Unis.

Nous pouvons interpréter ou classifier des informations fondées sur une référence préconçue. Une tasse classifiée comme moitié remplie peut également être classifiée comme une tasse à moitié vide. Comment ces questions peuvent-elles être résolues par l'approche organisationnelle ?

#### **2.2.4.1.4. Quelques langages d'organisation de contenus**

L'information doit être organisée afin de faciliter son analyse, sa synthèse, sa compréhension et sa communication. L'organisation d'information est importante parce qu'elle nous permet de :

- Gerer l'information plus efficacement.
- Communiquer l'information plus efficacement.
- Identifier les informations pertinentes et rejeter les informations inutiles.
- Identifier les tendances dans l'information recueillie.
- Synthétiser les informations hétérogènes pour une nouvelle connaissance.

Différents langages sont employée dans l'organisation de l'information. Les divers langages de représentation sont traités dans les sections suivantes. En fait, il y a une tendance vers les dispositifs d'intégration d'un langage à un autre pour permettre plus d'expressivité. Une organisation d'information définit un vocabulaire pour décrire des informations on parle alors de classification, d'agrégation, et de généralisation (Horrocks et al, 2000). Le modèle fondamental de données est souvent sans schéma explicatif, donc nous commençons notre présentation par des modèles de données pour modéliser des données semi-structurées, telles que l'OEM et le RDF.

#### **2.2.4.1.4.1. Object Exchange Model (OEM)**

Le modèle d'échange d'objet est un format pour l'échange des données semi-structurées. L'OEM a été utilisé dans divers systèmes de prototype de recherches pour échanger des objets entre les bases de données pour des tâches d'intégration de l'information. L'OEM est semi-structuré et auto-descriptif et se compose de quatre champs : identification, étiquette, données, et type.

Dans un modèle d'échange d'objet, l'échange de l'information est vu comme l'échange des objets. Selon (Papakonstantinou, 1995), chaque valeur d'information à échanger est représentée comme le label (ou étiquette) qui décrit sa signification. Par exemple, si nous souhaitons échanger la valeur de la température de 70 degrés centigrade, nous pouvons la décrire comme

(température-en-degré, integer, 70)

Là où la chaîne «température-en-degré» est une étiquette humainement

compréhensible le « integer » (nombre entier) indique le type de valeur, et « 70 » est la valeur elle-même. Dans ce cas chaque objet peut avoir sa propre étiquette. Par exemple,

(ensemble-des-poids, ensemble, {*wetg1*, *wetg2*})

*wetg1* est (poid-en-kilogramme, entier, 52)

*wetg2* est (ensemble-en-livre, entier, 122)

Une caractéristique principale d'OEM est qu'elle est auto descriptive. Nous n'avons pas besoin de définir en avance la structure d'un objet, et il n'y a aucune notion de classe fixe de schéma ou d'objet. Dans un sens, chaque objet contient son propre schéma. Par exemple, la « température-en-degré » au-dessus joue le rôle d'un nom de colonne, où l'objet à stocker serait le domaine pour cette colonne.

La différence entre un schéma de base de données et l'OEM est que, une étiquette ici peut jouer deux rôles: identifier un objet, et identifier la signification d'un objet (Papakonstantinou, 1995). Pour illustrer, considérons l'objet suivant:

(enregistrement-personnel, ensemble, {*cmpnt1*, *cmpnt2*, *cmpnt3*})

*cmpnt1* est (nom-person, chaine, "Francois")

*cmpnt2* est (bureau, entier, 333)

*cmpnt3* est (secteur, chaine, "vente")

#### 2.2.4.1.4.2. Topic Maps

Les **cartes thématiques**, (en anglais **Topic Maps**) constituent un outil très général de représentation des connaissances, dont le but est d'agréger autour d'un point unique d'indexation (appelé topic) toutes les informations disponibles concernant un sujet donné, et de relier ces points par un réseau sémantique de relations

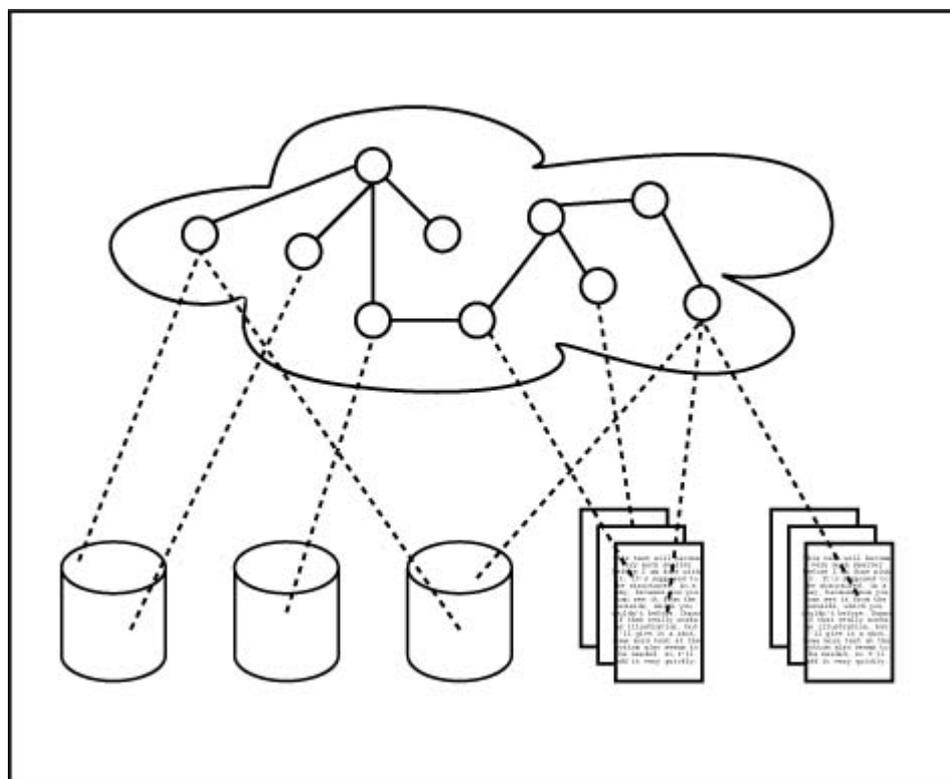
appelées *associations*.

Un «topic maps » représente une information en utilisant des « sujets» (*topics" en anglais*) qui représentent tout un concept, tels qu'une personne, un groupe de personnes, une couleur, un pays, une organisation, un module logiciel, un fichier individuel, des événements, des « associations» qui représentent les relations entre ces « sujets », et des « occurrences» qui représentent des relations entre des sujets et des ressources informationnelles qui s'y rapportent.

L'intérêt des Topics Maps est de définir des contextes et profils d'utilisateurs particuliers et de faciliter la fusion de topic maps provenant de sources des documents différentes.

Un «Topic maps» est un type de couche d'index ou d'information qui peut être construit séparément d'un ensemble de documents. Le «Topic maps» permet d'identifier des sujets et des relations dans l'ensemble de documents. Selon les spécifications de Topic Maps de XML (XTM1.0) de TopicMaps.org, l'objectif d'un «Topic maps » est de transmettre la connaissance au sujet des documents par une couche superposée sur des documents. Un «Topic maps» saisit les sujets principaux des documents et les relations entre les sujets, d'une manière indépendante de son implémentation. Les concepts clés dans les «Topic maps » sont les thèmes, les associations, et les occurrences. Un thème est une représentation d'un sujet. Les exemples de tels sujets pourraient être «recherche d'information », « indexation des documents électroniques », ou «rapport d'activités ». Les thèmes peuvent avoir des noms, des occurrences c.-à-d., les informations qui sont considérées pertinentes par rapport au sujet. Les thèmes peuvent participer aux relations (les associations), dans lesquelles elles jouent des rôles comme membres. Ainsi, les thèmes ont trois types de caractéristiques : noms, occurrences et rôles joués comme membres des associations. L'attribution de telles caractéristiques est considérée comme valide dans un contexte. Des «Topic maps » peuvent être fusionnés. La fusion peut avoir lieu à la discrétion de l'utilisateur ou de l'application (au temps d'exécution), ou peut être indiqué par

l'auteur de le «Topic maps » à l'heure de sa création.



*Figure 2.16 : Indexation dans Topic maps*

(Source <http://www.xml.com/pub/a/2002/09/11/topicmaps.html>)

#### **2.2.4.1.4.3. Web Ontology Language (OWL)**

La langue d'ontologie de Web de OWL est conçue à l'usage des applications qui doivent traiter la teneur de l'information au lieu de l'information de présentation juste aux humains. Le OWL actuellement est adopté par la communauté du Web, par exemple en décrivant des personnes dans FOAF (Friend of a Friend), qui est fondé sur une ontologie exprimée en OWL.

L'OWL facilite une plus grande traduction de contenu par la machine Web soutenu par le schéma de XML, de RDF, et de RDF (RDF-S) en fournissant le vocabulaire additionnel avec une sémantique formelle. Pour l'OWL se déployant sur le Web, un codage tout à fait bavard de RDF est employé. Il existe des

syntaxes plus concises, que sont plus faciles à lire par des humains, mais ne s'adaptent pas dans le cadre proposé pour l'enchaînement sémantique.

L'OWL a trois sous-langages de plus en plus expressifs : OWL Lite, OWL DL, et OWL complètement. Ces sous-langages peuvent être employés pour créer les ontologies qui sont légèrement limitées dans leur expressivité mais aussi pour **avoir** des propriétés informatiques.

## **2.2.4.2. Modèles fondés sur les modes de création d'annotation**

### **2.2.4.2.1. Les annotations automatiques**

Nous notons que l'annotation automatique est une annotation de type action et ce n'est pas de l'annotation de type objet.

Quand nous parlons de l'annotation automatique, nous entendons que c'est une action de création des annotations par la machine qui fonctionne dans un mode autonome, essentiellement sans l'interférence de l'homme.

La plupart des outils considérés comme des systèmes d'annotation automatiques ne sont que des systèmes d'information qui aident à classification des documents (cf, définition du document section 1.3). L'identification est faite avec un lien à une base de données des objets similaires ou avec des couleurs spécifiques. Un exemple est le **DOGMA** (Wyman et al, 2004) et l'annotation automatique de musique (Tumbull, 2005). **DOGMA** est un système de Web qui permet l'utilisation des recherches comparatives pour identifier et annoter des gènes dans un génome.

Dans le cas de l'annotation de musique, l'annotation inclut des étiquettes de classe trouvées en classifiant la musique et les mesures musicales à valeurs réelles en utilisant la régression. Les tâches de classification classifient la musique fondée

sur des concepts objectifs tels que l'identification de l'artiste ou l'instrument aussi bien que les concepts subjectifs tels que le genre ou le contenu émotif.

Des exemples de méthodes utilisées dans l'annotation automatique sont le « Support Vector Machines (SVM) », et le « Artificial Neural Network (NN) ». Ceux-ci ont été appliqués pour l'annotation automatique de la musique. La logique derrière le « Support Vector Machines » est de créer un modèle généralisé pour les données dispersées (Schölkopf et al, 1998). Nous notons que la disposition des données dispersées dérivés est fondée sur l'apprentissage de machine. L'automatisation est alors créée et fondée sur cette méthode.

*Un exemple de système d'annotation automatique: **GENETATOR***

Le « Genotator » est un outil permettant l'annotation automatique et la navigation dans les annotations (*automated sequence annotation and annotation browsing*). Dans l'application sur des documents ADN, le système effectue des analyses et présente le résultat en utilisant des codes couleurs pour les classes de document. Genotator fournit une manière d'identifier les régions significatives (par exemple, exons) dans une séquence. Les utilisateurs peuvent interactivement ajouter des annotations personnelles aux sortes d'étiquettes qui les intéressent.

Magenta	NNPP promoter predictions
Red	GenPept hits (using BLASTX): GenPept consists of all the GenBank coding regions translated to amino acids
Orange	EST hits (using BLASTN)
Yellow	Human repeat sequence hits (using BLASTN)
Chartreuse	xpound exon predictions
Green	GeneFinder exon predictions
Turquoise	GRAIL exon predictions
Dark Blue	Genie exon predictions
Purple	GenBank CDS (exons)
Magenta/Red/Orange	Open reading frames ( $\geq 150$ bases), colored by frame

Tableau 2.4 : Les couleurs des annotations en GENOTATOR

source <http://www.fruitfly.org/~nomi/genotator/genotator-paper.html>

#### 2.2.4.2.2. Annotation semi-automatique

Il est difficile d'imaginer des systèmes d'annotation entièrement automatiques. Si les activités d'annotation sont censées interpréter ou évaluer un document, ces actes devraient impliquer la participation humaine. L'interprétation de document dépend de plusieurs facteurs qui sont au delà de la représentation symbolique. Nous n'essayerons pas de présenter ces facteurs dans la représentation informatisée. Ces facteurs d'annotation sont psychologiques, sociaux, religieux, culturels et économiques. En fait, c'est pourquoi nous croyons qu'un acte d'annotation est un facteur non seulement du document mais du créateur d'annotation et du temps. L'interprétation de document dépend du temps, des publics visés et de l'objectif de l'interprétation.

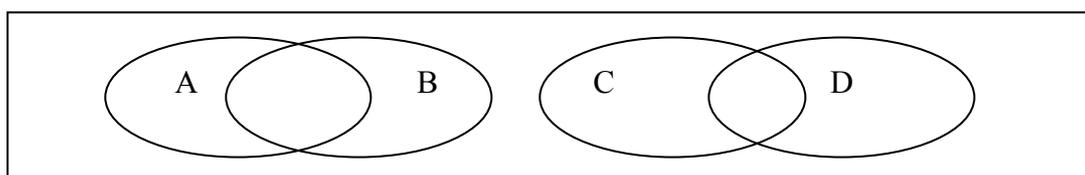
Il est plus facile de faire confiance aux processus d'annotations semi-automatiques où la machine est utilisée pour effectuer des actions répétitives et où les calculs mathématiques sont nécessaires dans la création d'annotation.

Les parties automatisées de système concernent généralement des concepts comme des concepts sémantiques, ontologiques ou linguistiques. La partie non automatisée est laissée au choix de l'utilisateur par rapport à son contexte d'utilisation du résultat et son expérience.

Plusieurs approches ont été utilisées pour associer les significations aux mots. Une des ces méthodes employées est la création d'une base de mots (dictionnaire) de type sémantique ou ontologique. Une autre méthode est d'utiliser et d'appliquer l'algorithme de matrice des mots dans les documents. Un exemple des méthodes utilisées pour analyser le contenu de l'information dans l'annotation semi automatique est le modèle « Latent Space Model ».

#### *Latent Space Models et Probabilistic Latent Space Models*

Le «Latent space model» (LSA) est un algorithme fondé sur l'algèbre linéaire. L'hypothèse du LSA est qu'il existe des caractéristiques communes entre deux acteurs, par exemple entre l'acteur A et B, ou C et D. L'hypothèse avance le fait qu'il n'y a pas d'interaction entre A et C, A et D, B et C et B et D, illustré pas le schéma suivant (Shortreed et al, 2005).



*Figure 2.16 : Interactions entre les acteurs dans LSA*

L'objectif est d'établir le lien entre la ligne de relation qui existe entre les différents groupes. Dans son application à l'analyse des documents, LSA décompose le

terme par-document en trois matrices par une «Singular Value Decomposition» (SVD) tronquée. Dans ce cas, les mots sont récupérés en groupes différents, alors LSA est employé pour établir les liens entre les groupes. PLSA (Probabilistic LSA) est une amélioration de LSA qui permet l'annotation fondée sur le processus de calcul de probabilités et spécifiquement, sur la distribution a posteriori des termes de vocabulaire.

### **2.2.4.2.3. Annotation manuelle**

L'annotation manuelle est l'utilisation des motivations et des expériences humaines pour faire l'annotation. Il est difficile de classer ce genre d'annotation parce que l'annotation humaine peut prendre différentes formes applicables à toute sorte de documents (cf. section 1.3.5 sur notre définition d'un document). L'annotation humaine peut être écrite, graphique ou vocale. Il est même possible de faire des annotations sous une forme codée compréhensible pour les invités. Ceci rend l'annotation manuelle unique par rapport à d'autres formes d'annotations. La signification attribuée à une annotation humaine peut changer selon le contexte.

L'annotation manuelle est la méthode la plus répandue pour la création des annotations en raison de la participation de l'homme dans le processus. Elle est également la plus subjective à cause des changements de l'action et du raisonnement humain dans le temps. Il est psychologiquement et socialement instable. Le centre d'application de cette étude est lié aux changements de l'annotation en raison du changement de l'état de l'annotateur avec le temps. Nous croyons que l'analyse des changements dans les annotations peut être vue comme l'indication de la personnalité impliquée dans l'annotation.

### **2.2.4.3. Annotation dans les technologies du Web**

Une annotation d'un document Web est une annotation en ligne liée à une ressource du Web (typiquement une page Web). A travers l'utilisation d'un système d'annotation du Web, un utilisateur peut ajouter des commentaires sans

modifier le document source. L'annotation sur les ressources Web est une couche sur la ressource existante. N'importe quelle couche d'annotation peut être rendue visible aux utilisateurs qui partagent le même système d'annotation et cachée aux autres.

#### **2.2.4.3.1. Les dispositions du Web**

Dans le cas de l'annotation fondée sur un serveur, l'annotation est effectuée sur le serveur du Web avant qu'elle ne soit envoyée au client. Il n'y a pas beaucoup de systèmes connus dans cette catégorie car cette approche exige des documents source d'être prétraités à l'avance pour avoir les crochets et nœuds nécessaires pour les annotations, et par conséquent, ne peut être généralisée pour annoter les documents sur le Web. CoNote (Davis, 1995) et notes virtuelles (Geyer-Schulz, 1999) sont deux exemples dans cette catégorie.

Le modèle CoNote est fondé sur le concept d'un document de groupe, qui est un ensemble de personnes qui partagent une collection de documents. Chaque personne dans un groupe a un certain rôle en ce qui concerne les documents dans ce groupe avec les droits d'accès progressifs. Les rôles possibles sont «observateur», «lecteur», «utilisateur» et «auteur». Un observateur peut afficher un document annoté, mais ne peut pas voir les annotations là-dessus. Un lecteur peut voir les annotations, mais ne peut pas ajouter quoi que ce soit sur le document. Un utilisateur peut lire et ajouter, et un auteur peut lire, ajouter, et supprimer les annotations.

Une personne peut avoir un rôle différent dans des groupes différents. Par exemple une personne pourrait avoir le rôle d'un auteur dans un groupe et être un lecteur dans un autre. Les groupes peuvent également définir un rôle par «défaut» qui fournit un certain niveau d'accès aux utilisateurs externes. S'il n'y a aucun rôle par défaut, alors les utilisateurs externes n'ont aucune autorisation d'accès aux documents ou aux annotations dans le groupe.

Bien que les annotations apparaissent à l'utilisateur en ligne dans le document, elles sont stockées dans une base de données séparée, et sont intégrées dans le document quand il est fourni à l'utilisateur. Ainsi le document lui-même n'est pas changé par les annotations. Ceci signifie que le même document peut apparaître dans plus d'un groupe de document, ou des groupes d'annotations peuvent apparaître dans un document.

Un groupe de document se compose d'un ensemble d'utilisateurs, chacun avec un rôle spécifié et d'un ensemble de documents. L'ensemble des utilisateurs est spécifié par une « **définition de rôles** ». La définition de rôle énumère chaque utilisateur et le rôle qu'il prend pour ce groupe de document. Les documents dans la collection sont spécifiés par une « **définition de document** », qui indique où les fichiers sont stockés. La définition du document peut énumérer les nœuds d'annotation dans le document.

Chaque document dans CoNote a un nom unique, le docid, qui est le nom composé du groupe du document et de la définition de document (Davis et Huttenlocher, 1995). De plus, chaque point d'annotation a un nom unique dans le document. L'auteur assigne ces noms en indiquant les emplacements potentiels pour l'annotation.

#### **2.2.4.3.2. Présentation d'information**

Dans le cas de ce genre de systèmes d'annotations, ils sont faits pour présenter l'information annotée dans un format spécifique. Le formatage est important en raison des contraintes fonctionnelles. Par exemple, la présentation d'annotation dans le domaine médical ne sera pas nécessairement la même qu'une représentation dans les systèmes aérospatiaux. Beaucoup de systèmes d'annotations développés pour des utilisations spécifiques dans des domaines spécifiques peuvent être classifiés dans cette catégorie. Nous notons que la majeure partie de ces types de système d'annotation repose sur un des modèles

tels que l'ontologie, la sémantique ou sur le modèle linguistique. Nous voulons seulement souligner que la présentation de l'annotation est d'une importance particulière dans la conception de son modèle.

### Exemple

#### GOA (Gene Ontology Annotation)

Le GOA est un projet géré par l'institut européen de Bioinformatique qui vise à fournir des attributions des produits de gène à la ressource de Gene Ontology (GO) sous la forme d'annotations. Le but de «Gene Ontology Consortium» est de produire un vocabulaire contrôlé dynamiquement et qui peut être appliqué à toutes les organisations qui travaillent dans le domaine de la biotechnologie.

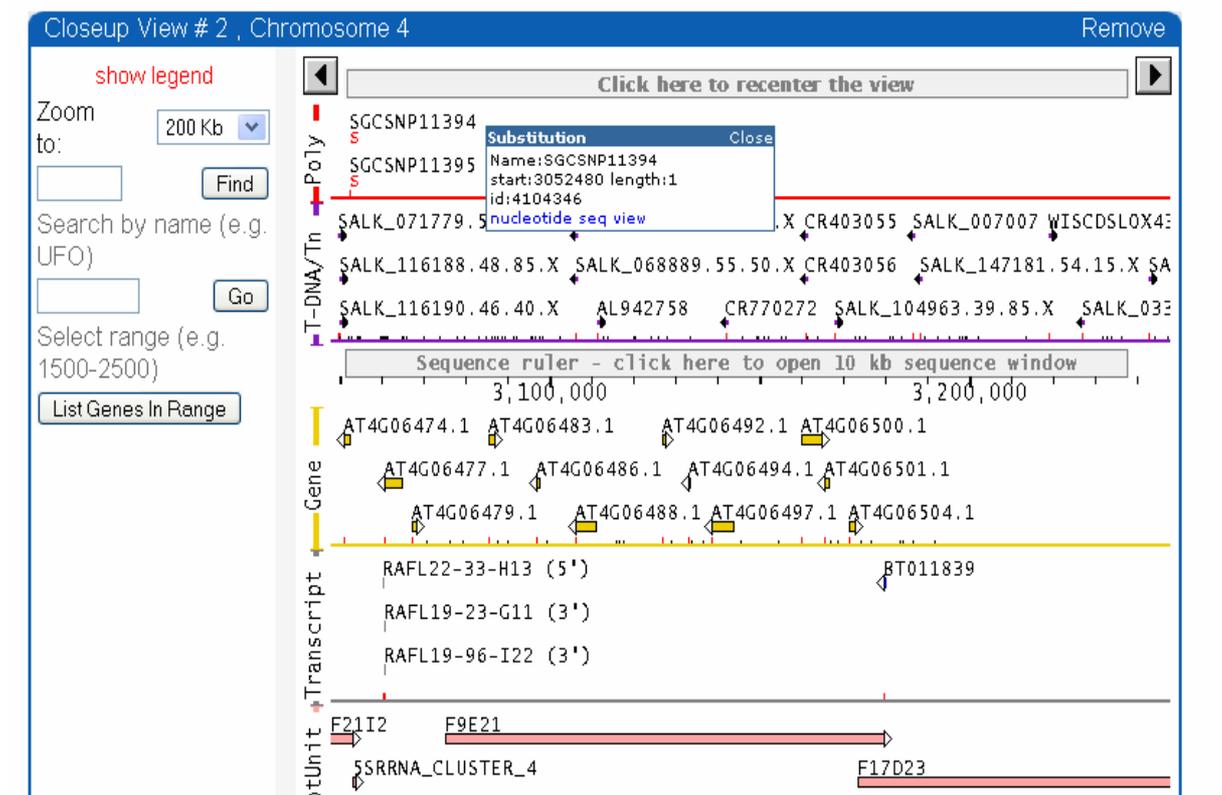


Figure 2.17 : Un exemple d'annotation basée sur le format (GOA)

Quelques outils d'annotation de domaine-spécifiques sont décrits dans la table en dessous:

Domaine	Nom de projet	Site internet
Génétique	GOAT : Gene Ontology Annotation Tool	<a href="http://goat.man.ac.uk/">http://goat.man.ac.uk/</a>
	Genetator	<a href="http://www.fruitfly.org/~nomi/genotator/">http://www.fruitfly.org/~nomi/genotator/</a>
Statistiques	SAS Annotation	<a href="http://www.math.yorku.ca/SCS/sssg/boxplot.html">http://www.math.yorku.ca/SCS/sssg/boxplot.html</a>
Médicale	LEADTOOLS	<a href="http://www.leadtools.com/sdk/medical/Medical-Addon-Annotations.htm">http://www.leadtools.com/sdk/medical/Medical-Addon-Annotations.htm</a>
Multimédia	VideoAnnEx	<a href="http://www.research.ibm.com/VideoAnnEx/">http://www.research.ibm.com/VideoAnnEx/</a>
SIG (GIS)	Arc4You	<a href="http://www.wlm.at/A4_Eng/A4anno.htm">http://www.wlm.at/A4_Eng/A4anno.htm</a>
Géologie	GEM	<a href="http://ess.nrcan.gc.ca/pubs/carto/downloads/adding_and_editing_annotation_in_gems.doc">http://ess.nrcan.gc.ca/pubs/carto/downloads/adding_and_editing_annotation_in_gems.doc</a>
Linguistique	ATLAS (Architecture and Tools for Linguistic Analysis Systems)	<a href="http://www.nist.gov/speech/atlas/">http://www.nist.gov/speech/atlas/</a>

*Tableau 2.5 : La table des systèmes d'annotation spécifiques*

### **2.2.4.3.3. Stockage d'information**

Le stockage des informations est un des facteurs importants dans la modélisation des systèmes d'annotations. Comment les annotations sont-elles organisées et comment peut-on les stocker et où on peut-on les stocker. Nous avons trois méthodes utilisées pour les lieux de stockage des annotations : le stockage local et le stockage à distance.

Dans le cas du stockage local, les annotations sont stockées sur la machine d'un utilisateur. Dans le cas du stockage à distance, les annotations sont stockées par le serveur web.

Plusieurs suggestions et développements sont en cours de réalisation pour les stockages des annotations. Parmi les technologies du stockage adoptées au cours des années pour les stockages d'informations on cite : XSL, DBF, TXT et CSV. Ces dernières sont associées avec les bases de données relationnelles. Dans le cas des bases de données liées avec les informations sur le Web, les technologies comme SGML, HTM, RDF, XML, HTML, TXT sont le plus utilisées.

DBF	DBase File. Format de fichier du SGBD dBase. Un des formats reconnus pour les bases de données relationnelles sur micro-ordinateur
CSV	Comma Separated Values. Valeurs séparées par des virgules (et par extension, par un caractère particulier). Il s'agit d'une forme très primitive, mais finalement très robuste, de base de données, où chaque ligne est un enregistrement où les champs sont séparés par un caractère prédéfini.
XLS	eXcel Spreadsheet. feuille de calcul du tableur Excel de Microsoft
HTM	HTM désigne en particulier un fichier contenant un document formaté en HTML (normalement on devrait utiliser HTML comme extension de fichier)
SGML	Standard Generalized Markup Language. Métalangage utilisé pour définir de façon générale des langages définissant des documents hypertextes de toutes sortes, normalisé sous le nom d'ISO 8879. HTML en est un dérivé (très) simplifié, XML aussi (mais en moins simplifié).
RTF	Rich Text Format. Format de fichier strictement textuel (norme ASCII ou ANSI), et contenant des codes de formatage du document. Ces codes sont aussi exprimés en ASCII ou ANSI. Le format RTF est utilisé par Microsoft comme code source des fichiers

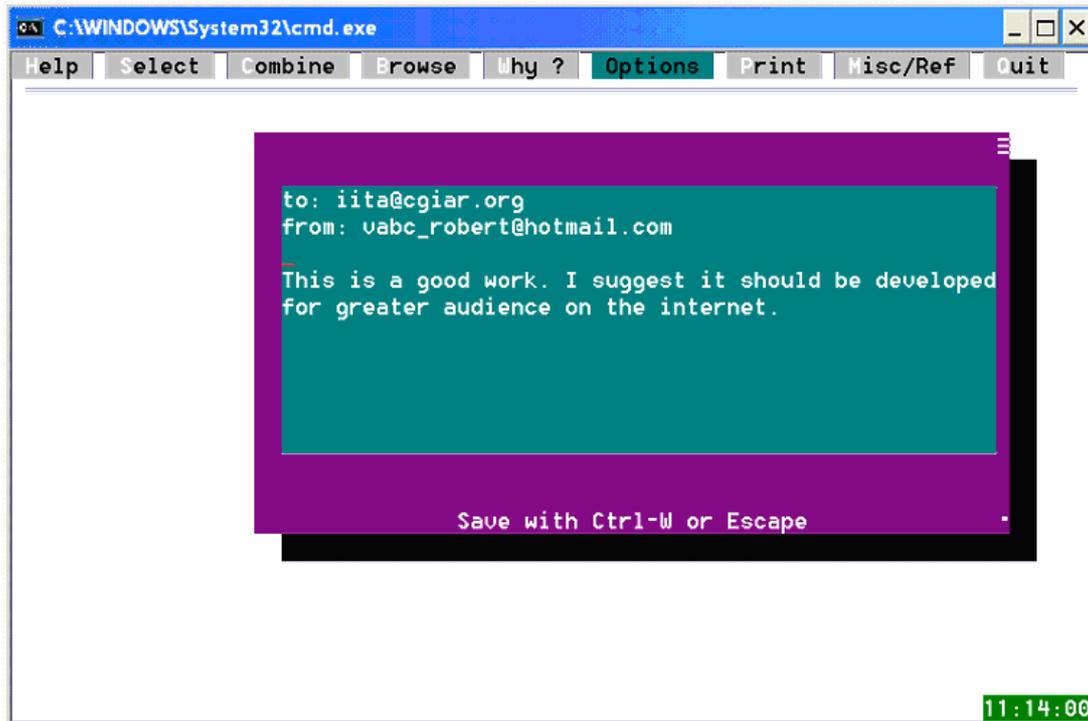
*Tableau 2.6 : Tableau des fichiers dans le stockage des informations*

Les formats utilisés pour le stockage des informations annotées dépendent la plupart du temps des annotations elles mêmes. Les facteurs peuvent inclure: comment accéder aux informations stockées ? Comment transmettre les annotations stockées pour son utilisation ? Quelle sera leur utilisation fondamentale ?

#### **2.2.4.3.4. Partage d'information**

Comme nous l'avons dit plus tôt dans section 1.3.6, l'annotation peut être utilisée pour évaluer ou récapituler l'information. Souvent le but des ces évaluations, ou sommaire, est le partage de l'information. Il y a des outils d'annotations qui ont été développés purement pour le partage de l'information. Quelques exemples de ces outils d'annotation incluent le commentaire dans Microsoft Word, NOTE dans LEXSYS et l'annotation dans la feuille de QuatroPro de Corel. Certains de ces outils ne favorisent pas le partage de l'annotation en temps réel. Ils sont faits de manière à ce que les annotations soient envoyées à un ou plusieurs récepteurs d'information par moyens électroniques. Nous décrivons le système d'annotation dans le logiciel LEXSYS.

LEXSYS (Legume Expert System) est un outil de partage d'information développé par l'institut international d'agriculture tropicale (IITA), Ibadan, Nigeria (Weber et al, 1997). Dans le système, il y a un outil d'annotation intégré nommé «NOTE» sous le menu d'option. L'essence de cette «NOTE» est de permettre aux chercheurs scientifiques dans le domaine agricole de partager les expériences par un envoi des notes ou des observations sur les traits ou sur des espèces spécifiques des plantes légumineuses. Des notes «note.DBF» et «note.FPT » faites par des utilisateurs sont communiquées au coordonnateur du projet par l'intermédiaire de courriel.



*Figure 2.18 : Un exemple d'annotation partagée dans le système d'informations  
LEXSYS*

L'une des utilisations de l'annotation dans ce système permet la collaboration entre les chercheurs dans le domaine d'agriculture. Par exemple, les utilisateurs en Italie peuvent envoyer leurs expériences aux utilisateurs de ce système au Brésil.

#### **2.2.4.4. Problèmes liés aux outils d'annotation sur le Web**

##### **2.2.4.4.1. Consistance des sites Web**

Le problème lié à la consistance des pages Web a été identifié dans (Miles-Board, 2004). L'auteur a remarqué que la plupart des outils d'annotation sur le Web sont essentiels pour la gestion de l'information. En plus, en raison de la nature changeant de la page Web, un des problèmes principaux avec l'annotation du Web est de savoir comment résoudre le problème de consistance des Sites Web ? Quelques systèmes d'annotation, tels que XLibris, supposent que les documents numériques annotés ne changeront jamais. Des annotations sont typiquement placées en utilisant des moyens très simples, tels que des ensembles de caractères ou des positions (x,y).

L'hypothèse est que le document de base n'est jamais modifié ainsi que les annotations. iMarkup par exemple n'exige pas explicitement que les pages doivent rester sans changement, mais travaille mieux quand il n'y a aucun changement de position. iMarkup utilise la localisation géométrique pour enregistrer le positionnement des annotations. Dans le cas où la page Web change, les localisations initiales des annotations pourraient être perdues. CoNote a essayé de compenser les modifications potentielles des pages Web en permettant aux utilisateurs d'annoter seulement des positions prédéfinies par le propriétaire de la page.

#### **2.2.4.4.2. Compatibilité**

Un problème majeur avec les outils d'annotation sur le Web est la question de compatibilité. Il y a plus de vingt types différents de navigateurs du Web. Les outils d'annotation sont conçus et fondés sur un concept et/ou un navigateur spécifique. Pour être capable de circonvenir ce type de problème, Amaya a développé son navigateur propre. Le travail (Vatton et al, 1999) remarquent que, Amaya peut tourner sur le navigateur Netscape. L'outil d'annotation Annotator a été conçu pour le navigateur Netscape. Hylighting tourne sur son propre navigateur (Lebow, 2003). ComMentor tourne seulement sur Mosaic Browser (Roscheisen et al, 1995). Un autre problème est que quelques uns des outils d'annotation ont été développées sur une plate-forme particulière. Par exemple, ComMentor fonctionne sur un navigateur spécifique (Mosaic). L'outil GrAnT suppose qu'il existe un seul espace de travail avec une base de données d'annotations sur la machine d'un utilisateur (Schickler, 1996). Web Annotator a été conçu comme un plugin sur le navigateur Internet Explorer (Reed et John, 2003).

### **2.2.4.4.3. Sécurité**

Un des problèmes des systèmes d'annotation sur le Web est celui de la sécurité. Parce qu'une annotation est souvent sous la forme de document partagé, elle doit être disponible pour tous les utilisateurs.

Nous voyons un cas particulier comme illustration pour le problème de sécurité lié aux annotations sur le Web. ThirdVoice, est un outil d'annotation sur le Web qui a encouragé des utilisateurs et des Webmasters d'Internet à partager des annotations sur les sites Web en liberté (Kumar, 2001). Le problème n'est pas nécessairement que le système ThirdVoice est envahissant, mais qu'il permet à n'importe qui d'ajouter des annotations - tels que le Spam - à un site Web sans permission du propriétaire de ce site (Carter et al, 2004).

Le ThirdVoice est une plugin de navigateur de Web sous Windows qui permet aux utilisateurs de choisir des textes sur une Page Web et d'y attacher des annotations, sous forme de note instantanée. Les autres utilisateurs de ThirdVoice en visitant le Site Web voient une ou plusieurs icônes sur la page Web qui signalent la présence des annotations. Le ThirdVoice ouvre un Site à l'attaque externe.

### **2.2.4.4.4. Coût**

Nous avons souligné que plusieurs outils d'annotations exigent des navigateurs spéciaux ou des systèmes d'exploitation spéciaux pour fonctionner. Les navigateurs spéciaux ou les systèmes d'exploitation spéciaux des coûts supplémentaires. En dehors du système d'exploitation, le coût de stockage des annotations dans les archives doit être intégré. D'après un rapport par Pogolinux<sup>24</sup> en 2006, le coût minimum par giga octets des outils de stockage StorageWare est de \$3.67. Les annotations sur un document augmentent avec le temps. Il est

---

<sup>24</sup> <http://www.pogolinux.com/storage/costpergigabyte.html> 13/08/2006

possible que l'amélioration des technologies de stockage diminue le coût de stockage, mais il est aussi vrai que la taille d'informations à stocker également augmente.

Après avoir présenté les modèles d'annotation ainsi que les problèmes identifiables, nous présentons les outils d'annotation existants dans les sections suivantes.

## 2.2.5. Les outils d'annotations spécifiques et leurs spécificités

### 2.2.5.1. Grants

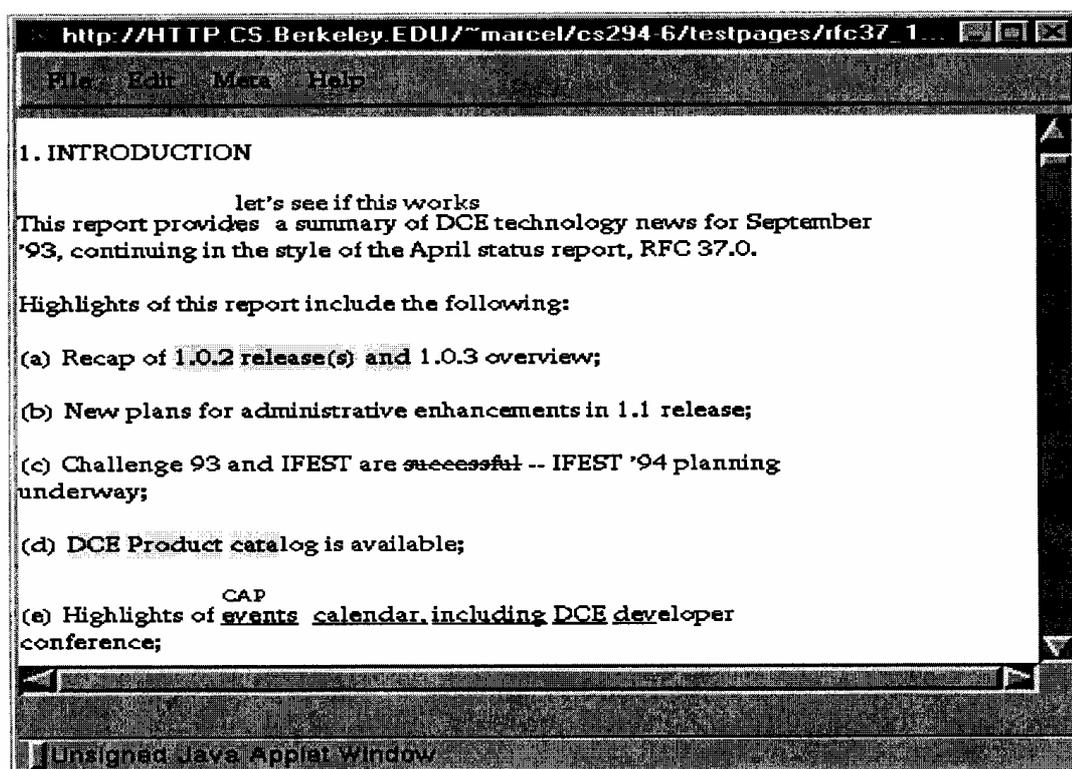


Figure 2.19 : Une fenêtre de création d'annotation par GrAnt

Le GrAnt (Group Annotation Transducer) est un système d'annotation conçu pour être situé sur un serveur proxy. C'est un système d'annotation des pages Web personnelles (Kornacker et Gilstrap, 1997). Un module du système interroge un serveur de meta-donnée pour les informations liées aux annotations concernant le

document HTML concerné. Ensuite le module fusionne les annotations avec la page Web avant d'envoyer le résultat au client de l'application. Le GrAnt permet des annotations en groupe.

L'annotation est enregistrée par des formulaires HTML et les résultats d'annotation sont retournés au client comme des ajouts aux pages HTML du document source. Bien que GrAnt soutienne l'attache des annotations à une zone de texte, le texte en question doit être écrit par l'utilisateur dans le champ approprié sous le formulaire de création d'annotation.

#### **2.2.5.2. One-Thousand Words (OTW)**

Le One-Thousand Word (OTW) est un système pour créer, enrichir et partager des documents sur le Web (Parente et Bishop, 2003). Le logiciel OTW permet à des utilisateurs d'annoter des images numériques trouvées sur les pages Web. En consultant une page Web, l'utilisateur peut interroger les informations primaires (images numériques) et leurs annotations. L'utilisateur peut écouter les annotations enregistrées sous forme de son. La page Web contenant les images n'est pas changée ou remplacée suite à l'annotation. Un module séparé recouvre les annotations et les images annotées et les affiche sur la page Web. Cet outil est particulièrement adapté aux mal-voyants.

Des images à annoter peuvent être chargées à partir d'un site Internet sur son éditeur intégré. Une fois qu'une image est ouverte, l'utilisateur définit une nouvelle annotation. Chaque annotation se compose d'un certain nombre de couches d'informations. Des arrangements tels que la voix à employer pour la synthèse texte-voix peuvent être configurés dans une couche. Les couches dans une annotation indiquée se composent d'un certain nombre de régions qui définissent des « hotspots » sur une Image. Ces « hotspots » prennent la forme de régions géométriques dessinées par l'utilisateur, ou des régions fondées sur des couleurs de Pixel. Des fichiers d'informations et de sons peuvent être associés à chaque région pour être affichés et joués respectivement quand l'utilisateur

clique dessus ou déplace le curseur sur le « hotspot ».

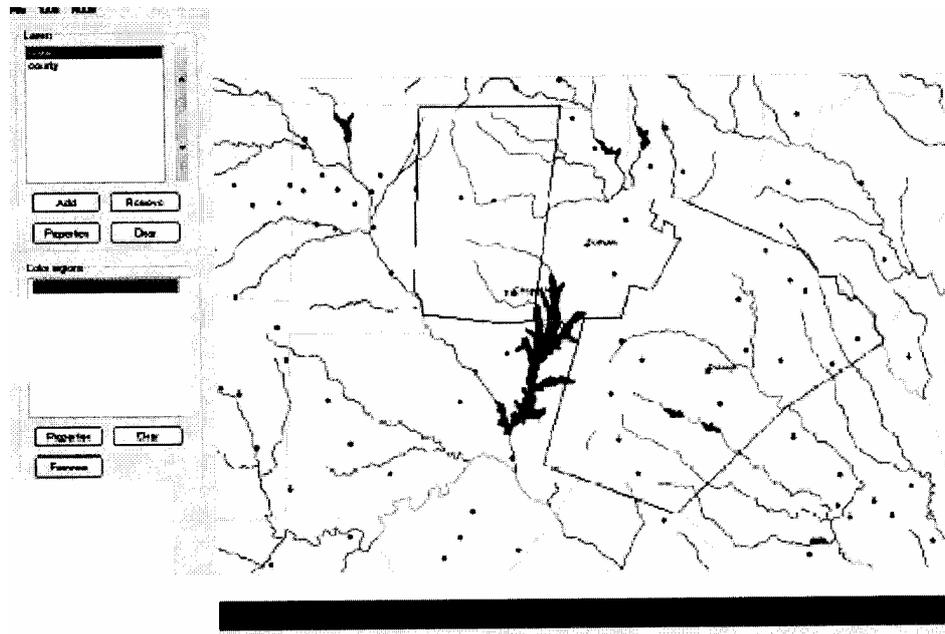


Figure 2.20 : Copie d'écran d'annotation d'une image sur OTW

Quand un utilisateur termine la création d'une annotation, il peut choisir de le sauvegarder localement ou de la télécharger dans un dépôt sur internet. Les régions définies par l'utilisateur sont écrites dans un fichier XML qui est alors sauvegardé avec les fichiers de support dans les archives.

### 2.2.5.3. AMAYA

Amaya a été développé sur le «framework » d'ANOTEA. Il a été développé pour être un navigateur de Web et en même temps un outil d'annotation. Nous ne sommes pas concernés par son efficacité comme outil de navigation. Il offre une possibilité pour l'annotation personnalisée des pages Web. Quelques versions d'AMAYA existent comme plugins pour des navigateurs spécifiques. Le navigateur d'AMAYA (ou les connexions pour les navigateurs d'Internet Explorer et de Mozilla) peut être configuré pour communiquer avec un serveur d'annotation donné. Le stockage peut être à distance ou local. Ce n'est pas possible d'utiliser deux

serveurs d'archivage simultanément. Dans les anciennes versions d'AMAYA, la référence du serveur d'archivage est indiquée dans le fichier de configuration d'AMAYA. Lors de l'affichage par un navigateur, les serveurs d'archivages sont interrogés pour les annotations qui ont été faites sur le document affiché. Les icônes d'annotation sont insérées dans le document pour montrer le texte annoté. Chaque annotation a son auteur, type et emplacement sur l'archivage. Des documents fortement annotés peuvent être filtrés selon les mêmes critères comme le nom d'auteur, le type d'annotation et le parcours d'archivage d'annotation. Un mécanisme de filtrage plus puissant est disponible et permet aux experts d'interroger l'archive, par exemple pour demander des annotations d'un type particulier, faite à une date donnée.

Bien que nous ne soyons pas concernés par son efficacité en tant que navigateur, un des problèmes principaux avec AMAYA est qu'il n'est pas capable de manipuler les pages Web avec des cadres. Par exemple, dans la figure suivante, la page Web LORIA, affiché par AMAYA est tordue parce qu'elle contient les cadres. AMAYA peut afficher une seule page Web et non pas l'ensemble des cadres figurant dans un site Web.

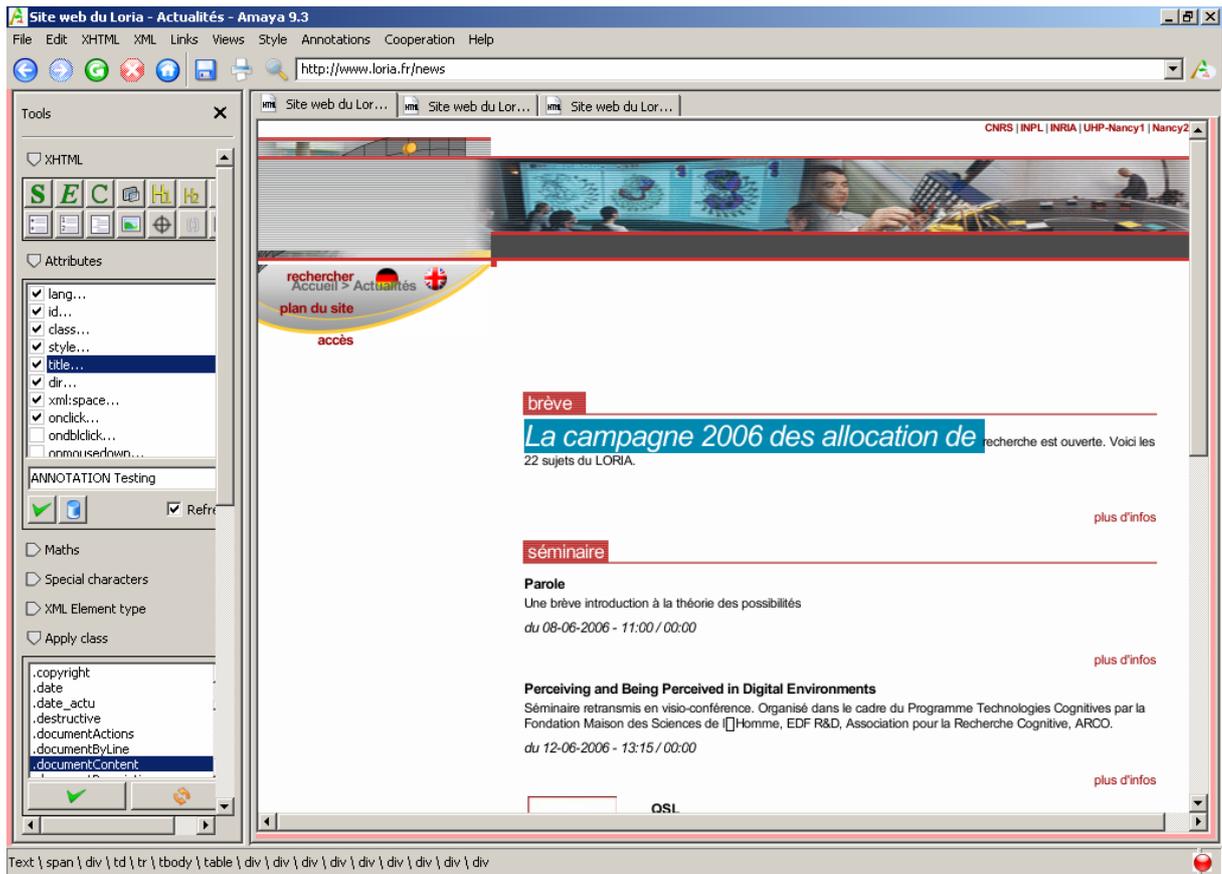


Figure 2:21 : Interface of AMAYA 9.3 sur [www.loria.fr/news](http://www.loria.fr/news) sans les cadres

#### 2.2.5.4. L'annotation GeoMedia

La recherche des documents multimédias et la représentation du contenu multimédia sont le défi des groupes de recherche de National Institute of Informatics, Japan<sup>25</sup>.

Le défi porte sur la difficulté d'annotation des documents multimédia contenant du texte, des images, de la vidéo. Ce sont les problèmes fondamentaux qui les ont poussés à utiliser un système d'annotation multimédia qui intègre les documents et les annotations dispersés géographiquement.

L'objectif de cette recherche est fondé sur le fait que des facteurs géographiques peuvent influencer les activités d'annotation. En dehors des facteurs géographiques,

<sup>25</sup> <http://research.nii.ac.jp/mma-req/description/need.htm>

la perspective GeoMedia pense que l'âge des annotateurs est un facteur dans ce genre d'annotation.

Le mot « geomeia » a une utilisation très large. L'armée canadienne a utilisé le mot comme méthode pour établir une base de données fédérée à travers plusieurs localisation afin de mettre en application ce qu'elle a appelé le « Spatial Information Management System» (Cobley, 2005).

### 2.2.5.5. CoNote

Le CoNote<sup>26</sup> est un système d'annotation développé à l'université de Cornell, EtatsUnis, qui est fondé sur l'HyperNews et écrit dans le langage Perl (Davis, 1994) (Ovsiannikov et al, 1999). Dans ce système, le modèle d'annotations partagées fournit le forum pour les communications entre les étudiants et les professeurs. Le CoNote a été évalué comme le meilleur forum de discussion par rapport aux autres médias électroniques tels que des newsgroups ou les blogs (Davis and Huttenlocher, 1995). La différence principale est que les documents annotés fournissent un contexte pour les discussions de groupe (Heck, 1999).



**Assignment 2**

- ◆ [Nice Page, Ben Adams\\* Date: Thu Sep 25 10:38:36 1997](#)
  - ◇ [Thanks, Chris Smith Date: Thu Sep 25 16:32:42 1997](#)
- ◆ [Comments for Assignment 2 Mike Martin Date: Thu Sep 25 19:51:56 1997](#)
  - ◇ [Great job David James Date: Fri Sep 26 02:10:33 1997](#)
  - ◇ [I got it to work . . . sort of Chris Smith Date: Fri Sept 26 13:34:46 1997](#)
  - ◇ [Thanks guys Chris Smith Fri Sept 26 09:58:08 1997](#)
- ◆ [Mr Smith's David Chi Fri Sept 26 10:51:39 1997](#)
- ◆ [Comment Carolyn J Small Date: Fri Sep 26 15:31:29 1997](#)
  - ◇ [Just a test Brian Boyd Date: Fri Sep 26 17:14:37 1997](#)
- ◆ [Very nice job Mike Mattel Date: Sun Sept 28 22:49:35 1997](#)
- ◆ [Grading Comments Telly Papas Date: Sun Oct 5 19:17:26 1997](#)

**Add annotation**

Figure 2 :22 : Copie d'ecran de CoNote

Nous présentons dans les sections suivantes la manière dont les annotations sont

<sup>26</sup> ([http://www .cs.cornell.edu/home/ dph/annotation/annotations.html](http://www.cs.cornell.edu/home/dph/annotation/annotations.html))

intégrées dans les documents.

### **2.2.6. L'intégration des annotations dans les documents**

Une annotation est forcément liée au document d'origine et souvent se trouve attachée au document. Dans ce cas, l'environnement dans lequel une annotation est liée au document influencera sa perception. Quand nous parlons de l'environnement d'une annotation, nous parlons de section et de contexte de document qui entoure le document et l'annotation. Ceux-ci peuvent être considérés comme «*l'écologie de l'annotation* ».

#### **2.2.6.1. Les emplacements des annotations dans un document**

Rappelons qu'une annotation peut se présenter sur diverses formes - marque sur un morceau de texte, commentaire, icône, etc. Le type et la taille de l'annotation adoptée peuvent influencer sa localisation sur le document. Un nuancier sur un texte sera naturellement attaché au-dessus d'une expression ou une phrase. Un surlignage est normalement situé au dessous d'une expression ou d'une phrase. La norme d'annotation est que des commentaires sont attachés dans la marge des documents.

Quand la forme d'une des annotations ne suit pas les normes (la convention), une interprétation différente peut être donnée par un lecteur de cette annotation.

#### **2.2.6.2. Les annotations non-intégrées sur les documents**

Nous avons montré qu'une annotation est liée au document source. Le lien est souvent sous la forme d'ancre. Deux types d'ancres peuvent être identifiés. Il y a les documents avec les ancres visibles, liées à l'annotation et ceux qui n'ont pas de liens visibles.

### **2.2.6.2.1. L'ancre visible**

Les ancras ne sont pas des annotations mais elles pointent sur l'objet du document annoté. Des ancras sont données sous la forme de flèches, de puces numérotées, des puces symboliques et de soulignés. Elles peuvent également être sous la forme d'ombres ou de signes sur les marges. Quand des puces numérotées, des puces symboliques ou des signes sont utilisés comme ancras d'annotations, ils peuvent être utilisés avec des légendes situées quelque part dans le document, sans lien direct avec la partie du document annotée. Dans le cas de surlignage, l'ancre pointe directement sur l'annotation dans le même document. Pour des documents non textuels, l'annotation est habituellement située sur l'objet. Dans le cas d'annotations sur le Web, les annotations sont faites à travers des ancras pour référencer les annotations sur le document.

### **2.2.6.2.2. Ancras non-visibles**

Certaines annotations sont physiquement séparées du document source. C'est souvent le cas quand les annotateurs veulent préserver le document source ou parce qu'ils ne souhaitent pas partager leurs annotations avec d'autres utilisateurs du document. Quand la taille d'annotation est relativement grande par rapport au document source, l'annotation peut être dissociée. Dans le cas où une ancre n'est pas rendue visible par l'annotateur. Elle devient très difficile de comprendre la signification et le but de l'annotation. L'ancre est là pour identifier la partie du document à laquelle l'annotation est associée. L'annotation sans ancre visible est utilisée souvent pour des annotations à usage personnelle. Dans ce cas, c'est seulement l'annotateur qui peut interpréter exactement la signification de celle-ci.

Nous revenons, dans la section suivante, sur l'une de nos problématiques qui concerne l'annotation dans le processus d'IE.

## **2.3. Utilisation et intégration d'annotation dans le processus d'IE**

Nous avons jusqu'ici considéré l'annotation dans la perspective de sa conception, de sa modélisation et de sa création, notre objectif dans ce travail est de savoir comment utiliser les annotations dans le processus d'IE.

### **2.3.1. Un regard sur annotation par son utilisation**

Si l'utilisateur veut se souvenir d'un document qu'il a lu, il donne une «trace» de manière simple (par exemple l'utilisateur peut souligner, marquer, etc.) ; la marque est apposée pendant le processus d'annotation et l'objet «annotation» est sauvegardé dans un dépôt d'annotation. Nous donnons dans les sections suivantes des exemples d'utilisations possibles des annotations.

Une annotation peut être utilisée pour la classification des documents ou pour la structuration des contenus des documents.

#### **2.3.1.1. Annotation pour la classification**

L'annotation de type classification est utilisée pour classer les documents comme des entités. L'objectif est de créer un classement des documents existants en se servant des annotations.

Dans les bibliothèques, les classifications sont faites en respectant des normes souvent prédéfinies. La classification fondée sur les annotations est faite selon les critères exprimés dans les annotations. Dans ce type d'utilisation, la frontière entre l'annotation et l'indexation est réduite.

L'annotation de type « classification » est très utilisée dans le domaine des sciences

de la vie. Par exemple, l'outil «AutoFACT» permet six classes de classifications par l'annotation pour séquencer les données génomiques [Liisa, 2005]. Un autre exemple est l'outil d'annotation automatique développé pour classer des publications sur le site web de PubMeds MedLine [Sittichai, 2004].

Furlest un autre outil pour classifier les informations sur internet. Le noyau de Furl permet de stocker le contenu disponible sur Internet et sauvegarder les documents à tout moment (par exemple, les articles, descriptions de produit, site, e-commerce etc.).

D'autres exemples incluent Plpage un service gratuit qui permet la gestion de favoris en ligne, l'importation et l'accès sur Internet à toutes les préférences (favoris), bookmarks, signets et de les partager avec d'autres utilisateurs. L'objectif principal de Plpage est pour la classification d'information.

### **2.3.1.2. Annotation pour la structuration**

Les annotations de structuration des documents concerne les différents éléments (la structure) des documents. La plupart des outils d'annotation sur l'Internet sont de ce type. Les exemples incluent, «annotation engine », «Hylighting », «Amayas» et «CritLink ». L'outil d'annotation tel que celui sur Microsoft Word est du type structuration.

### **2.3.1.3. Le public dans l'utilisation**

#### **2.3.1.3.1. Annotation pour un groupe d'utilisateur**

Dans un système d'annotation pour un usage public, l'accès peut être restreint ou ouvert. L'annotation peut être modérée ou non-modérée. Pour l'annotation modérée, un but est fixé ou un objectif est défini pour le cadre des annotations.

Dans un espace de travail d'annotation public, les membres dans la collaboration peuvent être invités ou ouvert au public (non contrôlé). Quels sont les critères pour la participation ? Quels sont les éléments à annoter ?

### **2.3.1.3.2. Annotation pour un usage personnel**

L'utilisation privée d'annotation a été un des objectifs principaux du système d'annotation sur le papier. Dans ce cas, l'individu décide de ce qui est à annotation, et comment. Les annotations précédentes qui ont peut être sur le document par les autres annotateurs ne sont pas souvent considérées.

## **2.3.2. Annotation et indexation**

### *Indexation de documents*

*« L'indexation consiste à identifier dans un document certains éléments significatifs qui serviront de clé pour retrouver ce document au sein d'une collection. Ces éléments comprennent le nom de l'auteur, le titre de l'ouvrage, le nom de l'éditeur, la date de publication et l'intitulé du sujet traité »<sup>27</sup>.*

Les annotations peuvent être faites pour indexer les documents. Nous pensons que, l'indexation des documents est un cas spécifique d'annotation. L'indexation des documents est un cas d'annotation qui doit respecter une règle. Les indexations peuvent être faites sur :

- les documents eux-mêmes,
- des parties de documents,
- des groupes de documents.

Par rapport à l'annotation, nous pouvons noter que l'indexation ne donne pas de sens à un document. Elle identifie les éléments principaux d'un document et utilise

---

<sup>27</sup> <http://www.pblconsulting.com/techltech.html> (16/03/2006)

ces éléments pour classer le document.

### **2.3.2.1. Annotation hors du contexte**

Nous devons affirmer qu'un annotateur annote un document selon le contenu d'un document. Une annotation hors de son contexte n'est pas conforme à la définition d'une annotation. Si l'information provenant d'un auteur n'est pas interprétée ou si l'annotation est hors contexte de document, alors il s'agit d'une altération du document, un embellissement, une variation ou un changement de style de l'objet ou une recréation du document.

### **2.3.2.2. Langage d'annotation**

Le langage utilisé pour le processus d'annotation peut être formel ou non. Dans ce contexte, l'importance est la communication entre l'auteur et le lecteur. Parfois ça peut être une forme de convention pré-établie entre ces deux acteurs (auteur, lecteur).

Parmi ces conventions, il s'agit des formes d'expression, des formes de codage, des formes symboliques qui permettent l'échange d'information entre les acteurs. Les règles linguistiques spécifiques ou les règles sémantiques formalisées sont parfois implicites ou explicites pour accomplir cette tâche de communication.

Nous pouvons identifier trois formes ou types de méthodes pour effectuer la communication dans le contexte d'annotation. Les règles suivantes sont possibles.

#### **2.3.2.2.1. Règle graphique**

Il existe plusieurs types de communication graphique qui peuvent être utilisés pour annoter un document. L'ensemble est constitué par des éléments de dessins, des éléments picturaux, des éléments d'estampe et des éléments photographiques. Quelle que soit la forme utilisée pour annoter un document, cette dernière doit

être compréhensible par le lecteur. Le point commun entre l'auteur et le lecteur d'annotation pour l'interprétation de ces types de communication graphique réside dans la capacité d'utiliser les mêmes conventions dans la perception de l'information.

#### **2.3.2.2.2. Règle orale**

L'usage des informations orales appliquées à l'annotation des documents est une autre alternative de communication dans le contexte d'annotation. Notre intérêt s'oriente vers l'emploi de texte électronique comme moyen d'annotation d'un document. Nous donnons en exemple la solution apportée par «*iMarkup28*<sup>28</sup> Solutions Inc» qui a développé et a commercialisé un logiciel (*iMarkup*) permettant l'annotation d'un document par la voix.

#### **2.3.2.2.3. Règle écrite : explicite ou codée**

L'annotation peut prendre une forme explicite ou codée. Quand l'annotation est codée, les individus qui disposent de ces informations doivent déchiffrer avant de lire et d'interpréter le message en question (annotation). La notion de codage et de son décodage en une information compréhensible est une autre forme d'annotation qui peut être réalisée automatiquement. Ainsi il est utile de se référer aux règles de codage/décodage de l'information pour retrouver la forme lisible du message d'origine grâce à diverses techniques (cryptographie, codage/décodage, ...).

### **2.3.3. Fonctionnement d'annotation sur un document**

L'annotation des éléments du document pennet les fonctions suivantes :

- *information, illustration, extension du document* : lorsque le lecteur rédige ses annotations, il devient alors rédacteur d'annotations, il peut utiliser ses annotations pour illustrer ces points de vue,

---

<sup>28</sup> <http://www.imarkup.com/21/09/2004>

- *forum*: en permettant à un ensemble de lecteurs de débattre sur le document,
- *interprétation et désambiguïsation de l'information* : l'objectif majeur de l'annotation est de désambiguïser le document pour un ou plusieurs utilisateurs de document annoté,
- *aide au processus rédactionnel* : en permettant d'indiquer des consignes de rédaction (corrections, remplacement d'informations...),
- *support de lecture* : (la mise en évidence de passages importants...) permettant l'appropriation du texte par le lecteur, l'annotation est alors le reflet de l'engagement par rapport à un texte d'un lecteur qui le personnalise (trace de lecture) afin de faciliter un futur retour ;
- *évaluation d'un travail* : permettant à un enseignant/tuteur d'évaluer et de corriger/ commenter un devoir d'un étudiant. Ce type d'annotation reflète l'avis de celui qui annote par rapport au travail effectué par le rédacteur du texte annoté,
- Pour *récapituler* les points dans le document. Pour évaluer le contenu de document,
- Pour *résumer* le document,
- Pour *relever* les points dans le document, pour la classification de document
- Pour la *restructuration* du contenu du document,
- Pour *différencier* un document par rapport aux autres document
- Pour *informer* les autres lecteurs du document, Pour répondre à une question liée au document,
- Pour *illustrer* le contenu du document
- Pour *créer une extension* du document,

#### **2.3.4. Granularité d'annotation**

L'annotation est faite sur un document. La question est «qu'est ce qu'un document ?». Il est possible de faire une annotation sur :

- une section de document,
- un document complet,
- une collection de documents.

L'agglomération d'un ensemble de caractères qui forme l'objet d'une annotation est ce qui nous appellerons «la granularité d'annotation ». Nous considérerons quelques types de granularités d'annotation sur un document écrit.

L'annotation de granularité fine est souvent pour corriger les orthographes. Ce genre d'annotation est souvent utilisé par des rédacteurs, les éditeurs et des personnes intéressées par la mise en forme du document.

Notre souci n'est pas principalement sur ce genre d'annotation. Nous pouvons supposer que d'autres recherches peuvent être effectuées sur la relation qui existe entre l'annotation de ce niveau de granularité et l'origine des langues de balisage (Markup Languages).

L'annotation peut être faite sur les mots isolés. Le but est généralement de corriger des erreurs orthographiques dans un document. Ce type d'annotation est très utile sur les documents brouillons. Dans ce cas, le nombre d'annotations est en générale gros, conséquent plusieurs paramètres joueront un rôle dans la détermination du nombre d'annotations. Certains des paramètres incluront (a) le niveau de connaissance approfondie de la langue, (b) l'usage du vocabulaire, (c) le style littéraire et (d) le domaine du document.

#### **2.3.4.1. Les valeurs d'annotation et leurs granularités**

La granularité d'annotation est le degré d'attention donné aux éléments d'un document pendant un processus d'annotation. L'annotation de granularité grossière ne considérera pas la structure ou le contenu du document. L'annotation d'une granularité plus fine donnera l'attention au contenu du document. Une annotation d'une plus haute granularité (grossie) considérera l'annotation des éléments des

articles plus détaillée. Une annotation de granularité grossière peut considérer plusieurs documents dans une localité comme un groupe de documents. Par exemple, nous pouvons voir une bibliothèque comme un seul document à annoter.

#### **2.3.4.2. Granularité du documents et type des annotations**

Nous pouvons identifier les niveaux différents du granularité du document et les mettre en correspondance avec les types d'annotation. Il convient de noter qu'un certain type d'annotation est mieux adapté à certains types de documents selon leur granularité. Par exemple, il est possible de regarder les annotations de type «classifications» sur un ensemble de documents hétérogènes ou les annotations de type « structuration» sur les orthographes du document.

Dans le chapitre suivant, nous présentons la proposition de notre modèle AMIE et ses applications.

### **2.4. Conclusion**

Il était nécessaire de situer ce travail dans les sciences de l'information et la communication. Nous avons pris compte que, le document, l'annotation et l'information peuvent être appliqués à plusieurs domaines des activités humaines. Si c'est le cas, l'application spécifique des termes peut être différente. Après avoirs définir clairement ces terminologies pour notre utilisation spécifique, nous avons défini notre conception spécifique « l'intelligence économique » et quelques autres terminologies relatives. Nous savons que d'autres vues sur ces derniers soumis peuvent être différentes du nôtre.

# Chapitre 3

*" Nos doutes sont des traîtres et nous font perdre, par crainte d'essayer, le bénéfice que souvent nous pourrions gagner.  
William Shakespeare<sup>29</sup>*

*Nos doutes et nos préoccupations (soucis) ont une seule solution : la solution croisée. William K<sup>30</sup>*

*Les doutes et les questions sont des sujets d'annotation<sup>31</sup>*

## **3. Proposition pour la modélisation et la réalisation d'un système d'annotation pour l'IE**

Nous présentons dans ce chapitre notre proposition pour la modélisation et la réalisation des annotations. Le chapitre est composé de trois sections. La première section porte sur les éléments nécessaires pour la modélisation d'un système d'annotation pour l'IE. Le modèle AMIE (Modèle d'annotation pour l'échange d'information, (en anglais Annotation Model for Information Exchange)), que nous proposons, sera présenté dans cette section. La deuxième section porte sur la réalisation d'un système d'annotation fondé sur notre modèle AMIE. La troisième section concerne l'application du modèle dans un système de recherche d'informations contenant des publications scientifiques, afin d'illustrer les fonctionnalités d'un système fondé sur notre modèle.

### **3.1. Schéma général d'un système d'annotation**

L'objectif général des outils d'annotation est d'attacher des informations autres que celles fournies par les auteurs des documents sources. Ces informations ne sont pas stockées sur le document annoté.

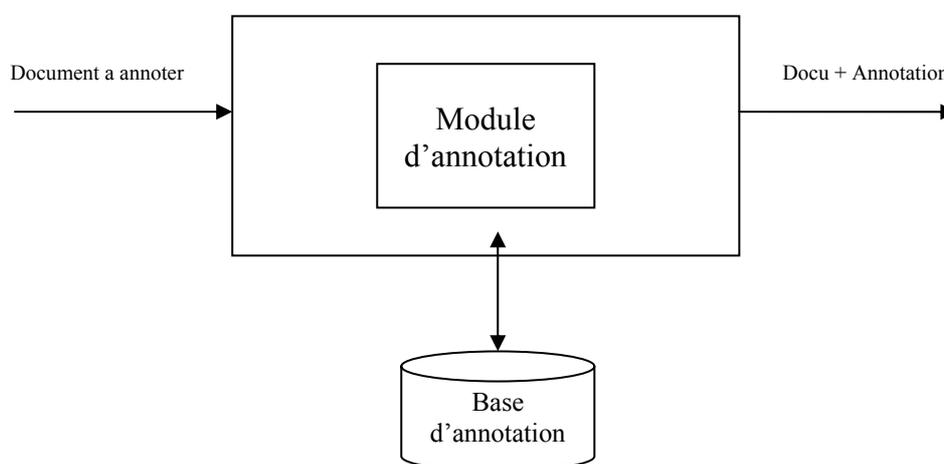
---

<sup>29</sup> Our doubts are traitors, and make us lose the good we oft might win, by fearing to attempt

<sup>30</sup> Our doubts and our cares have one single solution: The Cross solution

<sup>31</sup> Doubts and questions are subject of annotations

Rappelons que le concept d'annotation peut se comprendre par annotation-action et annotation-objet (cf. section 1.3). La plupart des modèles d'annotation peuvent être expliqués à l'aide du diagramme ci-dessous. Le module d'annotation qui met en œuvre les fonctions d'annotation-action reçoit en entrée un document et met les fonctions à la disposition de l'annotateur. Le module d'annotation envoie les annotations-objet à la base d'annotations ainsi que les références des objets du document annoté. Le module d'annotation assure aussi la fonction de visualisation des annotations puis qu'il peut retrouver les annotations associées à un document ainsi qu'aux objets du document avant l'affichage.



*Figure 3.1: schéma général d'un système d'annotation*

Nous avons présenté dans la section 2.4.3. que les annotations sont souvent stockées dans une base d'annotations locale ou à distance. Le stockage d'annotation est important pour nous parce qu'il garantit l'accès aux annotations stockées. Le stockage est aussi un facteur de partage des annotations. Les annotations stockées en local sont généralement destinées à l'usage personnel. Le stockage à distance des annotations favorise son partage, mais la sécurité de la machine et des annotations à distance est un risque.

Nous avons présenté des outils d'annotation existant dans la section 2.2 : GrAnt, One-Thousand Words, Amaya. Ils comportent des fonctionnalités similaires. Ces similarités sont présentées dans le tableau 3.1 et dans l'annexe.

	<b>GrAnt</b>	<b>OneThousandW ord</b>	<b>Amaya</b>	<b>CoNote</b>
<b>Objectif</b>	Usage personnel	Usage personnel	Collaboratif	Educatif
<b>Type de document</b>	Page Web	Image numérique	Page Web	Document Web
<b>Stockage d'annotation</b>	Distance	Local et distance	Local et distance	Local
<b>Affichage</b>	Tous les navigateurs	Tous les navigateurs	Ne supporte pas des « frames »	Spécifique
<b>Système d'exploitation</b>	Non spécifié	Non spécifié	Windows, Unix, MacOS	Spécifique

*Tableau 3.1 : Tableau comparatif des outils existants*

### **3.2. Notre choix des paramètres d'une annotation: Le modèle**

#### **AMIE**

Dans cette étude, notre préoccupation ne porte pas sur la modélisation de l'utilisateur ni sur son profil dans un processus d'annotation, mais plutôt sur le contenu de l'annotation et ses fonctions comme information à valeur ajoutée au document. Néanmoins, les annotations pourront être analysées pour découvrir ou observer les attributs ou les traits caractéristiques des annotateurs. Nous reviendrons sur ce point dans 3.6. Dans notre modèle d'annotation AMIE, nous représentons une annotation-objet par rapport à son contexte, l'annotateur et le document annoté. Ce modèle est présenté sous forme d'une structure arborescente :

- Annotation
  - Référence
  - Objet
  - Type
  - Contexte
    - Objectif
    - Type de collaboration
    - Problème recherche d'information
    - Problème décisionnel
      - L'objet
      - Le signal
      - L'hypothèse
  - Annotateur
    - Identité
  - Document
    - Titre
    - Mots clés
    - Auteurs
    - Résumé

Où :

**Référence** : L'attribut «référence» correspond à l'identifiant de l'annotation

**Objet**: La référence de l'objet du document annoté. Les objets peuvent inclure les sections d'un document, par exemple un mot, une phrase, un paragraphe, une image, etc... .

**Type** : Le «type» d'annotation est la façon de spécifier l'annotation. Nous avons retenu cinq types d'annotation que nous présentons ci-dessous.

- Les annotations de type « **commentaire** » sont indiquées à la marge avec un lien direct sur l'élément annoté.
- Les annotations de type « **symbole** » sont les annotations implicites car l'interprétation est associée au symbole utilisé.
- L'annotation de type « **typographique** » prend des formes soulignées.
- Les annotations de type « **référence** », attache des numéros à des éléments du document correspondant au numéro de référence de l'annotation.

**Contexte :** Le contexte caractérise le contexte de l'annotation. On trouve associés à cet attribut le type de collaboration, l'objectif et le problème de recherche d'information ou le problème décisionnel lié aux activités des annotations.

**Objectif :** «L'objectif» d'une annotation est la raison pour laquelle une annotation est faite. Les objectifs peuvent être **d'évaluer, faire une remarque, relever un point, approuver** un concept, **désapprouver** un concept, **résumer** le document, **classifier** le document, **informer** les lecteurs du document en question, **répondre à une question, illustrer** un concept, **clarifier** des ambiguïtés, **récapituler** des points dans le document, **poser une question** ou une **requête**.

**Type de collaboration :** Le type de collaboration peut être collectif ou individuel.

**Problème de recherche d'information :** La spécification du problème de la recherche d'information

**Problème décisionnel :** Le problème décisionnel lié à la recherche d'information. Nous avons retenu les paramètres de cet attribut proposé par (Bouaka, 2004) pour la représentation d'un problème décisionnel. Ces paramètres sont l'objet, le signal et l'hypothèse.

- l'objet de l'environnement,
- le signal émis par cet objet
- l'hypothèse que nous pouvons déduire de la détection de ce signal.

**L'annotateur** est la personne qui inscrit ses perceptions, ses interprétations, ses évaluations ou ses suggestions sur les éléments du document source par les objectifs indiquées ci-dessus. Dans un contexte d'IE, l'annotateur peut être le décideur, le veilleur ou toute autre personne de l'organisme.

Nous nous sommes inspirés des travaux sur la modélisation d'un problème décisionnel (Bouaka, 2004) pour déterminer les attributs pertinents pour représenter l'annotateur. Ces paramètres sont l'identité de l'annotateur (présentée ci-dessous) ainsi que ses expériences par rapport au projet traité.

**Identité** : L'identité d'un annotateur correspond aux attributs nécessaires pour l'identifier. Les attributs que nous avons retenus sont:

- Nom,
- Prénom,
- Adresse postale,
- Adresse électronique,
- Pays.

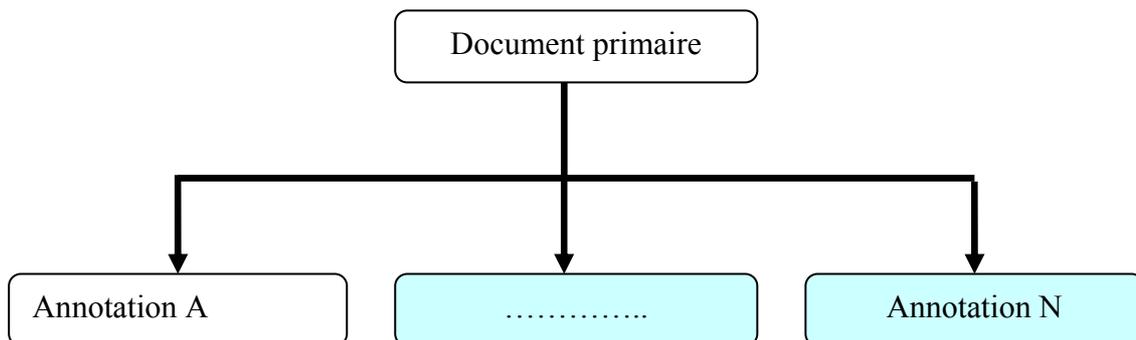
**Document**: L'attribut « document » regroupe des attributs du document annoté. Les paramètres du document que nous avons retenus sont le titre, les mots clés, les auteurs du document et le résumé.

### **3.3. Notre choix des types de document possible pour une annotation**

Comme nous avons indiqué dans la section 1.3, nous nous intéressons aux documents écrits. Nous identifions trois types de document écrits: document primaire, document secondaire, et annotation en tant que document.

#### **L'annotation sur un document primaire**

Un document primaire est pour nous l'œuvre directe d'un auteur, par exemple des livres et articles. Pour le travail d'annotation, il faut que les versions électroniques de ces documents soient accessibles directement.



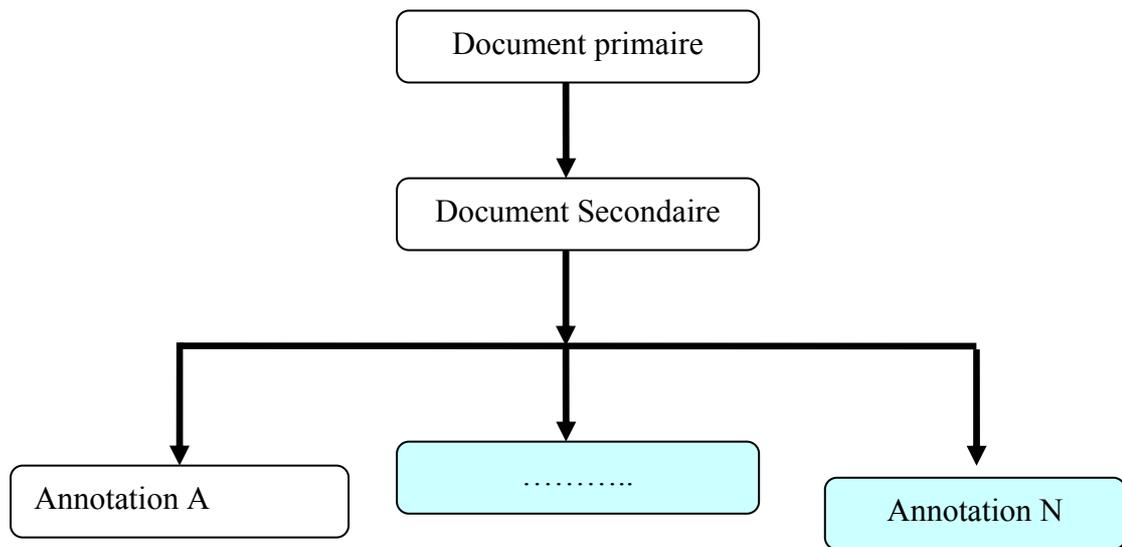
*Figure 3.1 : Annotation sur un document primaire*

Les annotations portent directement sur ces documents et le lien annotation document source est direct. C'est ce type d'annotation qui est le plus souvent mis en œuvre.

### **L'annotation sur un document secondaire**

Nous considérons un document secondaire comme un document dérivé du document primaire sans modification ou interprétation de leur contenu, par exemple les index, des extraits des documents.

Nous considérons que les annotations sur les documents secondaires peuvent être transposées sur le document primaire. La différence principale entre annotation sur un document primaire et annotation sur un document secondaire concerne le fait que tous les éléments du document primaire ne sont pas disponibles sur le document secondaire. De ce fait, il peut être difficile de comparer des annotations sur des documents primaires avec celles faites sur les documents secondaires.



*Figure 3.2 : Annotation sur un document secondaire*

### **L'annotation sur une annotation existante**

Nous avons montré que l'annotation est un processus continu au chapitre 1. Nous pouvons effectuer des annotations sur des annotations existantes dans un document. Dans ce cas l'annotation initiale est considérée comme un document. Nous rappellerons ces annotations « les annotations secondaires ». La question qui peut se poser est : comment réconcilions-nous la différence entre le document primaire et les documents avec les annotations en cascade?

Pour illustrer ce type d'annotation, rappelons notre exemple de «partage en mode transition» des annotations au chapitre 2, page 87, sur les annotations sur un document dans le domaine de la géographie qui porte sur le sujet d'une éruption volcanique.

Le problème ici concerne la difficulté de synthèse des annotations car elles sont de niveau de profondeur varié. Il serait difficile de comparer deux annotations de niveaux différents, ou même d'un même niveau.

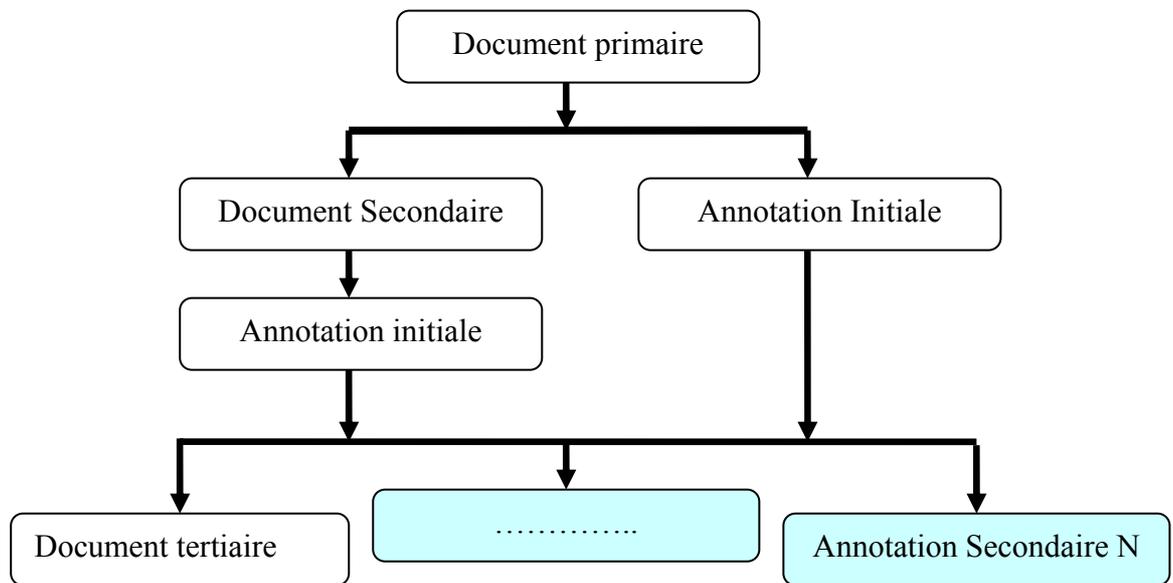


Figure 3.3 : Annotation sur des annotations

### 3.4. La modélisation du système d'annotation

Dans cette section, nous présentons ce que nous avons retenu comme concepts et fondements pour la modélisation d'un système d'annotation pour l'IE.

Nous pensons que le système d'annotation doit intégrer le modèle d'annotation AMIE ainsi que les opérations nécessaires pour effectuer les annotations.

#### 3.4.1. Caractéristiques fonctionnelles du système

Nous pensons qu'un système d'annotation doit permettre au moins les fonctionnalités suivantes :

- **sélection** : sélection des éléments du document à annoter,
- **spécification** : spécifier l'annotation,
- **association** : associer l'annotation à l'élément du document,
- **instanciation** : instancier les autres attributs de l'annotation, c'est-à-dire les attributs du modèle AMIE.

En considérant ce système comme un système d'information intégrant les fonctionnalités d'annotation, nous présentons dans la section suivante la description du système sous forme d'un schéma entité-association et du schéma relationnel correspondant.

### **3.4.2. Schéma Entité-Association et relationnel pour les données du système**

Notre conception d'un système d'annotation fondé sur le modèle AMIE se compose des entités : annotateur, document, lecture, annotation et session.

Un utilisateur du système peut ouvrir une **session** et lire un document d'intérêt. La session qu'il ouvre peut être une session de **lecture** ou d'**annotation**. La lecture d'un **document** peut engendrer un processus d'annotation. Dans le cas où un **utilisateur** n'est pas intéressé par annoter un document, sa session est terminée en enregistrant ses activités. Dans le cas où un utilisateur est intéressé par annoter un document consulté, ses activités d'annotation sont enregistrées. Remarquons que plusieurs annotations peuvent être faites pendant la même session par un utilisateur.

Nous avons choisi de représenter l'acte de lecture pour observer les comportements des utilisateurs en ce qui concerne leurs activités d'annotation, car on peut lire sans annoter.

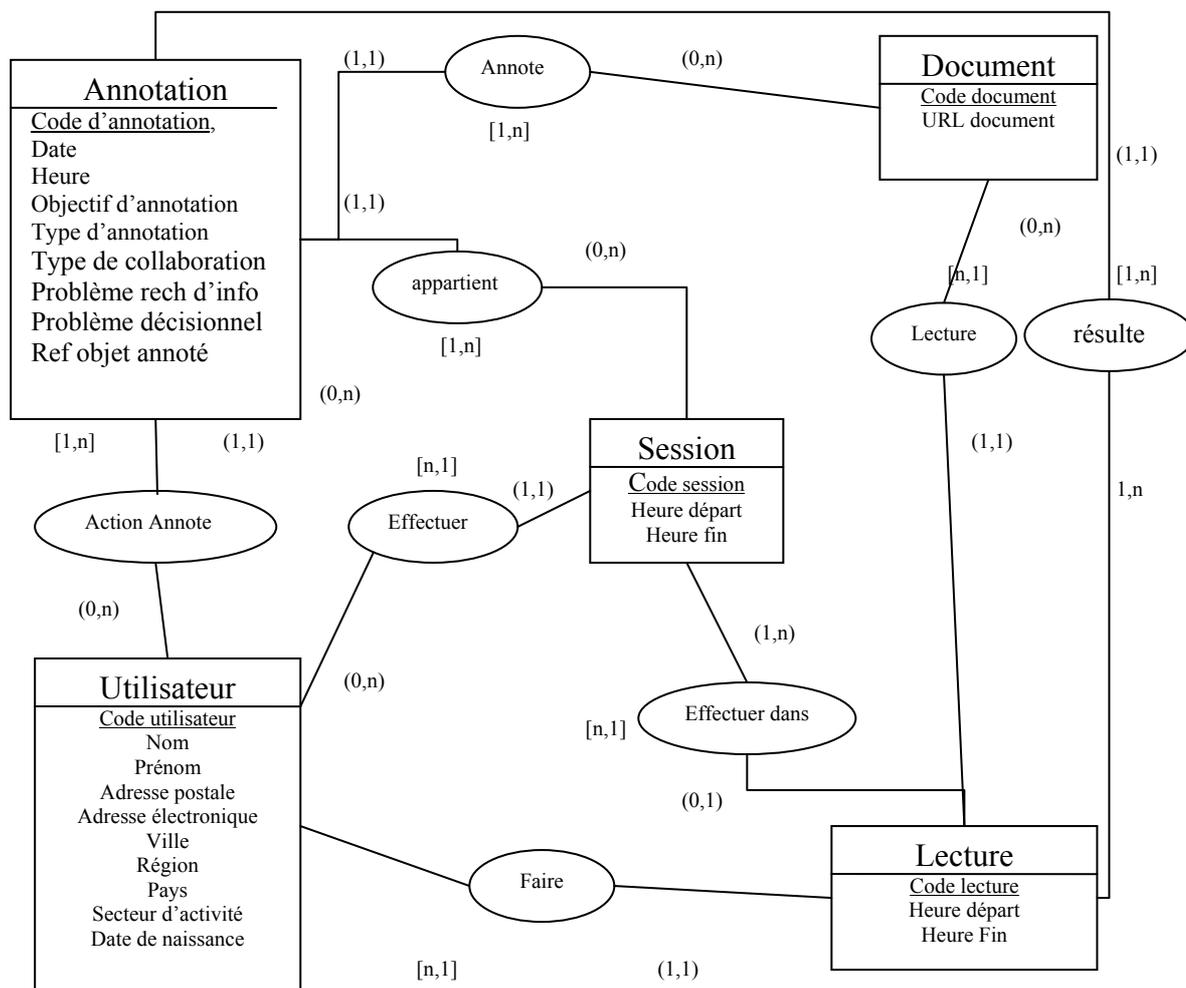


Figure 3.4: Schéma Entité-Association de notre système d'annotation

Le schéma Entité-Association pour le système est présenté par le schéma 3.5. Nous traduisons ce modèle comme ci-dessous:

1. Utilisateur-annotation

Un utilisateur peut faire 0 ou plusieurs annotations, la cardinalité est (0,n)

2. Annotation-utilisateur

Une annotation est faite un *seul* utilisateur, la cardinalité est (1,1)

3. Document-annotation

A un document peut être associé 0 ou plusieurs annotations, la cardinalité est (0, n)

4. Annotation-Document

A une annotation peut être associée 1 seul document, la cardinalité est (1,1)

5. Lecture-Utilisateur

Un utilisateur peut faire 0 ou plusieurs lectures, la cardinalité est (O,n)

6. Utilisateur-Lecture

Une lecture est faite pas un seul utilisateur, la cardinalité est (1,1)

7. Session-Utilisateur

Une session appartient à un seul utilisateur, la cardinalité est (1,1)

8. Utilisateur-Session

Un utilisateur peut effectuer 0 ou plusieurs sessions, la cardinalité est (O,n)

9. Lecture-Annotation

Une lecture peut produire 0 ou plusieurs annotations, la cardinalité est (O,n)

10. Annotation-Lecture

Une annotation peut être associée à une seule lecture, la cardinalité est (1,1)

11. Session-Lecture

Une session peut produire 0 ou plusieurs lectures, la cardinalité est (O,n)

10. Lecture-Session

Une lecture est faite dans une seule session, la cardinalité est (1,1)

## Schéma de relation

Annotation (Code d'annotation, date, Objectif d' annotation, Type, Type-collaboration, Problème recherche d'information, Problème décisionnel,ref objet annoté,Code utilisateur, code\_session, code\_document, code\_lecture)

Utilisateur (Code utilisateur, Nom, Prénom, Adresse postale, Adresse électronique, Ville, Région, Pays, Secteur d'activité, date de naissance)

Lecture (Code lecture, Heure départ, Heure Fin, Code\_document, Code\_utilisateur,Code\_session)

Document (Code document, url document, code\_session,code\_lecture)

Session (Code session, Heure départ, Heure fin, Code\_utilisateur)

## Graphe de relation de la collection de relations

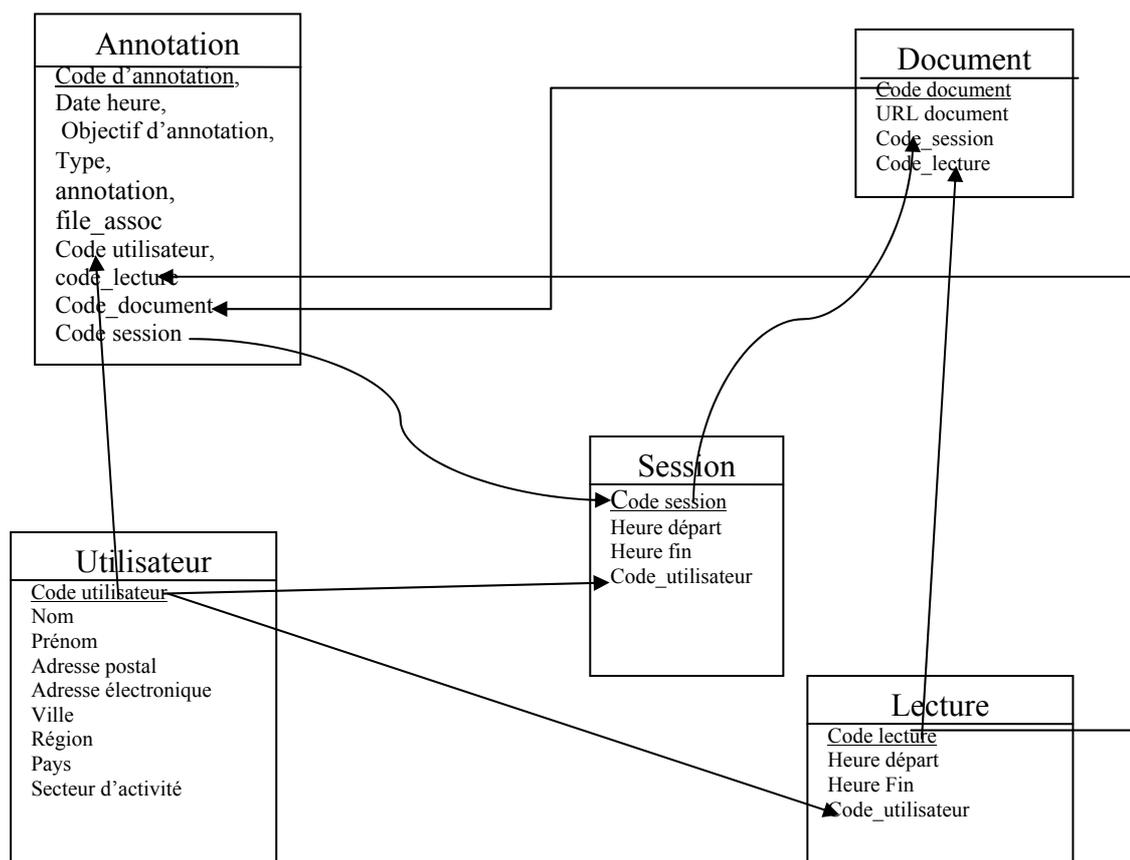


Figure 3.5 : Graphe de relation de la collection de relations

### 3.5. Schéma fonctionnel du système I-AMIE

Le système I-AMIE est un système que nous avons développé fondé sur le modèle AMIE. Le système permet la recherche d'information suivant le modèle EQuA<sup>2</sup>te mais en mettant particulièrement l'accent sur la fonctionnalité d'annotation. Rappelons ce qu'est le modèle fonctionnel EQuA<sup>2</sup>te.

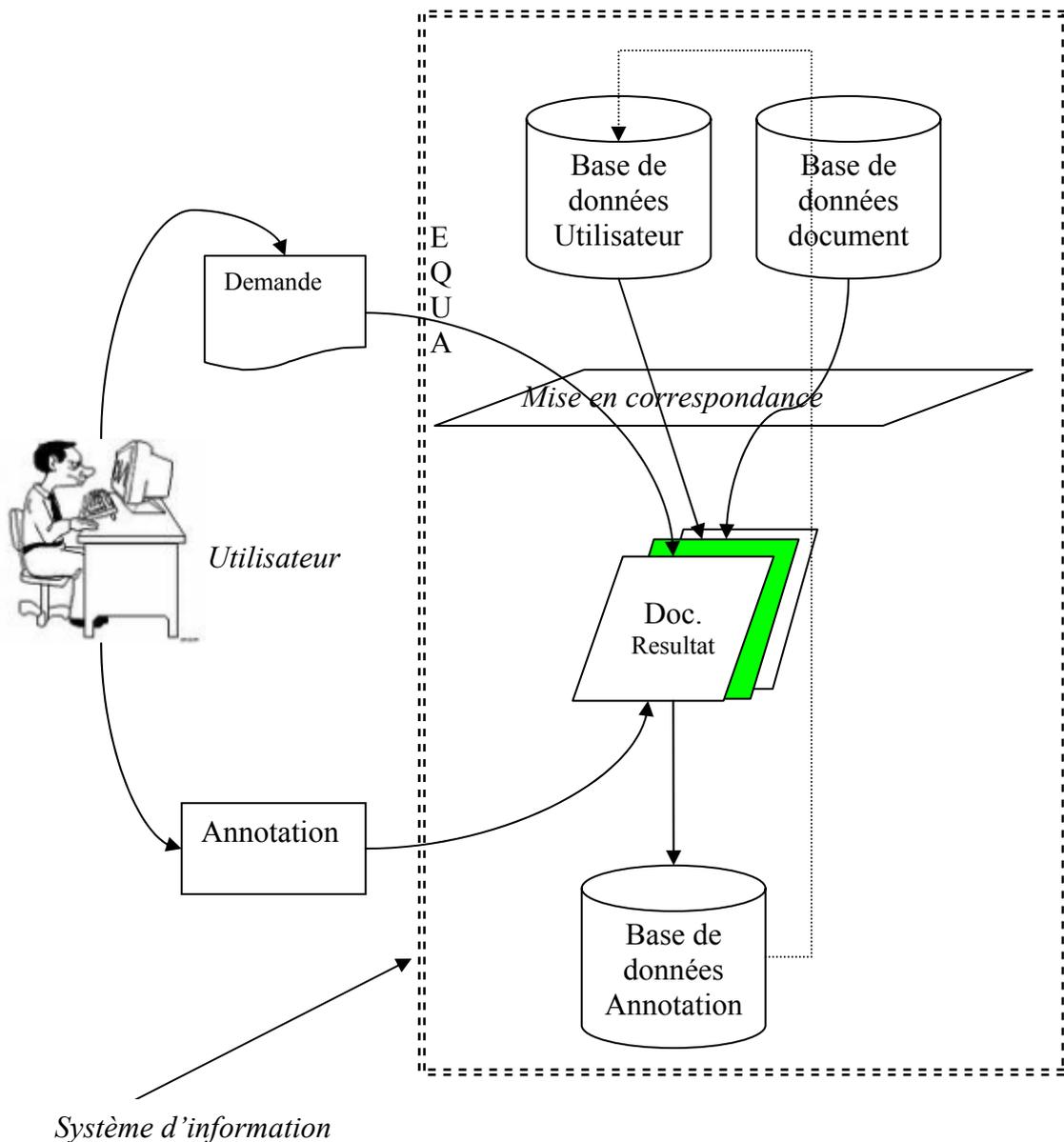


Figure 3.6 : Représentation du système I-AMIE

Nous nous inspirons de (David, 2005) pour expliquer le lien entre notre modèle AMIE et le modèle EQuA<sup>2</sup>te. Comme il s'agit d'utiliser un SI pour aider à la résolution d'un problème décisionnel, le modèle EQuA<sup>2</sup>te a été proposé pour faciliter le processus de résolution d'un problème décisionnel par un SI. Cette architecture est fondée sur les caractéristiques fonctionnelles qui permettent aux acteurs du processus d'IE (le décideur et le veilleur ou tout autre personnel de

l'organisme) d'évoluer dans les phases cognitives rencontrées dans le processus de résolution de problème (David, 2002).

L'acronyme EQUA<sup>2</sup>te est tiré de ces quatre phases : *Explorer, Interroger (Query), Analyser et Annoter*.

- **Explorer** l'entrepôt de données, c'est-à-dire, naviguer dans les données ; le système hypertexte illustre bien cette fonctionnalité. Dans les systèmes que nous proposons pour des applications dans le domaine d'IE, il s'agira de la navigation par des attributs des données du système. Par exemple, cliquant sur un attribut donnera toutes les valeurs de l'attribut, une forme de liste d'index. Et à partir de cette liste, l'utilisateur pourra obtenir l'intégralité du document ou de l'information.
- **Interroger** l'entrepôt, c'est-à-dire, utiliser des requêtes par le principe de recherche par le contenu, sous forme d'équations Booléennes. Les systèmes documentaires et la plupart des moteurs de recherche emploient cette approche pour la recherche d'information. L'inconvénient majeur est que le degré de pertinence de l'information dépend du niveau de connaissance de l'utilisateur sur les informations qu'il cherche ainsi que sur les fonctionnalités de spécification des valeurs des attributs.
- **Analyser** l'ensemble des données du système pour faire émerger des distributions, tendances, etc.. La qualité de l'analyse dépendra bien sûr du degré de compréhension du problème décisionnel ainsi que de la qualité de la transformation du problème décisionnel en indicateurs.
- **Annoter** les solutions proposées pour les contextualiser selon des critères personnels. Cette fonctionnalité permet d'adapter le système par le contenu informationnel à des particularités individuelles.

A partir de la Figure 3.6, un utilisateur envoie une demande au système d'information (délimité pas la ligne pointillée). Dans notre application, le système exploite une base de données de documents bibliographiques de l'équipe de

recherche SITE. La demande de l'utilisateur est mise en correspondance avec la base documentaire.

Par rapport aux travaux de l'équipe de recherche SITE, les trois premières fonctionnalités sont implémentées dans le système de recherche d'information développé dans l'équipe – à savoir le système METIORE.

Le système METIORE permet à l'utilisateur de pouvoir évoluer dans les différents niveaux d'habitudes évocatives (observation, abstraction élémentaire, symbolisation et raisonnement, et créativité) (David, 1999). C'est un système de recherche d'informations appliqué aux références bibliographiques. Le système est expérimenté sur une collection de 5000 références bibliographiques du laboratoire LORIA. Le système repose sur le concept que l'utilisateur doit pouvoir évoluer dans ses niveaux d'habitudes évocatives.

Notre travail porte sur l'aspect de la deuxième "A", dans l'EQUA<sup>2</sup>te, qui sont les annotations possibles par l'utilisateur.

Bien sur, l'annotation n'est pas nécessairement une expression de besoin en information mais elle peut participer à une meilleure compréhension du besoin en information suite à des annotations des résultats proposés par le système.

Nous reprenons ci-dessous trois des principaux processus du système I-AMIE.

**La demande** : La demande correspond à l'expression du besoin informationnel. Dans les applications existantes dans l'équipe de recherche SITE, trois types de demande sont implantés : Explorer, Interroger et Analyse.

**Mise en correspondance** : La mise en correspondance est effectuée par le système pour trouver des informations de la base de données documentaire qui correspondent au mieux à la demande de l'utilisateur. Pour se faire, le système peut utiliser la base de données utilisateur qui contient des informations sur l'utilisateur,

ses préférences et ses expériences afin d'avoir une meilleure compréhension de la demande exprimée par l'utilisateur.

Notre apport par rapport à la mise en correspondance est d'offrir la possibilité d'acquérir des connaissances sur l'utilisateur par le biais du process d'annotation. Ainsi la base d'annotation pourra participer à la mise à jour de la base de données utilisateur.

**Annotation** : Annotation est le commentaire de l'utilisateur après avoir lu et interprété (évalué) le document provenant de la base documentaire.

### **3.5.1. Le développement du système avec la technologie Internet**

Nous avons choisi de développer le système avec la technologie Internet, en employant l'architecture 3-tiers. Nous présentons ci-dessous très brièvement ce qu'est l'architecture 3-tiers.

L'architecture 3-tiers est composée de trois éléments, ou plus précisément dans ce cadre, de trois couches. En effet dans ce contexte, et dans la philosophie qui a guidé l'élaboration de cette architecture, il est plus adéquat de parler de couche fonctionnelle où à chacune d'elle est attaché un élément/entité logique. Dans l'architecture 3-tiers, il faut distinguer :

- Le niveau présentation (ou affichage) associée au client. C'est ce niveau qui permet à un utilisateur d'envoyer sa demande d'une ressource aux machines à distance. Généralement, un navigateur Web permet cette interface. La machine client est reliée au serveur Web par le réseau TCP/IP. Le raccordement est indépendant de ce que le client fait. Chaque fois qu'il a besoin d'une ressource (située sur le Web), l'utilisateur envoie sa demande au serveur Web à travers le protocole HTTP.

- Le niveau fonctionnel lié au serveur qui, dans de nombreux cas, est un serveur Web muni d'extensions applicatives. Dans notre cas, le serveur Web est aussi le serveur PHP. C'est ici que des commandes en PHP sont exécutées.

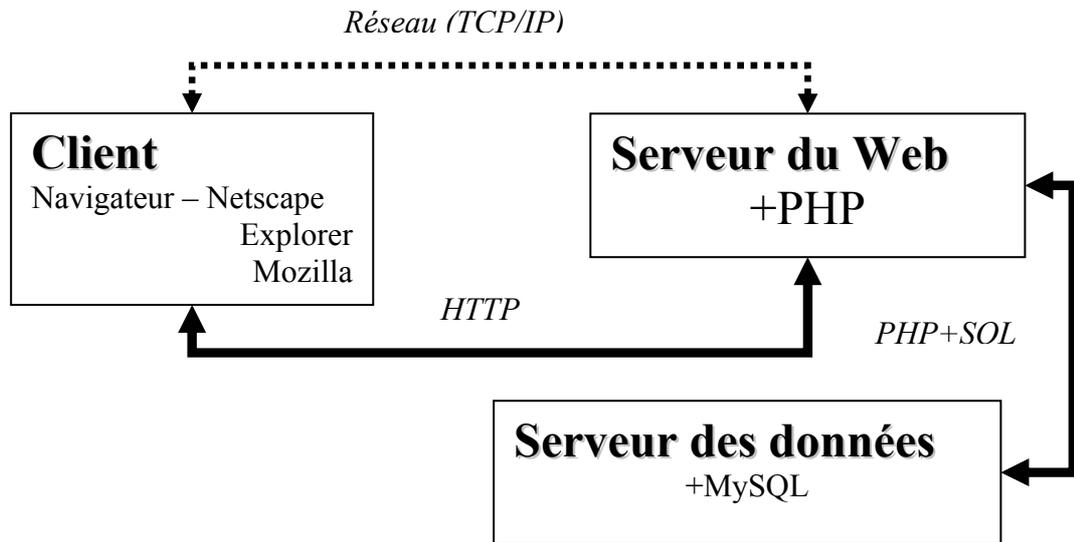


Figure 3.7 : Présentation de l'architecture 3-tiers

- Le niveau donné est lié au serveur de base de données. Les serveurs de bases données sont en charge du stockage et de la gestion des données organisées sous forme d'une base de données. Elle fournit l'information (comme donnée) au serveur Web par des commandes PHP/SQL.

Le deux principaux composants logiciels dans cette architecture sont PHP et MySQL ; que nous présentons très brièvement dans les deux paragraphes suivants.

## PHP

Nous avons choisi le PHP comme langage de programmation parce qu'il nous permet de réaliser notre objectif. PHP est un langage et un interpréteur, qui est intégré dans le serveur web ou bien compilé en version CGI. Il est capable d'accéder aux fichiers, d'exécuter des commandes et d'ouvrir des connexions réseaux. Le problème majeur de PHP est qu'il rend fragile la sécurité d'un serveur Web à cause

de sa capacité à accéder à des fichiers curant ainsi la possibilité de modifier les fichiers ou infecter le serveur de virus. De plus, une sélection rigoureuse des options de compilation et d'exécution permet d'obtenir un équilibre entre liberté et sécurité (Mehdi et al, 2005).

## **MySql**

MySQL est un système de gestion de base de données relationnel. Le SQL dans "MySQL" signifie "Structured Query Language" : le langage standard pour les interrogations. MySQL est « Open Source ». Le mot « Open Source » signifie qu'il est possible à chacun de d'utiliser et de modifier le logiciel. Tout le monde peut le télécharger sur Internet et l'utiliser sans payer aucun droit.

Pour notre travail, nous avons utilisé la version déjà installée dans le système du réseau de LORIA. Le serveur MySQL où nous avons installé notre base de données est <http://mysql.loria.fr>.

### **3.5.2. L'implantation du système I-AMIE au LORIA**

Nous avons implémenté le système I-AMIE en utilisant les ressources technologiques Internet du LORIA ainsi que des ressources documentaires de l'équipe SITE. Comme le système est développé sur une architecture 3-tiers, nous avons utilisé un serveur Web et un serveur de données. Ainsi le système peut être utilisé à partir de n'importe quel client / navigateur.

#### ***Le Serveur Web***

Nous avons pensé utiliser un serveur dédié (<http://metiore.loria.fr>) unique mais pour des raisons de sécurité, [metiore.loria](http://metiore.loria.fr) ne peut pas communiquer avec d'autres serveurs qui sont essentiels à l'opération du système comme le serveur Web LORIA. Nous avons décidé d'utiliser le serveur Web centralisé du Web LORIA ([www.loria.fr](http://www.loria.fr)).

## ***Serveur Base données MySql***

Nous avons utilisé le serveur de base de données mysql.loria.fr pour gérer la base de données de notre système. Nous aurions préféré avoir notre serveur MySQL propre extérieur plutôt que sur le serveur du LORIA centralisé. Le problème est qu'un serveur MySQL extérieur ne peut pas communiquer directement avec le serveur Web du LORIA. La communication entre un serveur MySQL et un serveur Web est essentielle pour le fonctionnement de notre système.

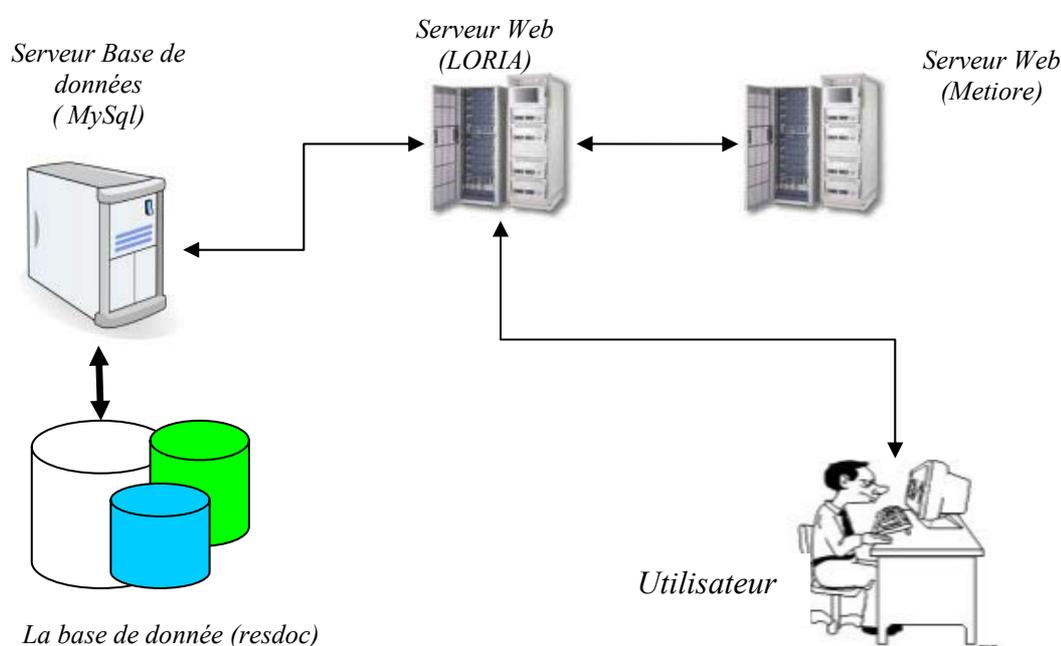


Figure 3.8: Les interactions entre les composants du système I-AIME

## **La gestion de la base de données**

La base de données que nous utilisons, *resdoc* se trouve sur le serveur mysql.loria.fr. Elle peut être gérée depuis seulement deux machines « *crossway.loriafr* » et « *lagos.loria.fr* » pour des raisons de sécurité. Cinq tables ont été créées dans la base de données (cf. figure 3.5). Les structures de ces tables sont présentées dans l'annexe 3.

### 3.5.2.1. Les composants logiciels du système I-AMIE

Dans cette section nous présentons les composants logiciels du système I-AMIE.

D'abord, l'utilisateur demande une connexion au portail qui sert d'interface entre l'utilisateur et le système (cf. ❶ figure 3.9). La méthodologie utilisée pour se connecter est la méthode de « login » avec un mot de passe. Après la connexion, l'utilisateur peut soumettre des demandes au moteur de mise en correspondance (cf. ❸ figure 3.9).

La demande correspond à l'expression du besoin en information de l'utilisateur. Les formes d'expression les plus courantes sont la navigation, l'interrogation et l'analyse globale de la totalité de la base d'informations. Ces formes ne sont pas généralement disponibles simultanément, ce qui explique l'intérêt du modèle EQuA<sup>2</sup>te qui justifie la nécessité de proposer ces fonctionnalités. La demande peut évoquer le type d'information, le format, la date et les autres attributs des documents recherchés. Le modèle Mirabel (Goria, 2004) développé dans l'équipe de recherche SITE-LORIA est bien adapté pour résoudre le problème de la spécification des besoins en information. Ce modèle a été proposé pour clarifier l'interprétation et la compréhension de la requête faite par un utilisateur.

Les modules du portail décompose la demande pour extraire les attributs des documents.

La demande peut porter sur les documents ou sur les annotations <sup>3a</sup> et <sup>3b</sup> de la figure 3.9). Dans ce dernier cas, une annotation est considérée comme un document. La demande est mise en correspondance avec la base concernée. Notons que cette mise en correspondance peut s'appuyer sur la base utilisateur pour la personnalisation des réponses à la demande. Rappelons que la base utilisateur peut être mise à jour en se servant des activités d'annotations de l'utilisateur afin de mieux cerner certaines de ses caractéristiques (cf section 3.5).

Le résultat de la mise en correspondance est stocké dans la base résultats (❹ de la figure 3.9). Cette phase peut nécessiter l'association des éléments d'un document résultat avec les annotations qui y réfèrent. Cette association peut être faite avant de stocker le résultat dans la base résultats (cf. ❹ de la figure 3.9).

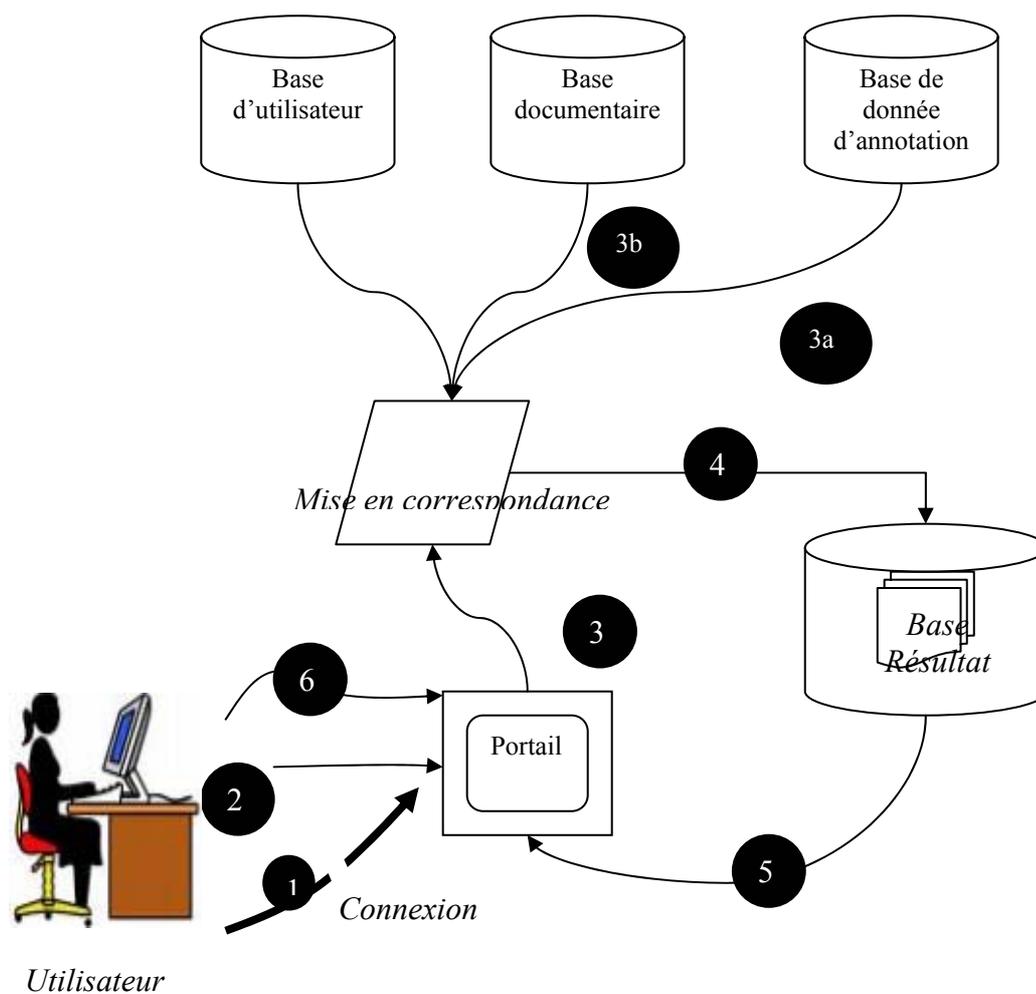


Figure 3.9 : Les composants logiciels de I-AMIE

Les résultats obtenus sont affichés sur le portail avec la possibilité d'effectuer des annotations (cf. ❺ de la figure 3.9). Notons qu'il est aussi possible de sélectionner les paramètres d'affichage des annotations, par exemple, affichage par type d'annotation, par date d'annotation, par annotateur, etc.. Enfin, nous avons prévu les fonctionnalités pour l'analyse des annotations. Dans ce cas, la demande aura été faite sur la base des annotations.

Rappelons que l'annotation permet à l'utilisateur d'apporter ses propres interprétations / évaluations sur les documents résultat.

Nous avons considéré la possibilité d'employer l'annotation comme un moyen de partage d'information au chapitre 2. Un outil d'annotation peut être adapté pour des objectifs de collaboration. Dans ce cas, l'architecture (Figure 3.9) du système I-AMIE peut inclure la recherche collaborative d'information. Trois modes sont possibles pour l'utilisation d'I-AMIE pour la recherche collaborative d'information :

- Mode autonome
- Mode observation
- Mode collaborative

Ces modes ont été proposés par (David, 1999) et appliqués sur deux systèmes de l'équipe SITE. Nous présentons succinctement ces modes dans les trois paragraphes suivants.

**Mode Autonome :** Dans le cas d'une utilisation autonome d'I-AMIE, un utilisateur a toutes les ressources à lui seul. Il n'a rien en commun avec d'autres utilisateurs. Il peut ne pas partager ses expériences ou ses annotations avec d'autres utilisateurs.

**Mode Observation :** Là où un utilisateur observe l'utilisation d'I-AMIE faite par les autres utilisateurs. Il n'est pas censé de faire sa propre annotation mais il peut tirer les bénéfices des annotations et des expériences d'autres annotateurs.

**Mode Collaborative :** Le cas d'une utilisation collaborative d'I-AMIE d'annotation est très important pour nous car nous essayons de résoudre un problème décisionnel qui peut être une tâche entre un décideur et un veilleur (ou entre deux personnes simplement). Dans ce cas-ci, la base de données d'annotation est commune aux deux utilisateurs en collaboration. Les portails auxquels ils accèdent sont différents. Comme illustré dans figure 3.13, le serveur de base de données et le Serveur Web sont communs aux deux utilisateurs car ils utilisent le même système I-AMIE. Les clients peuvent être évidemment différents selon le navigateur utilisé.

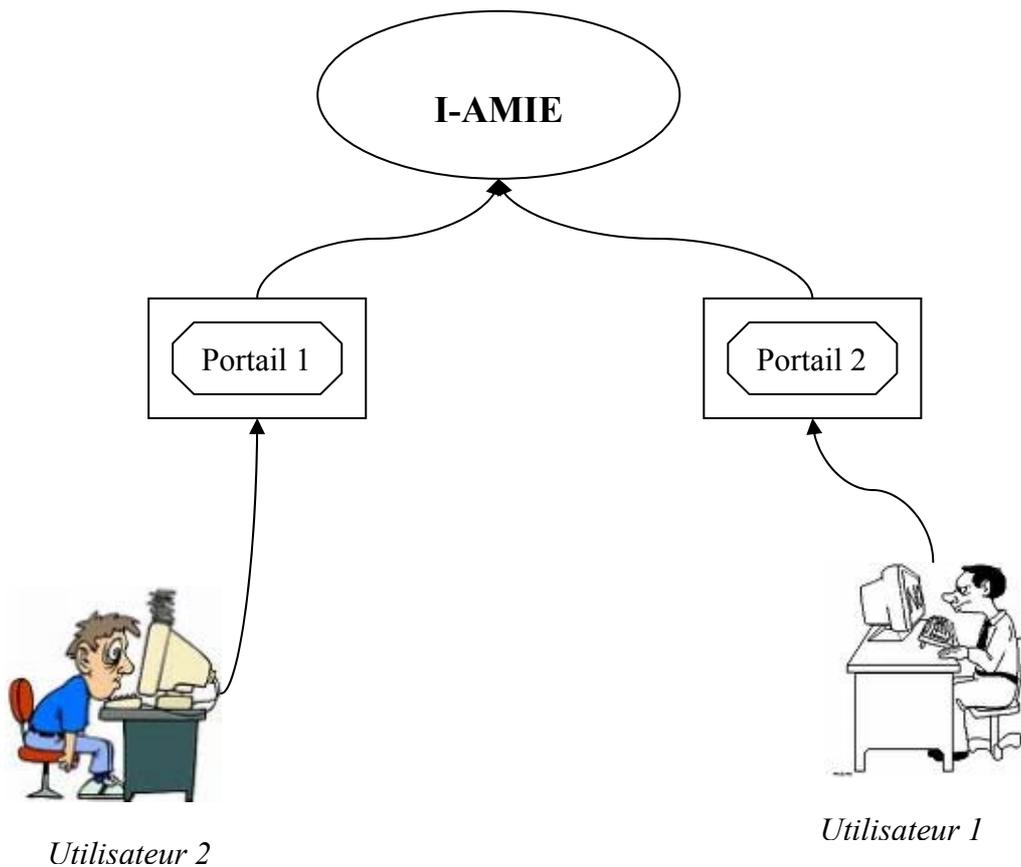


Figure 3 .10 : Utilisation de système I-AMIE pour la recherche d'information

Outre l'utilisation de I-AMIE pour la recherche collaborative, nous pensons que I-AMIE peut être utilisé en mode distribué selon quatre fonctionnalités que nous présentons ci-dessous.

#### *A. Raffineur*

Par cette fonctionnalité, I-AMIE peut être utilisé pour affiner une requête en recourant à la base d'annotation. Par exemple, une requête d'information sur "l'intelligence économique" à travers google peut être étendue vers la base de données d'annotation d'I-AMIE si l'utilisateur n'est pas satisfait de la réponse de google.

#### *B. Centralisateur*

Le système I-AMIE peut servir de fédérateur des annotations. Ainsi des annotations faites dans plusieurs domaines peuvent être centralisées et constituent ainsi une base

de connaissances.

### *C. Analyseur.*

Nous pouvons utiliser I-AMIE pour analyser les annotations qui viennent d'autres sources. Comme nous avons dit dans la section 3.5, des annotations et des documents sont mis en correspondance avec la demande de l'utilisateur. La base de document ou la base d'annotation peut aussi être le système METIORE.

### *D. Système multi-agent*

I-AMIE peut confier des traitements particuliers à des systèmes externes. Par exemple, le système SIMBAD (Sidhom, 2002) peut être sollicité pour traiter le résultat du système I-AMIE. Le système SIMBAD transforme cette information en attributs et valeurs. Ces attributs et ces valeurs peuvent enrichir la base de données d'annotation.

## **3.5.2.2. Les interfaces**

Nous présentons dans cette section des exemples d'interface du système I-AMIE illustrées par des copies d'écran. La première interface est l'interface de connexion (Figure 3.11). Un utilisateur qui est déjà inscrit dans le système devra entrer son pseudonym et son mot de passe. Cette interface est liée à la base de données de l'utilisateur. Quand l'identité de l'utilisateur n'est pas trouvée dans la base de données ou si le mot de passe ne correspond pas à celui de la base de données, une erreur est rapportée et le système est interrompu. Un nouvel utilisateur peut enregistrer son identité dans la base de données (Figure 3.13). Les interfaces du système ont été réalisées en français et en anglais. L'utilisateur peut sélectionner la langue de son choix.

Cette section de la thèse a pour objectif d'expliquer certaines fonctionnalités du système d'annotation qui ont été réalisées pour la mise en place du système d'I-AMIE. Nous voulons également mettre en évidence le rapport entre chaque section du système. Nous expliquons l'interface liée à la base de données. Nous illustrons également les rapports entre les différents éléments du système avec un schéma.

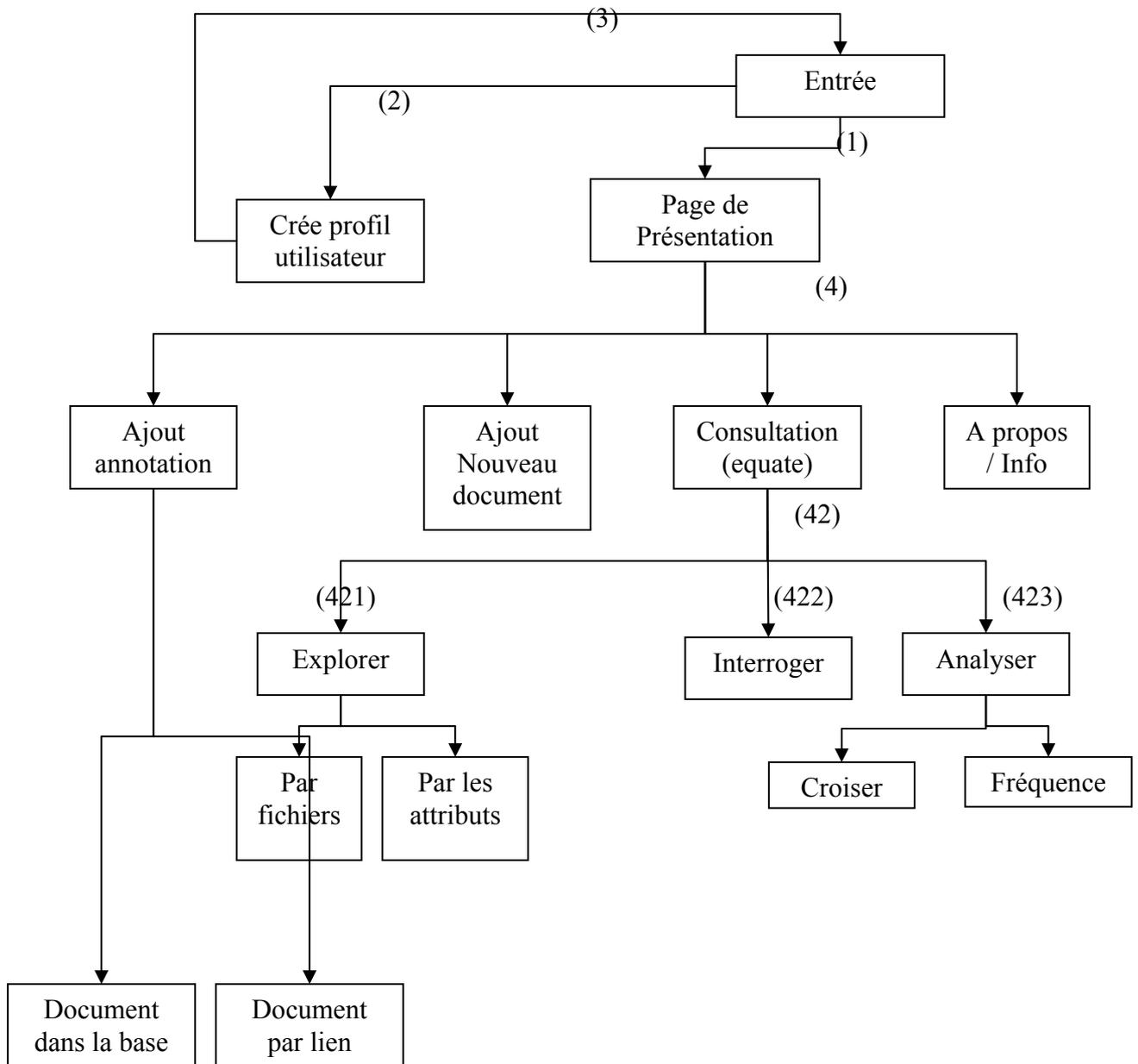


Figure 3.11: Schéma fonctionnel d'I-AMIE

11/11/2006

**Interface de connexion**

**Nom Utilisateur**

**Mot de Passe**

Nouveau utilisateur ? [Cliquer ici](#)

 [Français](#)
 [English](#)

Figure 3.12 : Ecran de login

L'accès au système est le « login ». Un utilisateur se connecte au système avec un identifiant et un mot de passe. Un utilisateur qui n'a pas un d'identifiant et un mot de passe correspondance peut créer son compte utilisateur (identifiant et mot de passe) (2). L'identifiant et le mot de passe associé sont sauvegardés dans la base de données. Après la création du nom d'utilisateur et du mot de passe, il peut alors se connecté au système (1). Quand l'utilisateur se connecte avec succès au système, une première page de présentation du système d'annotation s'affiche.

<b>Systeme d'annotation dans le contexte d'IE</b>		
<input type="radio"/> Acceder et annoter les documents		
<input type="radio"/> Acceder aux annotations precedentes		
<input type="radio"/> Ajouter un document sur la base document	<input type="button" value="Envoyer"/>	<input type="button" value="Annuler"/>
<input type="radio"/> A propos du systeme d'annotation		

Figure 3.13: Présentation de système d'annotation

La page de présentation du système d'annotation présente quatre types d'information (a) l'accès aux documents associés à une annotation et la création d'une l'annotation, (b) la consultation des annotations existante (c) ajout d'un nouveau document et (d) à

propos du système. Trois méthodes ont été adoptées pour consulter les annotations existantes. Ces méthodes sont : explorer (421), interroger (422) et analyser (433).

Bien que les interfaces soient multilingues, une seule base de données (resdoc) est utilisée dans le système.

### 3.5.2.2.1. La creation d'identifiant utilisateur

New User Creation	
Pseudoname	<input type="text"/>
Password	<input type="password"/>
Confirm Password	<input type="password"/>
FirstName	<input type="text"/>
Surname	<input type="text"/>
email address	<input type="text"/>
Address 1	<input type="text"/>
Address 2	<input type="text"/>
Postal code	<input type="text"/>
Region	<input type="text"/>
Country	<input type="text" value="-selection-"/>
Sector of activity	<input type="text" value="- Make selection-"/>
Date of Birth DD/MM/YYYY	<input type="text"/>
Personal Remarks	<input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

Figure 3.14 : Ecran de création d'un nouvel utilisateur

L'interface de création d'un compte utilisateur permet la création d'un compte utilisateur dans la base de données utilisateur. Les paramètres de l'utilisateur comme nous les avons montrés sur la figure 3.5 sont renseignés ici. L'utilisateur est libre de saisir ses informations personnelles comme le nom, adresse postale, etc. Il doit créer un pseudonome et un mot de passe avec lesquels il pourra se connecter. C'est

obligatoire qu'un utilisateur crée un pseudonyme. En effet, le pseudonyme est la référence à toutes les activités de l'utilisateur dans le système. Ce dernier (l'utilisateur) peut envoyer ses informations à la base utilisateur en cliquant sur « Submit » ou il peut réinitialiser ce qu'il a saisi avec « Reset ».

### 3.5.2.2.2. La creation d'annotation

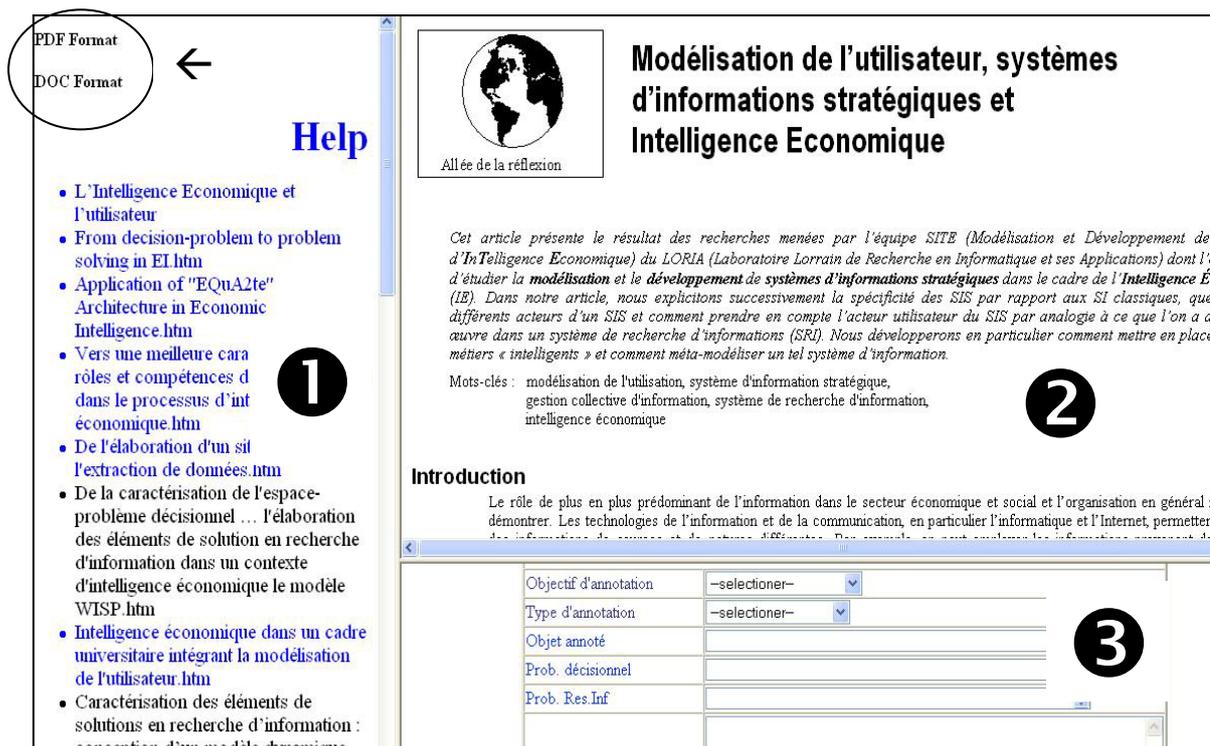


Figure 3.15 : Interface de création d'annotation

Quand un utilisateur est connecté au système, le portail est présenté avec une page de trois cadres. Le premier cadre ❶ du côté gauche contient les références des documents primaires qui sont stockés sur metiore.loria.fr. Trois formats de fichiers sont stockés dans metiore.loria.fr (format DOC, PDF et HTML). L'utilisateur peut choisir le type de format qui lui convient dans le cadre ❶ à gauche (indiqué par ← dans Figure 3.15). L'utilisateur peut alors cliquer sur un document parmi les liens présentés dans le cadre gauche ❶. En fait, le lien permet d'afficher le document dans son intégralité dans le cadre ❷ en haut à droite. L'utilisateur peut alors annoter le document dans le cadre ❸ situé juste en bas du cadre de document. Le cadre ❷ contient initialement une brève information sur le système d'annotation et des liens liés

à nos travaux d'équipe (Figure 3.16). L'utilisateur peut cliquer sur un lien ou sur le cadre ❶ pour accéder à son contenu.

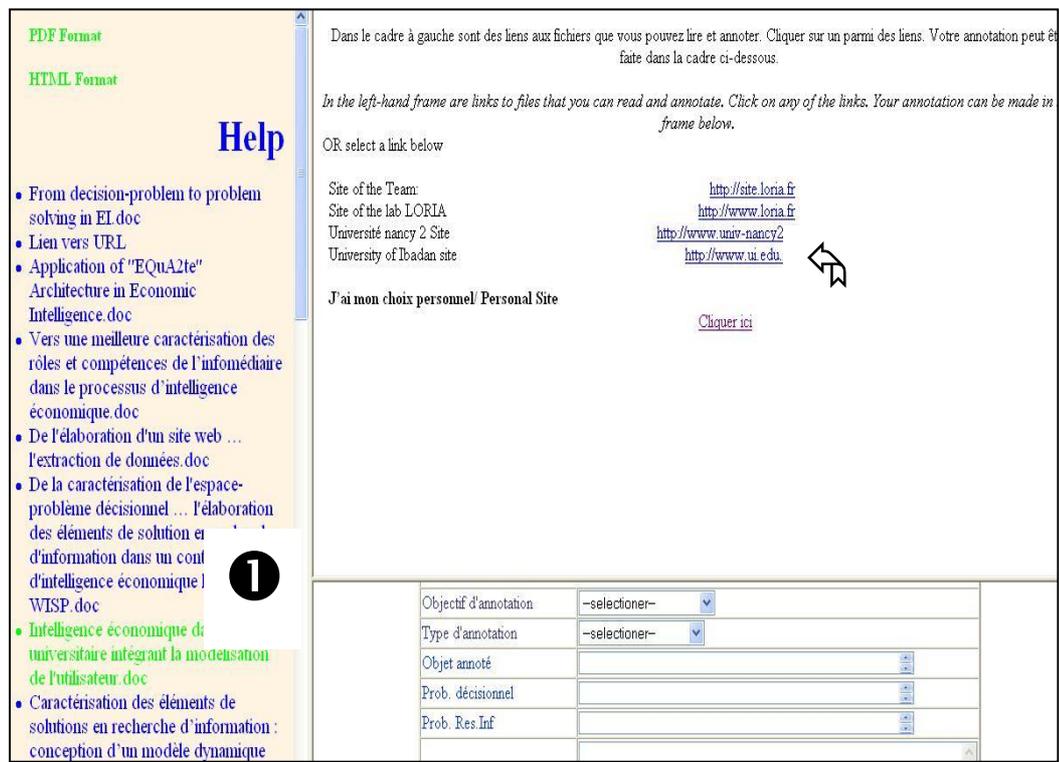


Figure 3.16 : Premier écran dans la création d'annotation

Le cadre ❷ peut être chargé avec la page Web de l'utilisateur (en format complet comme <http://www.pagepersonnel.com>) en cliquant sur le lien pages personnalisées (☞ dans Figure 3.16). L'utilisation de ce lien n'est pas toujours sûr compte tenu de la surcharge des pages Web ou de l'affichage des pages Web graphiques.

Supposant que l'utilisateur clique sur un lien dans le cadre ❶ « *Modélisation de l'utilisateur, systèmes d'informations stratégiques et Intelligence Economique.htm* », le fichier lié sera affiché dans le cadre ❷.

<b>Création d'annotation</b>	
Objectif d'annotation	-selectioner- ▼
Type d'annotation	-selectioner- ▼
Objet annoté	<input type="text"/>
Prob. décisionnel	<input type="text"/>
Prob. Res.Inf	<input type="text"/>
Annotation	<input style="width: 100%; height: 50px;" type="text"/>
Section annotée	-selectioner- ▼
Type collaboration	-selectioner- ▼
<input type="button" value="Envoyer Annotation"/> <input type="button" value="Lire simplement"/>	

*Figure 3.17 : Zone de création d'annotation*

Dans la zone de création d'annotation représentée sur la Figure 3.17, l'utilisateur est invité à faire son choix à partir des options prédéfinies dans section 3.2. Par exemple, il pourra sélectionner : « Objectif d'annotation », « Type d'annotation », « Section annotée » et « Type de collaboration » etc. Nous aurions aimé avoir une situation où le document à annoter soit directement inclus dans l'interface de création d'annotation, c'est-à-dire dans le cadre 3. Ainsi nous aurions directement les références du document à annoter. Nous n'avons pas pu réaliser ceci vu les contraintes du temps.

Après avoir choisi les options, l'utilisateur doit compléter les valeurs en relation avec le « Problème décisionnel », « Problème de recherche d'information » et sa propre annotation.

L'utilisateur peut envoyer l'annotation faite à la base d'annotation avec « Envoyer Annotation ». Nous avons combiné la « lecture de document » avec l'annotation dans la même interface. C'était pourquoi nous avons le « Lire simplement » sur l'interface. Cette option n'a pas été suffisamment développée.

### 3.5.3. Consultations des annotations existantes

#### 3.5.3.1. Affichage des annotations

Nous n'avons pas pu réaliser les modules permettant à l'utilisateur de choisir les modes d'affichage des annotations. En revanche nous avons développé un module qui permet de lister le contenu du fichier d'annotation comme représenté sur le schéma 3.18.

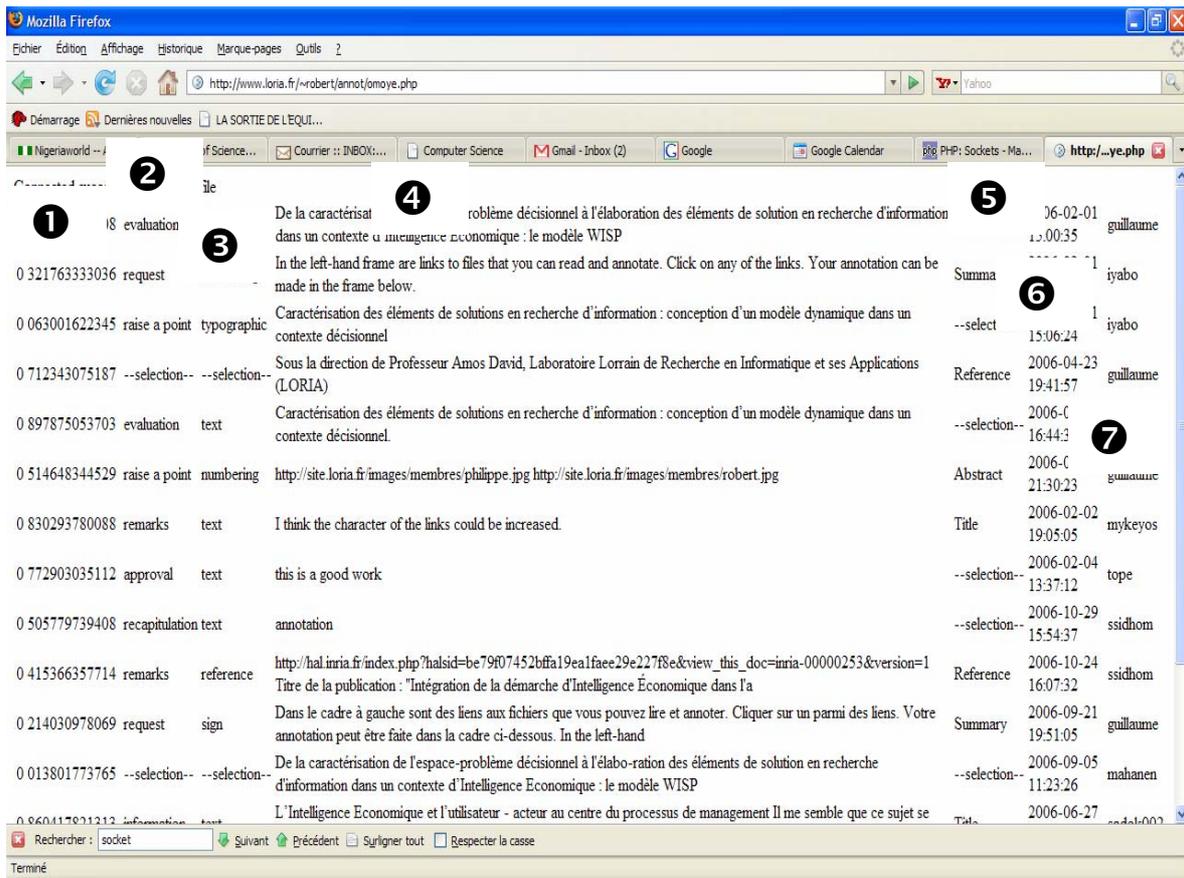


Figure 3.18 : Ecran de liste des annotations dans la base

Nous présentons la signification de chaque colonne dans la tableau 3.2.

Colonne	Signification
1	Reference d'annotation
2	Objectif d'annotation
3	Type d'annotation
4	Annotation propre
5	Section du document annoté
6	Date d'annotation
7	Pseudonym d'utilisateur

Tableau 3.2 : Signification des colonnes de l'écran 3.18

Après avoir présenté le système I-AMIE par ses fonctionnalités et pas ses composants logiciels, nous présentons dans la section suivante des exemples de domaines d'application aussi bien de la base d'annotation que du système I-AMIE.

### 3.5.3.2. Accès et interrogation de base d'annotation

#### 3.5.3.2.1. Explorer

Figure 3.19 : Interface pour explorer le contenu de la base

Quand « cliquer pour explorer par les attributs » est choisi, un utilisateur peut accéder aux occurrences spécifiques d'un attribut dans le système. Par exemple, un utilisateur peut vouloir savoir quel est les types de différents « Problèmes décisionnels » qui existent dans le système ou les différents types de « Problème de recherche de « Problème de recherche d'information » saisie. L'utilisateur peut alors cliquer sur n'importe lequel des types affichés pour obtenir la fréquence de l'occurrence choisie.

<b>Exploration de la base d'annotation</b>	<input type="radio"/> Objectif <input type="radio"/> Type d'annotation <input type="radio"/> Section annote <input type="radio"/> Probleme decisionnel <input type="radio"/> Probleme recherche d'information <input type="radio"/> Objet annote <input type="radio"/> Type collaboration	Envoyer	Annuler

Figure 3.20 : Interface pour sélection une variable a exploré

Au lieu d' « explorer » le système par des attributs, un utilisateur peut « explorer » le contenu de des bases de données dans le système. Par exemple, il peut analyser les utilisateurs qui sont stockées dans le système.

### 3.5.3.2.2. Interroger

#### Logique de recherche d'information

Interroger permet à l'utilisateur de choisir plusieurs attributs pour avoir leur recurpement. Par exemple, un utilisateur peut vouloir savoir l'occurrence de « objectif d'annotation » type « évaluation » avec la « section d'annoté » type « titre » ensemble. Il est possible de choisir « objectif d'annotation » type « évaluation » avec « objectif d'annotation » type « remarque ». Un utilisateur peut croiser jusqu'à quatre paramètres dans ce cas. Pour le croisement des attributs et des valeurs, nous avons utilisé les opérateurs Booléens ET et OU.

**recherche par:**

Auteurs tous types	<input type="text"/>	Et	<input type="button" value="↓"/>	<input type="button" value="👍"/>
Titre	<input type="text"/>	Et	<input type="button" value="↓"/>	
Sujet	<input type="text"/>	Et	<input type="button" value="↓"/>	
Note de thèse/mémoire	<input type="text"/>			

Figure 3.21 : Interface de recherche des livres de bibliotheque d'université Nancy 2

<b>Interrogation d'annotation</b>		
objectif	Remarques ▼	<input type="radio"/> ET <input type="radio"/> OU
type	Typographique ▼	<input type="radio"/> ET <input type="radio"/> OU
section	Phrase ▼	<input type="radio"/> ET <input type="radio"/> OU
coltype	Autonome ▼	
<input type="button" value="Envoyer Requete"/>		<input type="button" value="Reset"/>

Figure 3.22 : Un essai comparable aux autres systèmes de recherche d'information

Nous avons été inspirés par les systèmes d'information existants, comme le système de recherche des ouvrages mise en place par la bibliothèque de l'université de Nancy 2. L'avantage de cette option est la possibilité de sélectionner les opérateurs Boolèen après chaque attribut comme dans la Figure 3.22. Ce n'est pas une méthodes efficace de rechercher d'information dans une base d'information. La logique derrière ce genre de combinaison des opérateurs Boolèens est défectueuse. Les intentions des utilisateurs ne sont pas représentées.

Par exemples, si nous avons trois valeurs A, B et C pour trois attributs, la représentations des trois parametres peuvent prendre une des format suivants :

(Paramètre A « OU » Paramètre B) « OU » Paramètre C ..... 1  
 Paramètre A « OU » (Paramètre B « OU » Paramètre C) ..... 2  
 Paramètre A « OU » Paramètre B « OU » Paramètre C ..... 3

Dans la représentation dans Figure 3.22, les trois équations sont identiques. Mais nous sommes convaincus que les trois équations sont totalement différentes selon les théories des ensembles.

### Les méthodes de représentation possibles

**Méthode 1 :** Une méthode pour résoudre le problème est de représenter exclusivement les combinaisons des attributs. Une représentation de ce genre

exclusivement impliquera 6 possibilités de combinaison pour quatre attributs utilisant « ET » entre deux paires d'attributs seulement (exemple (A+B)+(C+D)). Pour utiliser la combinaison de « ET » et « OU » entre deux paires d'attributs, nous avons plus de 1290 possibilités. C'est au delà de notre objectif dans ce travail.

**Méthode 2 :** Une autre méthode pour résoudre ce genre de problème de représentation est de couper la représentation en étapes. La première étape peut être comme  $(A+B)=K$  et  $(C+D)=L$ . Après la première étape, nous pouvons avoir la deuxième comme  $(K + L)$ . Nous estimons que cette approche peut être trop compliquée et difficile à maîtriser par la plupart des utilisateurs du système.

**Méthode 3 :** Une autre approche pour résoudre ce problème est de permettre aux utilisateurs de saisir leur demande comme il souhaite. Un problème est associé à cette méthode. Cette méthode encourage les erreurs syntaxe. Puisque les utilisateurs peuvent ne pas saisir les requêtes justes.

### Notre approche

<b>Interrogation d'annotation</b>	
attribut1	--selectioner-- ▼
attribut2	--selectioner-- ▼
attribut3	--selectioner-- ▼
attribut4	--selectioner-- ▼
<input type="button" value="Envoyer requête"/>	<input type="button" value="Annuler"/>

Figure 3.23 : Interface de spécification d'une interrogation

Nous avons employé l'approche de deux niveaux en choisissant les attributs et les variables correspondantes dans la recherche d'information dans la base de données d'annotation. Le premier niveau est de choisir les attributs qui doivent être associés

aux variables. L'utilisateur doit au minimum un attribut et au maximum quatre attributs. Quand plus d'un attribut est utilisé, il s'agit du même attribut (répétitive) ou un combinaison de plus d'un attribut. Par exemple, nous pouvons avoir « objectif », « objectif », « objectif » et « objectif » avec l'opérateur « OU » pour localiser quatre objectifs différents dans la base de données d'annotation. Nous pouvons également utiliser quatre « objectif » avec l'opérateur « ET » pour trouver quatre objectifs différents dans la base de données d'annotation. Nous pouvons aussi utiliser attribut « objectif » avec attribut « section annoté » pour trouver l'existence d'un « objectif » avec une « section annotée » du document.

Après avoir choisie ces attributs, pour la deuxième étape, l'utilisateur indique les valeurs associées aux attributs. Nous notons que les valeurs fondés sur les valeurs « prédéfinies » pendant la création d'annotation. Nous avons mis en place la possibilité d'un seul opérateur à la fois.

<b>Interrogation d'annotation</b>	
objectif	--selectioner-- ▾
type	--selectioner-- ▾
section	--selectioner-- ▾
coltype	--selectioner-- ▾
<input type="radio"/> ET <input type="radio"/> OU	
<input type="button" value="Envoyer"/>	<input type="button" value="Annuler"/>

*Figure 3.24 : Interface d'interrogation pour choisir les variables à croiser*

La réunion des "attributs" et les "valeurs" indiqués sont affichés comme dans la Figure 3.25.

Nous avons 11 occurrences dans la base
1:-Annotation "De la caractérisation de l'espace-problème décisionnel à l'élaboration des éléments de solution en recherche d'information dans un contexte d'Intelligence Economique : le modèle WISP."
-----
2:-Annotation "Caractérisation des éléments de solutions en recherche d'information : conception d'un modèle dynamique dans un contexte décisionnel."
-----
3:-Annotation "Caractérisation des éléments de solutions en recherche d'information : conception d'un modèle dynamique dans un contexte décisionnel."
-----
4:-Annotation ".http://site.loria.fr/images/membres/philippe.jpg http://site.loria.fr/images/membres/robert.jpg."
-----
5:-Annotation "I think the character of the links could be increased."
-----
6:-Annotation ".this is a good work."
-----
7:-Annotation ".annotation."
-----

Figure 3.25 : Exemple de résultat d'une interrogation utilisant « OU »

### Opérateurs prolongés

D'autres opérateurs associés à la recherche d'information incluent NOT, XOR et NEAR.

Commande	Signification	syntaxe	Explication
NOT	SAUF	\$a    \$b	Trouve un mot \$a où il n'y a pas \$b. Exemple "information"    "intelligence"
XOR	XOR	\$a XOR \$b	Trouve \$a ou \$b mais où l'existence des deux mots ensemble n'est pas acceptés. Exemple "information" XOR "intelligence".
NEAR	Proximité	\$a NEAR \$b	La possibilité de localiser un mot, une expression qui est à côté d'un autre mot. Par exemple "Information" near "intelligence". La proximité dépend de la préférence des utilisateurs.

Tableau 3.3 : Les autres opérateurs qui ne sont pas utilisés

Nous n'avons pas employé NOT, XOR et NEAR parce que l'incorporation de ces opérateurs dans notre système peut prendre beaucoup de temps dans la recherche de chaque élément dans une base de données d'information textuelle. Elle exige également beaucoup de temps pour que nous écrivions les programmes qui mettent en application les logiques.

### 3.5.3.2.3. Analyser

Analyse d'annotation	
J'aimerais voir les	Objectif d'annotation <input type="button" value="v"/>
Mes paramètres pour le croisement sont:	
Prob. décisionnel	<input type="text"/> <input type="button" value="▲"/> <input type="button" value="▼"/>
Prob. Res.Inf	<input type="text"/> <input type="button" value="▲"/> <input type="button" value="▼"/>
Annotation	<input type="text"/> <input type="button" value="▲"/> <input type="button" value="▼"/>
<input type="radio"/> ET <input type="radio"/> OU	<input type="button" value="Envoyer"/> <input type="button" value="Annuler"/>

Figure 3.26 : Interface pour saisir les textes a « analyser » dans la base

L'objectif d' « analyser » est de permettre un utilisateur de saisir sa propre requête sur trois attributs. Les attributs concernés sont « Problème décisionnel », « Problème de recherche d'information » et les « Annotation » textes. Les valeurs saisies sont recherchées dans la base d'annotation. Nous avons utilisés les mêmes opérateurs comme dans le cas d' « INTERROGER ».

Dans l'interface Figure 3.25, nous pouvons être intéresser dans les « objectifs d'annotation » et les « annotation » sur « Problème décisionnel » lié a « TGV », transformé a la recherche d'information « information ». L'interface nous permettre de trouver toutes les annotation et l'objectif d'annotation liés au problème décisionnel » TGV ou les « problème de recherche d'information » est informationnel.

### 3.5.4. Diverse

#### 3.5.4.1. Alimentation de la base de document

Avec le temps, les documents dans la base de données devront être alimentés pour les meilleures informations. Deux types de documents sont stockés dans notre base de données de document. Nous avons les documents de source stockés dans notre serveur de projet (<http://metiore.loria.fr>) et les documents stockés par référence. Les documents sur le serveur de projet peuvent être mis à jour par les individus désignés. Nous avons fait le nécessaire pour les mises à jour des documents par référence. Dans ce cas, un utilisateur du système peut ajouter des documents en se référant à la base de données de document.

Ajouter un document sur la base	
Nom d'utilisateur	<input type="text"/>
Lien complete de document	<input type="text"/>
Titre de document (complete)	<input type="text"/>
Commentaire sur document	<input type="text"/>
<input type="button" value="Confirm"/> <input type="button" value="Annuler"/>	

Figure 3.27 : Interface pour alimenter la base de document

On s'attend à ce que l'utilisateur fournisse son identifiant. Seulement les utilisateurs enregistrés peuvent ajouter les nouveaux documents au système. On s'attend à ce qu'il fournisse une référence complète par lien et titre du document en question. Il peut également écrire quelques remarques au sujet de ces documents.

### 3.5.4.2. Filtration des documents affichés

Autrefois, la liste de documents pour l'annotation devient si longue qu'il devient pratiquement difficile que un utilisateur identifie le document intéressant pour lui. Nous avons pris la disposition de filtrer les documents à la convenance de l'utilisateur.



The image shows a web interface titled "Parametre doc." with a yellow background. It contains two text input fields, each labeled "1. Contenu doc." and "2. Contenu doc." with small up/down arrow icons to their right. Below these fields are two radio buttons: "ET" and "OU". At the bottom, there are two buttons: "Envoyer" and "Annuler".

Figure 3.28 : Interface pour spécifier les paramètres de filtration

L'utilisateur peut imaginer les documents intéressants en fournissant un ou deux identités des documents qui peuvent être dans le titre du document. Il peut combiner deux identités avec l'opérateur booléen de "ET" ou "OU". Les identités trouveront les documents pertinents au utilisateur.

### 3.5.4.3. Recherche Plus

Un utilisateur dans la recherche d'un document dans la base de document, envoie sa requête avec des paramètres spécifiques. Dans le cas où il ne peut pas localiser le document d'intérêt dans la base, nous avons imaginé qu'il peut reconduire sa recherche aux autres bases de documentaire (notamment base d'annotations). Nous avons fait le nécessaire pour que les utilisateurs puissent leur recherche des documents dans la base des annotations qui ont été faites sur les documents. L'utilisateur peut également rechercher des documents dans la base de données utilisée par google.com. L'implantation complète de la recherche dans la base de google.com n'a pas été accomplie en raison du temps.

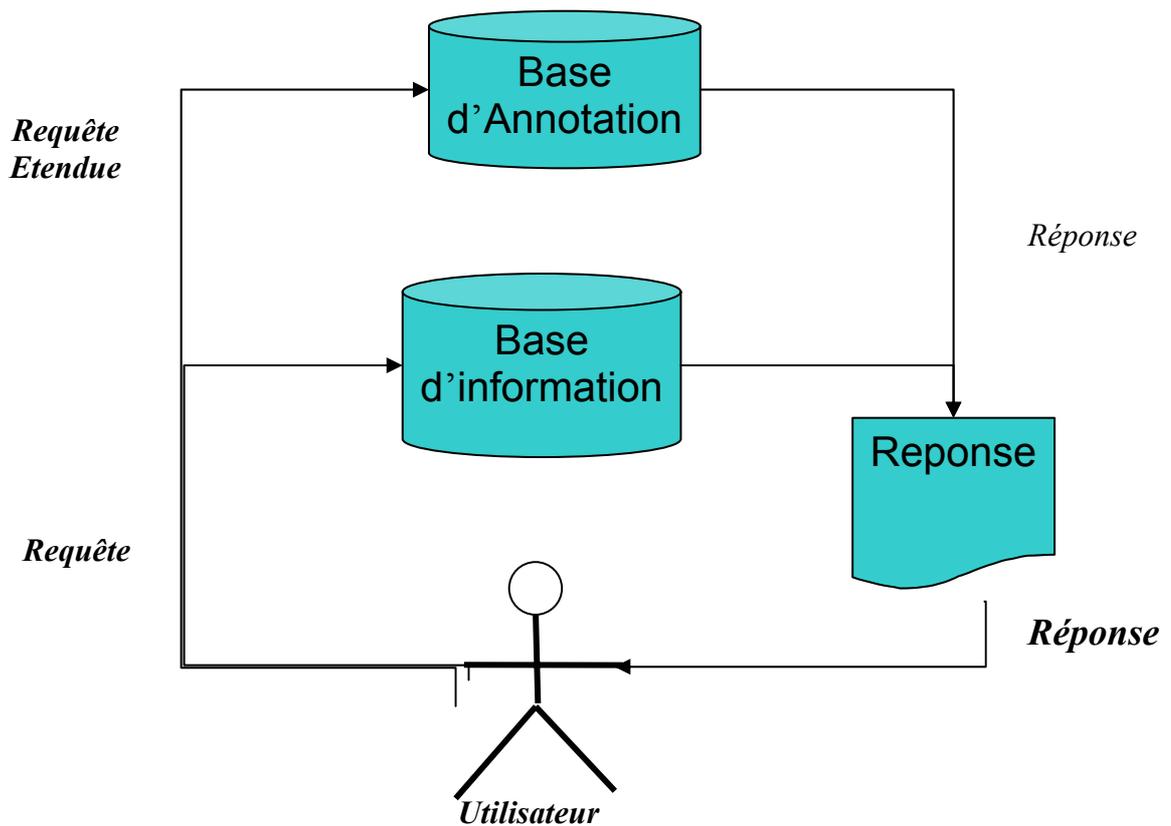


Figure 3.29 : Recherche d'information avec annotation

### 3.6. Domaines d'application du système I-AMIE

#### 3.6.1. Application des annotations dans un contexte d'IE

Pour appliquer le système I-AMIE en IE, nous considérons les annotations avec les paramètres: période, utilisateurs et documents. L'ensemble des annotations sur un ou plusieurs documents, par un ou plusieurs utilisateurs, pendant une période peut être utilisé pour la résolution de problèmes de recherche d'informations dans les processus d'intelligence économique.

Les annotations peuvent être analysées par croisement des attributs, par exemple avec la formule suivante :

$$\iiint_x dUdTdD$$

Où  $dU$  correspond à l'ensemble des utilisateurs,  $dT$ , et  $dD$  les périodes des annotations et les documents annotés respectivement (cf. (1) dans tableau 3.4).

Un ou plusieurs de ces paramètres peuvent être fixés et d'autres peuvent changer. Quand les trois paramètres sont fixés, cela concerne à l'ensemble des annotations d'un utilisateur. Dans le cas où aucun des trois paramètres n'est fixé, nous nous intéressons à toutes les annotations possibles sur tous les documents, par tous les utilisateurs (*ligne 8 dans le tableau 3.4*).

On peut s'intéresser aux annotations faites par un utilisateur particulier sur un document particulier pendant une période (*ligne 7 dans le tablea 3.3*). L'objectif peut être de voir les réactions de l'utilisateur en face d'un événement. Cela peut être représenté par :

$$UD \int dT$$

Nous pouvons nous intéresser à la perception diverse ou aux compréhensions diverses d'un groupe de personnes sur un sujet particulier (document) pendant une période donnée (cf. ligne 4), représenté par :

$$DT \int dU$$

**Ligne 4 :** Dans le cas des informations liées à l'utilisateur, nous pouvons nous intéresser aux perceptions générales d'un individu dans une période sur les sujets (documents). Les annotations faites par cet individu pendant la période sur plusieurs documents peuvent révéler des informations importantes. Cela peut nous donner la tendance générale des événements autour de lui (son environnement) et ainsi nous permettre de mieux le caractériser (cf. ligne 6), représente par .

$$UT \int dD$$

D'autres attributs peuvent être utilisés pour le croisement des informations :

- **l'annotateur** avec **l'enjeu** pour connaître les annotateurs qui sont intéressés pour les enjeux spécifiques.
- **l'objectif** avec **l'annotateur pour prendre connaissance des annotateurs qui** font les annotations avec les objectifs spécifiés.
- **le document** avec **l'objet** d'annotation pour savoir les raisons des annotations.
- Etc....

	contexte d'annotation	paramètres fixé			Représentation
		utilst	doc	périod	
1	Toutes les annotations de tous les documents de tous les utilisateurs				$\{dUdDdT$
2	Toutes les annotations de tous les utilisateurs sur tous les documents pendant une période spécifique			X	$T\{dUdD$
3	Toutes les annotations de tous les utilisateurs d'un document spécifique		X		$D\{dUdT$
4	Toutes les annotations de tous les utilisateurs sur un document dans une période spécifique		X	X	$DT\{dU$
5	Toutes les annotations d'un utilisateur sur tous les documents	X			$U\{dDdT$
6	Toutes les annotations d'un utilisateur sur tous les documents pendant une période spécifique	X		X	$UT\{dD$
7	Toutes les annotations d'un utilisateur sur un document	X	X		$UD\{dT$
8	L'annotation d'un utilisateur sur un document pendant une période spécifique	X	X	X	UDT

Tableau 3.4 : Synthèse des exploitation des annotations dans le contexte d'IE

En changeant une ou plusieurs variables, ou en ajoutant ou enlevant quelques paramètres de document ou dans les paramètres de l'utilisateur, nous pouvons donner un nouveau regard à ce modèle. Cela peut nous permettre d'adapter le modèle pour un usage dans un certain nombre de disciplines.

Nous donnons dans les sections suivantes quelques exemples de domaines d'application.

### **3.6.2. Application dans le domaine bibliographique**

Nous pouvons appliquer ce modèle pour un usage bibliographique lorsque nous voulons suivre l'utilisation des documents ou l'habitude des utilisateurs dans une bibliothèque. Dans ce cas, nous voulons voir l'évolution dans l'utilisation du matériel publié ou l'évolution des utilisateurs. Nous devons seulement ajouter une ou deux propriétés matérielles publiées en tant que paramètres du document. Si il est possible d'enregistrer toutes les utilisations des documents par tous les utilisateurs de tout temps il est possible de suivre les évolutions d'utilisation des documents dans le temps.

### **3.6.3. Application dans le domaine d'administration juridique**

Comment interprétons-nous l'effet d'une loi promulguée? Comment pouvons-nous savoir si un ordre public est bien accepté? Le patrimoine collectif et la loi peuvent être considérés comme des documents et les réactions à ces derniers comme des annotations. Par exemple, nous pouvons appliquer ce modèle pour évaluer les impacts de la loi d'immigration de la France. Comment contrôlons-nous les problèmes des « banlieues »? Pouvons-nous considérer les « annotations » par tous les utilisateurs sur la loi? Dans le cas où nous avons des réactions publiques sur un ordre, il peut être intéressant de voir ce qui se produit dans le public. Il convient de noter que le public se compose de groupes divers et d'idéologies diverses. Certains membres du public peuvent être dans l'opposition, certains d'entre eux peuvent être du côté du gouvernement, dans l'opposition ou simplement neutres. En particulier les

commentaires sur une loi peuvent être d'une importance considérable après qu'une loi a été promulguée.

#### **3.6.4. Gestion de développement « Open Source »**

« Open source » est une méthode de développement de logiciel par plusieurs programmeurs géographiquement dispersés. Le logiciel est souvent distribué gratuitement.

Les contributions des programmeurs dans un système « Open Source » sont considérées comme des annotations sur les programmes existants. Nous pouvons identifier quatre éléments dans un développement « Open Source » (a) un utilisateur est un membre d'espace de collaboration (b) le répertoire de programmes (c) la base de données d'annotation (d) le temps.

Un utilisateur dans une espace de collaboration aura normalement son identification (codeutilisateur). Il utilise le code d'identification pour demander un programme particulier du dépôt de logiciel. L'utilisateur reconnaît le programme comme tâche commune s'il est un membre de la collaboration. Il est libre d'éditer ou de modifier le programme. Au lieu d'envoyer la modification de programme comme un remplacement dans le dépôt, nous proposons qu'il l'envoie comme paramètre de sa contribution dans la base d'annotation. Une nouvelle version du programme est créée avec un nouveau code de version tandis que la version ancienne du programme est maintenue dans le répertoire.

Le nouveau programme est déposé au répertoire avec un nouveau code de version. L'identité de l'utilisateur est stockée dans l'espace de travail avec la date, la période et les codes du programme sur lequel il a travaillé (l'ancien code et le nouveau code). La base de données d'annotation reçoit la proposition en tant qu' « annotation » ainsi que les paramètres liés à la proposition.

### **3.6.5. Application dans les travaux de ressources ouvertes**

Dans une ressource de domaine public, trois entités sont importantes pour nous : l'utilisateur, la ressource et l'activité commune. Plusieurs activités peuvent être importantes pour nous dans un système de ressource du domaine public. Certaines de ces activités incluent l'étiquetage (tagging), l'édition, la publication, la contribution et l'annotation. Nous pouvons favorablement comparer ces activités avec l'activité d'annotation dans un domaine public.

Les utilisateurs des ressources publiques sont généralement dispersés géographiquement. Ils sont souvent liés par la technologie comme l'Internet ou liés par un espace public (comme un panneau d'affichage). Le point commun chez les utilisateurs est l'intérêt qu'ils ont pour les ressources communes. Les ressources peuvent être une passion pour un objet ou une activité comme la photographie, la natation ou un document spécifique. En général, il n'y a aucune qualification ou chose exigée de l'utilisateur, excepté son intérêt pour l'objet en commun.

Nous caractérisons la zone de travail avec deux composants : le serveur de ressource et des activités en commun. Le serveur de ressource peut être l'Internet ou un système d'affichage tandis que l'activité en commun peut être unique ou multiple. Pour la simplicité, nous identifions des activités en commun en tant qu'étiquetage (tagging), modification, contribution et annotation. Nous présentons trois exemples pour montrer comment notre modèle peut être appliqué dans une zone de travail commun.

#### **3.6.5.1. WIKI**

L'une des ressources du domaine public sont des wikis. Un wiki est un type de site Web qui permet aux utilisateurs d'ajouter, d'enlever, d'éditer et de changer la plupart du contenu d'un site Web facilement, parfois sans la nécessité d'être inscrit dans le portail de ce site Web. Cette facilité d'interaction et d'opération met les wikis dans une position d'avantage pour être des outils efficaces pour l'écriture en collaboration.

Le terme wiki peut également se rapporter au logiciel en collaboration lui-même (moteur de wiki) qui facilite l'opération d'un tel site Web ou de certains sites Web spécifiques de wiki, y compris le site Web wiki, WikiWikiWeb, et les encyclopédies en ligne telles que Wikipedia.

Il existe des millions de wikis locaux qui sont construits autour des utilisateurs locaux sur des serveurs de Webs locaux qui exigent un compte personnel sur le serveur de systèmes de wiki. Un exemple de ces wikis locaux est <http://wiki.loria.fr>.

Quelques langages de programmation présentent, des plugins, des modules ou des API pour rendre la création des ressources de domaine public (comme des wikis) facile. TCL/TK a présenté ce qu'ils appellent le « wikit » qui est mis en application par le CGI ou autonome. Le projet de Javapedia est une exécution des wikis dans Java avec des classes telles que WebChanges, WebIndex, WebStatistics, WebPreferences, CreateANewPage.

Quelques wikis sont développés pour faciliter les discussions professionnelles au lieu de l'utilisation des blogs. Certains de ces derniers incluent, VBWiki, FoxWiki, SQLWiki, [sql.wikis.com](http://sql.wikis.com), [www.tiddlywiki.com](http://www.tiddlywiki.com), [wiki.wxpython.org](http://wiki.wxpython.org), etc...

### **3.6.5.2. DELICIOUS**

« [del.icio.us](http://del.icio.us) » est un site web permettant de sauvegarder et de partager ses pages Internet marqués et de les classer selon le principe de « folksonomie » par des tags. Il fut créé fin 2003 par Joshua Schachter dans le but de sauvegarder ses marque-pages personnels.

L'interface du site repose sur du HTML simple, ce qui rend le site facile d'utilisation. [del.icio.us](http://del.icio.us) propose également de syndiquer son contenu par RSS (Rich Site Summary) et repose sur la technologie des tags. Les tags, sous la forme d'un mot, par exemple: Sports, Cinéma, Internet... permettent de retrouver facilement les différents sites ayant un rapport avec le mot du tag. Les tags sont choisis par l'utilisateur lui-même, ce qui

lui permet de gérer entièrement ses marques pages.

Dans «delicious», un site Web d'intérêt est identifié comme le centre de communication pour un utilisateur. Ce Site Web a les propriétés d'URL, l'étiquette (tag) et l'annotation (commentaire). L'étiquette placée sur cet site Web peut être liée à une autre site Web utilisant la même étiquette ou une autre étiquette liée au même site Web. Par exemple, un emplacement peut être étiqueté (taggué) avec le mot « ordinateur », une autre personne peut étiqueter le même site Web avec le mot « technologie ». Il est à souligner que le même mot « technologie » peut être lié à un autre site Web. L'essence de ces étiquettes est de pouvoir voir les faisceaux des mots utilisés pour classifier les sites Web.

Les étiquettes (tags) sont simplement des labels pour URLs, choisis pour aider un utilisateur dans la récupération postérieure des ces URLs. Les étiquettes ont l'effet additionnel pour regrouper les URLs. Il n'y a aucun ensemble de catégories fixes ou de choix officiellement approuvés. Vous pouvez utiliser des mots, des acronymes, des nombres, n'importe quoi d'important pour un utilisateur, sans se demander si c'est important pour un autre.

Le problème avec ce type d'étiquetage est que les individus sont libres d'utiliser n'importe quelle langue ou symbole pour étiqueter un site Web. Il peut devenir très discordant de ne pas avoir un vocabulaire contrôlé pour étiqueter des sites Web. Nous pouvons avoir un centre de contrôle attaché au site Web pendant l'étiquetage. Ce centre de contrôle peut agir en tant que thesaurus pour contrôler l'utilisation de mots dans l'étiquetage. Les mots utilisés sont alors considérés comme les annotations.

### **3.6.5.3. Flickr**

Flickr est un site Web qui permet aux utilisateurs de télécharger, partager des photos et d'étiqueter chaque photo avec des mots descriptifs (Marlow et al, 2006). Les autres utilisateurs peuvent alors faire des recherches sur ces étiquettes, leur permettant de trouver et de présenter leurs observations sur les photos. La communauté active et les

dispositifs de partage de Flickr ont attiré des millions d'autres utilisateurs. En dehors du fait que les utilisateurs externes peuvent rechercher des photos dans la base de données, ils peuvent également avoir la permission de classer des photos qui appartiennent aux autres utilisateurs (Hidderley et Rafferty, 2006). Flickr est à la fois un site et un système d'échange de photos.

Flickr se fonde sur l'interprétation personnelle des images. Le problème avec un système comme Flickr est que les images sont interprétées et fondées sur la perception des couleurs et d'autres considérations environnementales. Par exemple, vous pouvez explorer des photos sur Flickr un article bien choisi comme les images classifiées en tant que « boy ». Les mots comme "girl", "kid", "eyes", "play", "woman", "people", "love" sont étiquetés avec "boy"<sup>32</sup>. Comment pouvons-nous connaître la relation entre deux concepts reposant sur les images ?

Un autre problème est le problème du vocabulaire et de l'orthographe. Par exemple le mot "fils" peut être utilisé pour qualifier un «garçon» ou « garçons ». Puisque nous étiquetons les images et non les textes, un autre problème inhérent est l'internationalisation. Comment réconcilions-nous des mots dans les langues différentes ?

Nous pouvons appliquer notre modèle d'annotation pour résoudre certains de ces problèmes. Dans ce cas, chaque étiquette est considérée comme une annotation et est envoyée à la base de données d'annotation. Les étiquettes sont envoyées à la base de données d'information seulement après qu'ils auront été contrôlés. Avec le concept de notre modèle, il sera possible d'identifier l'identité des utilisateurs apportant des contributions.

---

<sup>32</sup> <http://flickr.com/photos/tags/boy/clusters/>

### **3.6.6. Application globale**

Il est possible d'appliquer ce modèle à d'autres situations de gestion de l'information, de recherche de l'information et de travail en collaboration. Les conditions importantes sont de pouvoir associer l'objet central (cf. définition du document) avec les paramètres nécessaires et que les activités associées puissent être réduites à une sorte d'annotation si les conditions sont remplies. Nous pensons que les applications possibles de modèles sont nombreuses.

### **3.7. Le test et l'analyse du système I-AMIE**

Le premier objectif dans l'analyse de la mise en oeuvre du système I-AMIE était de comprendre la perception et la facilité d'utilisation auprès du public. Trois ensembles de personnes ont été invités à utiliser le système : les étudiants de première année de master en information scientifique et technique (b) les étudiants préparant une licence en informatique à l'université d'Ibadan, Nigeria et (c) le grand public. Le grand public inclut les utilisateurs ayant accès à Internet dans le monde entier.

Le questionnaire intitulé «Système d'annotation en IST (Intelligence économique) » a été envoyé aux utilisateurs du système I-AMIE. Le questionnaire avait pour objectif de connaître les problèmes liés à l'utilisation du système d'annotation sur <http://www.loria.fr/~robert/annot>. Les questions qui ont été présentées sont regroupées en à cinq groupes pour identifier la perception et connaître la possibilité d'améliorer du système. Les six groupes de questions sont :  
(a) Général (b) Contenu (c) Présentation (d) Efficacité (e) Appel et convenance et (f) Facilité d'utilisation.

Le questionnaire complète est présenter dans l'annexe 1.

L'entretien avec les utilisateurs montre que les utilisateurs n'ont pas suffisamment utilisé le système pour comprendre son fonctionnement. Ils pensent que la présentation des informations dans le système est logique et pertinents. En revanche,

ils ont suggéré d'ajouter quelques exemples de formulaires (remplir). Selon eux, le système provoque une compréhension profonde du domaine d'IE.

# Chapitre 4

*La vie est enregistrement humain. Les annotations peuvent être les enregistrements d'utilisations de la vie.*

## 4. Conclusions générale

### 4.1. Conclusion

Rappelons que notre hypothèse est que nous pensons que l'annotation devrait contribuer à la transformation de l'information collectée en des informations à valeur ajoutée qui seront plus adaptées pour la prise de décision. Associées à cette hypothèse, nous avons identifié les problématiques suivantes :

- a. La recherche de l'information, préalable à une prise de décision, peut-elle être renforcée par des outils d'annotation ?
- b. Peut-on guider, par un modèle approprié, le processus d'annotation qui peut soutenir la recherche d'information ?
- c. Comment les informations annotées peuvent-elles être collectées pour aider à la prise de décision ?
- d. Comment l'information annotée peut-elle être utilisée pour la prise de décision ?

Nous avons retenu quatre principaux composants dans le cadre de notre étude : le document, l'utilisateur, l'annotation et la période des annotations.

Un « document » a été défini dans un sens très large de sorte que notre étude puisse être appliquée à une grande variété de domaines. Le document a été défini comme un « conteneur » de l'information, précisément comme une trace des activités humaines. Pour que nous puissions formuler un modèle d'annotation pour la recherche d'information dans des processus d'IE, nous avons établi notre travail sur des documents électroniques écrits. Plusieurs de leurs caractéristiques ont été identifiées pour notre étude. Nous avons pris en compte le format et les autres caractéristiques

bibliographiques des documents.

Un document est la base de n'importe quelle annotation. Le mot « annotation » a été considéré comme un processus (une action) et également comme un objet. Nous avons montré que l'annotation est une référence à un document, et la majeure partie du temps une annotation se trouve sur le document de base. Le lien entre le document annoté et l'annotation ne constitue pas nécessairement un lien physique. Nous avons démontré que nous pouvons avoir différents types d'annotation tels que des graphiques et des écrits. Notre souci a porté sur l'annotation écrite. Nous avons montré que l'annotation peut être comparée à plusieurs autres terminologies notamment l'indexation. Les différences et les similitudes entre l'annotation et ces terminologies ont été présentées dans la section 1.4.

Nous avons souligné l'importance du rôle d'un utilisateur dans un processus d'annotation. L'annotation sans référence à son créateur n'a pas de valeur dans le processus de la prise de décision.

Nous considrons « l'utilisateur » d'un document du point de vue de lecture et de l'annotation. L'annotateur est d'abord un lecteur de document avant de devenir un annotateur. Nous ne nous sommes pas attachés au profil d'un créateur de document parce que ses caractéristiques n'ont pas d'importance dans notre travail. Nous avons précisé qu'un annotateur pouvait être le créateur du document annoté.

La « période » d'annotation est l'attribut temporel de l'annotation, exprimée par une date et heure. Le même utilisateur ne peut pas forcément faire le même type d'annotation à des périodes différentes.

Nous avons étudié deux différentes approches d'annotation : l'une utilisée pour la création et l'autre utilisée pour l'organisation du contenu d'annotation. Le fondement de ces approches porté sur des cadres et des modèles spécifiques d'annotation. Les cadres comme le Common Annotation Framework de Microsoft, ANNOTEA, Linguistic Annotation Framework et Extensible MultiModal Annotation framework

ne sont pas suffisants dans la création et l'organisation des annotations pour la recherche d'information. Nous avons examiné quelques modèles existants pour la création d'annotation, parmi lesquels certains sont fondés sur l'organisation du contenu d'annotation (ontologique, sémantique) et d'autres sur la technologie. Puisque nous n'avons pas trouvé un modèle satisfaisant pour notre besoin, nous avons formulé notre propre modèle.

Notre modèle s'appelle AMIE (Annotation Model for Information Exchange), le modèle d'annotation pour l'échange de l'information. Le modèle est fondé sur le fait qu'une annotation peut être faite pour prendre en charge les caractéristiques du document, de l'annotateur et la période de l'annotation. Nous croyons que ces caractéristiques sont suffisantes pour rendre une annotation exploitable dans le cadre de la recherche d'information pour la prise de décision. Chacune de ces caractéristiques a été considérée en détail. Nous avons proposé une architecture fondée sur ce modèle qui permet la création, le stockage et l'exploitation des annotations dans un environnement d'IE. Nous avons développé les bases de données associées.

Nous avons mis en application ce modèle sur une base de données bibliographiques disponible dans l'équipe de recherche SITE-LORIA. Nous avons noté que ce modèle peut être appliqué dans plusieurs secteurs. L'application générale du modèle considère un objet en tant qu'un document et son utilisation ou les observations sur le document comme les annotations. L'application de ce modèle peut être classifiée dans les groupes suivants :

- Système d'évaluation,
- Système de rétroaction,
- Système de collaboration,
- Système pour le partage des expériences par les annotations,
- Système de partage des ressources distribuées,
- Système d'aide à la décision,
- Système d'observation des usages des documents.

Nous avons également étudié certaines applications spécifiques de ce modèle comme dans le cas de delicious, flickr et wikis.

## 4.2. Perspectives

### Ecologie d'annotation

Certaines autres considérations non négligeables dans la création et l'utilisation de l'annotation pour la recherche de l'information peuvent inclure les questions suivantes :

- Où les documents sont-ils utilisés ? (localisation)
- Quels sont les objets nécessaires pour rendre le document utilisable ? (technologie)
- Quel traitement serait nécessaire pour rendre le document utilisable ? (processus)
- Quel a été l'état initial et final du document avant et après un usage ?

Il peut être nécessaire de savoir exactement où un document a été utilisé. En effet, quelques influences culturelles et géographiques peuvent être impliquées dans l'utilisation d'un document. Par exemple, tablant sur notre définition générale du document et de l'annotation, nous pouvons étudier les annotations parmi les sociétés musulmanes. Pour eux, il est interdit d'écrire sur le coran. Serait-il possible d'intégrer toutes nos propositions si nous considérons le document « coran » dans les pays musulmans ?

Nous ne pouvons pas indiquer comment une annotation sur un document sera utilisée mais il peut être important de savoir exactement comment un document a été utilisé. On peut utiliser un document pour demander une action ou juste pour l'information. Un document peut être utilisé pour poser une question etc. Il peut être important d'énoncer avec précision le contexte de l'utilisation d'un document dans certains cas.

Certains types de document ne peuvent être utilisés seuls sans connaître d'autres documents qui sont nécessaires à leur compréhension. Comme nous le savons, les

livres sont en rapport avec d'autres documents. Les documents secondaires (comme les catalogues dans les bibliothèques) dépendent de leurs sources primaires. Ceci implique que le document secondaire ne peut être utilisé seul dans certaines circonstances.

Quel traitement serait nécessaire pour rendre un document utilisable ? Nous pouvons être intéressés par les processus qu'un document utilisé a subis avant qu'il puisse être utilisé dans le processus d'une annotation. Dans la plupart des cas, nous préférons une situation où rien n'est fait sur le document original avant son utilisation. Il est important de noter que ce ne sont pas tous les documents qui sont dans cette situation. Il y a des documents qui doivent être traduits d'une langue à l'autre avant leurs utilisations.

Nous devrions également pouvoir indiquer l'état initial et l'état final d'un document avant et après son utilisation.

### **Diversité de points de vue sur les documents**

Ce travail n'a pas essayé d'appliquer certains des résultats fondamentaux dans certaines disciplines comme les mathématiques et la bibliométrie. L'attention a été portée sur la façon dont la recherche d'information pour l'IE peut être améliorée avec un modèle d'annotation. Un regard critique sur l'annotation demande une application de technologie sémantique. Par exemple, nous avons établi que deux individus n'annoteront pas le même document la même manière. Deux individus peuvent parfois annoter le même document différemment simplement de par les terminologies utilisées. Par exemple, un utilisateur peut annoter un document avec le mot « géographie ». Ce même document peut être annoté avec le mot « science de la terre » par un autre utilisateur. Dans ce cas, la base d'annotation contiendra le mot « géographie » et « science de la terre ». Les deux mots sont essentiellement les mêmes. Il est donc impératif que les travaux ontologiques et sémantiques puissent être adaptés pour augmenter les résultats de ce travail.

Une autre point qui peut être considéré dans les processus d'annotation concerne les questions liées au langage. Comment pouvons-nous réconcilier le problème de syntaxe de langue pour augmenter le stockage et la récupération de l'information dans une base de données d'annotation ? Typiquement, la morphologie de la langue peut considérablement influencer la signification de l'information dans les bases d'annotation. En effet, un problème caractéristique d'annotation est un langage utilisé par l'annotateur. Par exemple, dans la mise en oeuvre de ce travail faite en anglais et en français, nous pouvons avoir deux utilisateurs différents qui utilisent deux langues différentes. Comment réconcilions-nous les mots en provenance de langues différentes ?

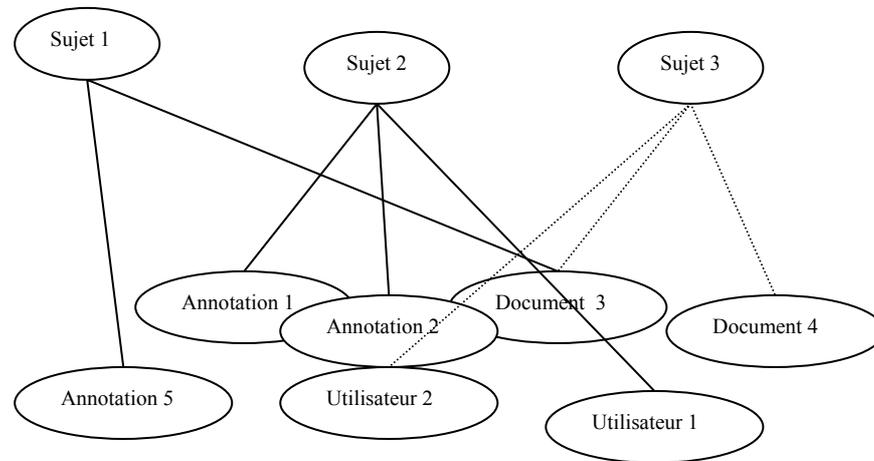
Nous pouvons également vouloir distinguer toutes les convergences ou divergences dans les annotations d'un document. Par exemple, comment pouvons-nous utiliser la convergence de l'annotation ou ses divergences pour authentifier les interprétations et informations ajoutées par des différents utilisateurs d'un document ?

### **Exploitation de la base d'annotations**

Nous ne nous sommes pas intéressés à la façon d'analyser les résultats issus de la question envoyée à une base de données d'annotation en terme de pertinence. Bien sûr nous avons précisé que nous pouvons utiliser SQL pour évaluer le contenu de la base de données d'annotation. Il est même possible d'utiliser des outils comme <http://metiore.loria.fr> pour exploiter la base de données d'annotation. Nous soulignons que le contenu du résultat est laissé à la discrétion de l'utilisateur et d'un travail supplémentaire.

Des approches différentes ont été suggérées comme la façon de rechercher de l'information dans une base de données en exploitant une base d'annotations. L'application de Topic Maps peut être intéressante dans la recherche d'information dans la base d'annotations. Nous croyons que quand la base de données d'annotation devient grande elle peut être difficile de trouver exactement l'information cherchée. Nous présentons ce que peut être le scénario probable de la recherche d'information

dans une base d'annotations avec Topic Maps.



*Figure 4.1 : La mise en correspondance des sujets avec les utilisateurs et les annotations*

Si le concept de topic maps est appliqué dans la recherche d'information dans les bases d'annotation, il sera possible de voir d'un coup d'oeil le nombre d'annotations par sujet particulier. Nous pouvons également voir comme sur le schéma *figure 4.1*, les utilisateurs qui sont intéressés par un sujet. Nous voyons dans le schéma comment les sujets sont associés aux annotations ou les sujets avec les utilisateurs. La différence entre cette approche et notre système est que, les sujets sont indépendants des attributs.

Une autre possibilité qu'on pourrait appliquer est le concept des SUMS (System for Universal Media Searching) pour la recherche dans les bases d'annotations. Comme point de départ SUMS utilise le principe des questions: Quoi ? Qui ? Où ? Quand ? Comment ? et Pourquoi ? Le « Qui » mène à la connaissance des personnes. La question « Quoi » mène aux objets ; La question « Où » mène aux endroits. La question « Quand » mène aux événements et aux schémas chronologiques. La question « Comment » mène aux instructions. Les questions « Pourquoi » mènent aux raisons et aux causes. Le choix d'une question mène aux listes, dont les sous-ensembles sont alors choisis pour atteindre d'autres listes et les ordres.

## **Situation actuelle dans la perspective**

Nous avons pu démontré les utilisations possibles de ce modèle à plusieurs domaines. Notamment, nous avons eu des publications dans des conférences pour démontrer les applications possibles du modèle. Une publication avec un professeur d'université de Grand Valley State, Allendale, Michigan, Etats-Unis a été réalisée, portant sur l'application de notre modèle pour l'évaluation des participations des étudiants dans une tâche de programmation en commun (Evaluating Student Participation in Open Source Software Development with an Annotation Model, *The Fourth IASTED International Conference on Knowledge Sharing and collaborative engineering*, KSCE 2006, November 29-December 1, 2006, St. Thomas, US Virgin Islands, USA, 2006).

Nous recherchons également une manière d'évaluer le contenu de l'annotation en utilisant des concepts de « Topic Maps » intégrant SUMS (System for Universal Multimedia Searching).

Nous n'avons pas pu développer tous les modules nécessaires pour l'évaluation du système I-AMIE pour deux raisons principales :

- Pour évaluer le système, nous avons besoin des annotations par plusieurs annotateurs et pendant une période assez longue. Vu les contraintes de temps nous n'avons pas pu avoir plusieurs annotations d'un même annotateur sur une période suffisamment longue. Ainsi, nous n'avons pas pu observer l'évolution des habitudes des annotateurs.
- La programmation est techniquement lourde à réaliser, ce qui nécessite davantage de temps de développement.

Notre hypothèse était que l'annotation devrait contribuer à la transformation de l'information collectée en des informations à valeur ajoutée qui seront plus adaptées pour la prise de décision.

Nous avons prouvé avec notre modèle AMIE et le système I-AMIE que c'est possible de contribuer à la transformation de l'information collectée par les annotations pour donner de la valeur ajoutée aux informations collectées pour la prise de décision (dans chapitre 3).



# Bibliographies

(ACSIOME, 1989)

ACSIOME, 1989, Modélisation dans la conception des systèmes d'information avec exercices commentés, MASSON, Paris

(Almeida, 1995)

D'almeida Fabrice, 1995, *Images et propagande*, Castennan-Giunti, Pages 91-95

(Ansoff, 1965)

I. Ansoff, 1965, Corporate Strategy: An Analytic Approach to Business Policy for Growth and Expansion. McGraw-Hill.

(Bargeron et al, 2001)

Bargeron David, Gupta Anoop, and Brush Bernheim A.J., 2001, A Common Annotation Framework, Technical Report MSR-TR-2001-108, *Microsoft Research*, Microsoft Corporation, Redmond, W A 98052

(Bérubé, 1991)

Bérubé Louise, 1991, Terminologie de neuropsychologie et de psychologie du comportement, Les Éditions de la Chenelière, Montréal, Pp 176

(Bird et al, 2000)

Bird Steven, Day David, Garofolo John, Henderson John, Laprun Christophe and Libennan Mark, 2000, ATLAS: A Flexible and Extensible Architecture for Linguistic Annotation, *Proceedings of the Second International Conference on Language Resources and Evaluation*, pp. 1699-1706, Paris: European Language Resources Association, 2000

(Bouaka, 2004)

Bouaka Najoua, 2004, *Développement d'un modèle pour l'explicitation d'un problème décisionnel : un outil d'aide à la décision dans un contexte d'intelligence économique*, Thèse doctorat de l'université Nancy 2, Sciences de l'information et de la communication, PP 209

(Bouaka and David, 2003)

Bouaka Najoua, and David Amos, 2003, Modèle pour l'Explicitation d'un Problème Décisionnel : Un outil d'aide à la décision dans un contexte d'intelligence économique. Editor: Amos, David, *Intelligence Economique : Recherches et Applications* Nancy, 14, 15 avril 2003, URL: <http://www.inist.fr/iera/fichiers/ierall.pdf>

(Bourgeois, 1980)

L. J. Bourgeois. Strategy and environment: a conceptual integration. *Academy of management review*, vol. 5, n°1, p. 25-39, 1980.

(Bourret, 2004)

Bourret Christian, 2004, Data concerns and challenges in health: Networks, information systems and electronic records, *Data Science Journal*, Volume 3, 17

September 2004, Pp 96-113.

(Boutelle, 2003)

Boutelle, Steven W., 2003, "Anny Network Centric Operations and Warfare: Putting Power to the Edge" MILCOM 2003 Transformational Strategies Panel. Boston, MA. 16 October 2003,  
[http://www.afcea.org/pastevents/milcom2003/Boutelle\\_files/09/09/2006](http://www.afcea.org/pastevents/milcom2003/Boutelle_files/09/09/2006)

(Boutelle, 2004)

Boutelle Steven W., 2004, Making the Move to a Network Centric Army, AUSA Institute of Land Warfare Breakfast, 09 Sept 2004,  
<http://www.army.mil/CIOG6/references/briefings/Boutelle/MILCOM2004Keynote.ppt>

(Bringay et al, 2004)

Bringay Sandra, Barry Catherine, and Charlet Jean, 2004, Les documents et les annotations du dossier patient hospitalier, *Information-Interaction-Intelligence*, Volume 4, n°1, Juillet 2004

(British Police, 2003)

British Transport Police, 2003, Effective Approaches to the Removal and Prevention of Vandalism and Graffiti on Public Transport- Final Report: The Case Studies and Emerging Issues, Department for Transport Prepared by Crime Concern: London, Pp 54

(Brusilovsky, 1996)

Brusilovsky Peter, 1996, Efficient technique for adaptive hypermedia, Intelligent hypertext: Advances techniques for the world wide web. Lecture notes in computer science, 1326, Berlin, Springer-Verlag, 12-30

(Brush et al, 2002)

Brush B., Bargeron, D., Grudin, J., and Borning A., Gupta, A., Supporting Interaction Outside of Class: Anchored Discussion vs. Discussion, online, *Proceedings of Computer Support for Collaborative Learning (CSCL 2002)*, <http://newmedia.colorado.edu/cscl/>, 2002, visited on January 2003

(Buckland, 1998)

Buckland Michael, 1998, What is a document ? *Journal of the American Society for Information Science*, Volume: 48, (9) Pp: 804-809

(Cadiz et al, 2000)

Cadiz U, Gupta Anoop et Grudin Jonathan, 2000, Using Web Annotations for Asynchronous Collaboration Around Documents, *CSCW'00*, Philadelphia, Microsoft Research, Collaboration & Multimedia Group, Date: December 2-6, 2000

(Campbell, 2003)

Campbell Eoin, 2003, Creating Accessible Online Content Using Microsoft Word, XML Workshop Ltd: Irish Learning Technology Association EdTech 2003, Waterford, Ireland, Pages: 1- 10, Date: 23May, 2003

- (Carayon, 2003)  
Carayon Bernanrd, 2003, Intelligence économique, compétitivité et cohésion sociale Pp: 79, Date: June 2003
- (Carter et al, 2004)  
Carter Scott, Churchill Elizabeth, Denoue Laurent, Helfman Jonathan and Nelson Les, 2004, Digital Graffiti: Public Annotation of Multimedia Content, CHI 2004, April 24-29, 2004, Vienna, Austria. <http://www.fxpal.com!publications/FXPAL-PR-03-263.pdf>
- (Casallas-Gutierrez, 1996)  
Casallas-Gutierrez Rubby, 1996, *Objets historiques et annotations pour les environnements logiciels*, Thèse de Doctorat à l'Université Joseph Fourier Grenoble I, May 1996.
- (Cauvin et al, 1994)  
Cauvin, P. and Cailloux, G., 1994, Les types de personnalités: La méthode MBTI Edited by ESF, ESF - Psychologie - Psychanalyse, Paris, Pg 11-32
- (Chien et al, 2001)  
Chien Shu-Yao, Tsotras Vassilis J. and Zaniolo Carlo, 2001, Efficient Management of Multiversion Documents by Object Referencing, *Proceedings of the 27th VLDB Conference*, Roma, Italy, 2001
- (Chou, 2004)  
Chou Wu, Dahl Deborah A., Johnston Michael, Pieraccini Roberto and Raggett Dave, 2004, EMMA: Extensible MultiModal Annotation markup language, W3C Working Draft 14 December 2004
- (Clay, 2005)  
Clay, Shirky, 2005, Ontology is overrated: Categories, Links and Tags, *Journal of Networks, Economics and Culture*, May 31, 2005
- (Coadic, 1994)  
Coadic, Yves-François. Le, 1994, La science de l'Information, PUF, Paris
- (CEE/ONU, 1995)  
Communication des délégations Autrichienne et Suisse, 1995, Recommandation CEE/ONU no 25, Relative à l'utilisation de la norme des nations unies concernant l'échange de données informatisé pour l'administration, le commerce et le transport (edifact/onu), Nations Unies, Pages: 1-9, 18-19 septembre 1995
- (Cook, 2004)  
Cook John, Leathwood Carole and Oriogun Peter, 2004, Online conferencing with multimedia students: monitoring gender participation and promoting critical debate *The e-Journal of the LTSN-ICS*, <http://www.ics.ltsn.ac.uk/publications/issue2/cook/006.html> (24/04/2006)

- (Corbley, 2005)  
 Corbley, Kevin P., 2005, Spatial Information Management  
[http://www.military-geospatial-technology.com!article.cfm ?DocID=861](http://www.military-geospatial-technology.com!article.cfm?DocID=861) (20/02/06)
- (Crossen et al, 2001)  
 Crossen, A., Budzik, J., Warner, M., Birnbaum, L., and Hammond, K., 2001, XLibris: An Automated Library Research Assistant, *Proc. 2001 International Conference on Intelligent User Interfaces*,  
 URL:<http://www.infolab.northwestern.edu/infolab/downloads/papers/paper10103.pdf>
- (Darmon, 1993)  
 Darmon, R. Y., 1993, Management des ressources humaines des forces de vente, Economica, Paris
- (Darses, 1994)  
 Darses, F., 1994, Gestion des contraintes dans la résolution de problèmes de conception, Université de Paris VIII, Saint-Denis.
- (Davis et al, 1995)  
 Davis James R. and Huttenlocher Daniel P., 1995, Shared Annotation for Cooperative Learning,  
[http://www .cs.cornell.edu/home/dphlannotation/annotations.html](http://www.cs.cornell.edu/home/dphlannotation/annotations.html) (14/07/2005)
- (David et al, 2001)  
 David Amos., Bueno David and Kislin Philippe, 2001, Case-Based Reasoning, User model and IRS. *The 5th World Multi-Conference on Systemics, Cybernetics and Informatics - SCI2001. International Institute of Informatics and Systemics (IIIS)*. (Orlando, USA). 2001. [http://isdml.univ-tln.ftt/PDF/isdml11isdml1a98\\_amos.pdf](http://isdml.tln.ftt/PDF/isdml11isdml1a98_amos.pdf), 31/01/2005
- (David, 2005)  
 David Amos L'Intelligence Économique et Les Systèmes d'Information: Problématiques et approches de solutions, in "Veille stratégique: Mise en oeuvre et valorisation de la veille stratégique en entreprise", Algérie Télécom, Alger, Algérie, Juin 2005
- (Davison, 1995)  
 Davison Robert M., 1995, A Survey of Group Support Systems: Technology and Operation, Dept of Information Systems, City University of Hong Kong. Hong Kong, Pages: 1-38
- (Delecroix , 2005)  
 Delecroix Bertrand, 2005, La mesure de la valeur de l'information en Intelligence Economique: Application à la mise en place de solutions pour accroître la plus-value d'information élaborée dans le contexte d'un intra net, Thèse doctorat de Université de Marne-la-Vallée, Information Scientifique et Technique,  
[http:// quoniam. univ-tln.ftt/jurys/Delecroix/DelecroixThese.pdf](http://quoniam.univ-tln.ftt/jurys/Delecroix/DelecroixThese.pdf) 11/08/2006
- (Denoue et al, 2000)

Denoue, L. and Vignollet. L., 2000, An annotation tool for Web browsers and its applications to information retrieval. *Proceedings of RIA 0200, Apr. 2000*. <http://www.fXpa1.com/people/denoue/publications/ria0200.pdf>. 31/03/2005

(Conseil Switzerland,2005)

Département Fédéral des Affaires Etrangères, 2005, Rapport annuel 2004 du Conseil fédéral sur la participation de la Suisse au Conseil de partenariat euro-atlantique et au Partenariat pour la paix, Partnership for Peace, Switzerland

(Desmontils et al, 2003)

Desmontils E., Jacquin C, and Simon L., Vers un système d'annotation distribué, 2003, <http://www.sciences.univ-nantes.fr/irin/Nie!RR/RR-IRIN2003-01.pdf> 17/07/2004

(Deweese, 1995)

Deweese Peter A., 1995, Social and economic incentives for smallholder tree growing: A Case Study from Murang'a District, Kenya, Food and Agriculture Organization of the united nations, Uppsala, Sweden

(Dohar, 2003)

Dohar, Vic, 2003, Adding and Editing Annotation Using GEMS, Natural Resources Canada, ESS Info Division, Publishing Services Subdivision, Pages: 1-10

(Dou et al, 2005)

Dou H., Leveillé, V., Manullang, S. and Dou JM Jr, 2005, Patent analysis for competitive technical intelligence and innovative thinking, *Data Science Journal*, Volume 4, 31 December 2005, Pg 209-236, <http://www.jstage.jst.go.jp/article/dsj/4/0/209/ydf>

(Dou et al, 1987)

Dou H., Hassanaly P., Latela A., and Milon M., 1987, Etudes de cas: le traitement de l'IST par les indicateurs scientométriques, *Bulletin des bibliothèques de France*, Vol. 32, No 3, pp. 220-226

(Dubois et al, 1997)

Dubois Didier, Fargier Hélène, Prade Henri, 1997, Decision-Making under Ordinal Preferences and Comparative Uncertainty, *Journal of Uncertainty in Artificial Intelligence*

(Dumas, 2006)

Dumas Philippe, 2006, V<sup>E</sup>TIC & Territoires : Quels développements ? *Territoire & Mondialité*, Université de Franche comté, Besançon, 9-10 juin 2006, <http://isd.m.univ-tln.fr/PDF/isd.m26/4.Dumas.pdf>

(Ellul, 1967)

Ellul Jacques, 1967, Histoire de la propagande. Paris : Presses Universitaires de France, 1967

(Encyclopedica, 2006)

Encyclopedria Britannica, 2006, Physical science, Encyclopedria Britannica Premium Service. URL: <http://www.britannica.com/eb/article?tocId=32545>. 08/06/2006

(Farrell et al, 1992)

Farrell R.A., C.J. Pearson, and L.C. Campbell, 1992, Relational databases for the design and construction of maintainable systems in agriculture. *Journal of Agricultural Systems*, Volume: 38, Pages: 411-423.

(Farrell et al, 2005)

Farrell C., Ross and Maness R. Thomas, 2005, A relational database approach to a linear programming-based decision support system for production planning in secondary wood product manufacturing, *Journal of Decision Support Systems*, Volume: 40, Issue: 2, Pages: 183-196, Date: August 2005

(Francica, 2004)

Francica J oe, 2004, Intergraph's GeoMedia WebMap Publisher - Product Overview Date: May 29,2004, URL: <http://www.locationintelligence.net/articles/S8S.html>

(Futtersack et Bolf, 1999)

Futtersack Philippe and Bolf Didier, 1999, XML and related standards for data warehouses Paris, URL: <http://www.infloom.com/lgcaconfs/WEB/philadelphia99/futtersack.HTM#NI>

(Garshol, 2002)

Garshol, Lars Marius, 2002, What Are Topic Maps URL: <http://www.xml.com/pub/a/2002/09/11/topicmaps.html> 07/04/06

(Geyer-Schulz et al, 1999)

Geyer-Schulz Andreas, Koch Stefan and Schneider Georg, 1999, Virtual Notes: Annotation on the WWW for Learning Environments, *Proceedings of the 5th Americas Conference on Information Systems*, Pages: 136-138, <http://citeseer.ist.psu.edu/263780.html> 24/03/2006

(Gonzalez-Barahoma, 2000)

Gonzalez-Barahoma Jesus M., 2000, Free Software/Open Source: Information Society opportunities for Europe? Working group on Libre Software, version 1.2, <http://eu.connecta.it>

(Gonzalez, 2005)

Gonzalez Guadamuz Andrés, 2005, Legal challenges to open source licences, 2:2 SCRIPT-ed 301, <http://www.law.ed.ac.uk/ahrb/script-ed/vol2-2/challenges.asp> 19/07/2006

(Goria et al, 2004)

Goria, Stéphane and Geffroy, Philippe, 2004, Le modèle MIRABEL : un guide pour aider à questionner les Problématiques de Recherche d'Informations. *Veille Stratégique Scientifique et Technologique - VSST'2004 (Toulouse)*, 2004. 5p.

(Goria, 2006)

Goria Stéphane, 2006, *L'expression du problème dans la recherche d'informations : application à un contexte d'intermédiation territoriale*, Janvier 2006, Thèse doctorat de l'université Nancy 2, Sciences de l'information et de la communication, Pp 6S6

(Gupta et al, 2001)

Gupta, David Barger; A.J. Bernheim Brush; Anoop, 2001, A Common Annotation Framework, *Microsoft Research*, Redmond, Pages: 1-22, November 2001

(Gustavo et al, 1998)

Gustavo Alonso, Fabio Casati, Harumi Kuno and Vijay Machiraju, 1998, Web Services: Concepts, Architectures and Application, Springer-Verlag, Berlin, Pg: 115-120

(Hammer et al, 1995)

Hammer Joachim, Garcia-Molina Héctor, Ireland Kelly, Papakonstantinou Yannis, Ullman Jeffrey, and Widom Jennifer, 1995, Information translation, mediation, and mosaic-based browsing in the TSIMMIS system, *Proceedings of the 1995 ACM SIGMOD Record*, v.24 n.2, p.483, May 1995, International Conference on Management of Data

(Handschuh et al, 2003)

Handschuh S., Staab S., and Volz R., 2003, On deep annotation, *Proceedings of International World Wide Web Conference*, Pp 431-438.

(Handschuh and Staab, 2003)

Handschuh S. and Staab S., 2003, Annotation for the semantic Web, *Frontier In Artificial Intelligence and Application*, Vol 96, IOS Press, Amsterdam, Pp 46-60.

(Harris, 2003)

Harris, Nomi L., Genotator: 2002, A Workbench for Sequence Annotation, Lawrence Berkeley National Laboratory, Human Genome Informatics Group Berkeley

(Harvard, 2004)

Harvard Family Research Project: Harvard Graduate School of Education, 2004, Detangling Data Collection: Methods for Gathering Data. Out of School Time Evaluation Snapshots, number 5, Pages: 1-6  
<http://www.gse.harvard.edu/hfip/content/projects/afterschoollresources/snapshot5.pdf> 24/03/06

(Hayes et al, 1998)

Hayes, J. and Allinson, C. W., 1998, Cognitive style and the theory and practice of individual and collective learning in organisations, *Journal of Human relations*, Volume: 51, Issue: 7

(Heck et al, 1999)

Heck R. M., Luebke S. M., and Obermark C. H., 1999, A Survey of Web Annotation Systems,  
[http://www.math.grin.edul-rebelsky/Blazers/Annotations/Summer1999/Papers/survey\\_pa\\_per.html](http://www.math.grin.edul-rebelsky/Blazers/Annotations/Summer1999/Papers/survey_pa_per.html) 10/04/2006

(Hearst et al, 1998)

Hearst M.A., Schölkopf B., Dumais S., Osuna E., and Platt J.. 1998, Trends and Controversies - Support Vector Machines, *IEEE Intelligent Systems*, 13(4): 18-28, 1998.

(Hidderley et al, 2006)

Hidderley Rob and Pauline Rafferty, 2006, *Flickr* and Democratic indexing: Discipling desire lines, *Advances In Knowledge organization*, Volume: 10, Pp: 405-411, 4 - 7 July 2006

(Heiner et al, 1999)

Heiner Jeremy M., Hudson Scott E., Tanaka Kenichiro, 1999, The Information Percolator: Ambient Information Display in a Decorative Object, *ACM Symposium on User Interface Software and Technology*, Issue: November, Pp: 141-148

(Horrocks et al, 2000)

Horrocks, D., Fensel, J. Broekstra, S., Decker, M., Erdmann, C., Goble, F., van Harmelen, M. Klein, S. , Staab, R., Studer, and E. Motta, 2000, The Ontology Inference Layer OIL,  
<http://www.cs.man.ac.uk/~horrocks/Publications/download/2000/oil.pdf>

(Horrocks, 1998)

Horrocks Ian, 1998, Using an expressive description logic: FaCT or fiction? *Proceedings of KR-98*, Pages: 636-647

(Horrocks et al, 1999)

Horrocks Ian, Sattler U., and Tobies S., 1999, Practical Reasoning for Expressive Description Logics, *Proc. 6th Int '1 Conf Logic for Programming and Automated Reasoning (LPAR 99)*, Volume: 1705, Pp: 161-180, Lecture Notes in Computer Science

(Ide et al, 2003)

Ide Nancy, Romary Laurent and de la Clergerie Eric, 2003, International Standard for a Linguistic Annotation Framework,  
URL: <http://acl.ldc.upenn.edu/W/W03/W03-0804.pdf>

(loyal et al, 1996)

loyal Paul M., 1996, Industrial espionage today and information wars of tomorrow, *19th National Information Systems Security Conference*, Baltimore

(Kahan et al, 2001)

Kahan José, Koivunen Marja-Ritta, 2001, Annotea: an open RDF infrastructure for shared Web annotations, *Proceedings of the IOth international conference on World*

*Wide Web*, Hong Kong, Pages: 623 - 632

(Kahan,2001 b)

Kahan, I., Koivunen, M., Prud'Hommeaux, E., and Swick, R. 2001, Annotea: An Open RDF Infrastructure for Shared Web Annotations, *Proceeding of the WWWIO International Conference*, Hong Kong, May 2001, URL: <http://www10.org/cdrom/papers/488/index.html>

(Kislin et al, 2003)

Kislin P., David A., Peguiron F., 2003, Caractérisation des éléments de solutions en recherche d'information: conception d'un modèle dynamique dans un contexte décisionnel, *4ème Congrès ISKO-France. L'organisation des connaissances: approches conceptuelles*. (Grenoble, France). 2003. Grenoble, Pages: 13

(Kislin et David, 2003)

Kislin Philippe and David, Amos, 2003, De la caractérisation de l'espace-problème décisionnel à l'élaboration des éléments de solution en recherche d'information dans un contexte d'intelligence économique: le modèle WISP. *Conférence Intelligence Economique: Recherches et Applications*, Edited by Amos, David, IERA'2003. (INIST, France), Nancy

(KnowTIPS, 2006)

KnowTIPS, 2006, Spectrum College and Consulting Services Ltd, KnowTIPS Online Conference, <http://knowplace.ca/moodle/mod/resource/view.php?id=4606> (13/06/06)

(Knauf et David, 2004)

Knauf Audrey and David Amos, 2004, Vers une meilleure caractérisation des rôles et compétences de l'infomédiaire dans le processus d'intelligence économique. *Veille Stratégique Scientifique et Technologique - VSST'2004* (Toulouse). Poster

(Kolmayer et Pevrelong, 1999)

Kolmayer Elisabeth, Pevrelong Marie-France, 1999, L'émergence du document dans un processus de capitalisation des connaissances, *Communication aux Journées "Org & Co", SFSIC*, Aix-en-Provence, 3-5, June 1999

(Kornacker et Gilstrap, 1997)

Kornacker Marcel and Gilstrap Ray, 1997, An Annotation Service for the World Wide Web CS294-6 Project <http://ratogi.net/archive/unused/class/cs294-6/> (27/05/2004)

(Kumar, 2001)

Kumar A., 2001, "Third Voice Trails Off...", [http://www.wired.com/news/business/O.1367\\_42803,00.html](http://www.wired.com/news/business/O.1367_42803,00.html), Wired News, April 2001

(Lebow et Lick, 2001)

Lebow D. G. and Lick D., 2001, HyLighting: A new approach to distributed learning. *Annual convention of the Association for Educational Communications*

*and Technology*, Atlanta, GA, November 7-10, 2001

(Lebow et al, 2003)

Lebow, D., Lick, G., and Marks, P., 2003, Introducing Hylighting: Learning to learn with interactive annotation technology, *Florida Educational Technology Conference*, Orlando, 2003, February

(Leidner et Elam, 1993)

Leidner, D.E. and Elam J.J., 1993, Executive information systems: Their impact on executive decision making, *Journal of Management Information Systems*, Winter 1993-94, vol110, Issue: No. 3, Pages: 139-155.

(Lemyre et Willemant , 2006)

Lemyre Pierre-Paul and Willemant Richard, 2006, The legal issues surrounding free and open source software: Challenges and solutions for the government of Québec, <http://www.cirano.qc.ca/pdf/publication/2006RP-04.pdf>

(Lewis et Keleman, 1989)

Lewis, L.F. and Keleman, K.S., 1989, Experiences with GDSS Development: Lab and Field Studies, *Journal of Information Science Principles and Practice*, Volume: 16 Issue: 3, Pages: 195-205

(Léveillé, 2000a)

Léveillé, Valérie, 2000, *De l'organisation des données dans les systèmes d'information : Réalisation d'un outil de gestion de données hétérogènes et formelles appliqué à la veille technologique*, Doctorat en sciences de l'université de droit, d'économie et des sciences d'Aix-Marseille, Marseille, Pages: 192

(Lionel, 1910)

Lionel, Giles M.A., 1910, *The Art of war of Sun Tzu*, The oldest military Treatise in the world, translated from Chinese with an introduction and critical notes, Puppet press classic

(Liu et al, 1999)

Liu, K.D., Wu, H.Q., Pong, Y.J., Liu, K.C. and Guo, C.Q., 1999, *Unascertained Information, Mathematical Treatments and Applications*, Scientific Press, Beijing

(Lortal et al, 2005b)

Lortal G., Lewkowicz M., Todirascu-Courtier A., 2005, Annotation: textual media for cooperation, *International Workshop on Annotation for Collaboration*, Paris, November, 24-25, 2005

(Lortal et al, 2005b)

Lortal Gaëlle, Lewkowicz Myriam and Todirascu-Courtier Amalia, 2005, Modélisation de l'activité d'annotation discursive pour la conception d'un collectif support à l'herméneutique; *Jaulent M-C. Actes de la conférence JC 2005*, Pg 169 - 180

(Maes et Potter, 2003)

Maes Stéphane H. and Potter Stephen, 2003, Requirements for EMMA: W3C Note 13 January 2003, URL: <http://www.w3.org/TR/2003/NOTE-EMMAreqs-20030113/>

(Marlow et al, 2006)

Marlow Cameron, Naaman Mor, Boyd Danah and Davis Marc, 2006, Position paper, tagging, taxonomy, flickr, article, ToRead, *WWW 2006 Collaborative Web Tagging Workshop*, Edinburgh, Scotland

(Martinet, 1989)

B. Martinet et J. Ribault. La veille technologique concurrentielle et commerciale. Les éditions d'Organisation, 1989.

(Martre, 1994)

Martre, H., Intelligence économique et stratégie des entreprises, Rapport du commissariat Général au Plan, Paris, *La documentation Française*, 1994, pp 17,18

(Martin, 2004)

Martin David, 2004, EDIFACT: the international standard for EDI Contribution to eLib Study on Document Requesting Standards  
URL: [www.ukoln.ac.uk/dlis/models/models l/serials-mtg/serials-req-edi.doc](http://www.ukoln.ac.uk/dlis/models/models%20l/serials-mtg/serials-req-edi.doc)

(McGonagle et Vella, 1990)

McGonagle, J.J. and Vella, C.M., 1990, Outsmarting the Competition: Practical Approaches to Finding and Using Competitive Information, Sourcebooks, Naperville, Illinois

(McNurlin et Sprague, 1989)

McNurlin B.C. and Sprague R.H. 1989, Information systems in practice, 2nd Edition, Upper Saddle Miner, US, Prentice-Hall International

(Mehdi et al, 2005)

Mehdi Achour, Friedhelm Betz, Antony Dovgal, Nuno Lopes, Philip Olson, Georg Richter, Damien Seguy and Jakub Vrana, 2005, Manuel PHP  
URL: <http://php.benscom.com/manual/fr/index.php> 12/03/2006

(Miles-Board, 2004)

Miles-Board, T., 2004, *Everything Integrated: A Framework for Associative Writing in the Web*. PhD thesis in Electronics and Computer Science, University of Southampton

(Mucchielli et al, 1998)

Mucchielli Alex, Corbalan Jean-Antoine, Ferrandez Valérie, 1998, Théorie des processus de la communication, Collection "Sciences de la communication", Armand Colin, Paris

(Mucchielli et Guivarch, 1998)

Mucchielli Alex and Guivarch Jeannine, 1998, Nouvelles méthodes d'étude des communications, Collection "Sciences de la communication", Armand Colin, Paris

(Müller et Strube, 2003)  
Müller Christoph and Strube Michael, 2003, Multi-Level Annotation in MMAX, *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan: 198-207.

(National, 1996)  
National Communication Systems Technology & Standards Division, 1996 Telecommunications: Glossary of Telecommunication terms, General Services Administration, Information Technology Service, Federal Standard 1037C URL: [http://www.its.bldrdoc.gov/fs-1037/dir-018/\\_2613.htm](http://www.its.bldrdoc.gov/fs-1037/dir-018/_2613.htm)

(National, 1992)  
National Joint Committee for the Communicative Needs of Persons with Severe Disabilities, 1992, Guidelines for meeting the communication needs of persons with severe disabilities, (1992, March), *Asha*, 34, 1-8.,  
<http://www.asha.org/NJC/njcguidelines.htm> (08/06/2006)

(Nicholas et al, 1996)  
Nicholas Charles, Mayfield James and Sasaki James, 1996, Category theory as a foundation for document processing, Computer Science Department, University of Maryland, Baltimore, Pg 13

(Ovsiannikov et al, 1999)  
Ovsiannikov L, Arbib M.A. and McNeill T.H., 1999; Annotation Technology, *International Journal of Human-Computer Studies*, Vol 50, pp 329 - 362

(Paoli et al, 2003)  
Paoli Clément, Dou Henri, Dou Jean-Marie Jr and Maninna Bruno, 2003, La constitution d'indicateurs brevets par domaines technologiques, *Cahiers de la documentation*, N° 2, 2003, <http://www.imcsline.com/Docimcs/CdDocjuin2003.pdf> 11108/2006

(Papakonstantinou et al, 1995)  
Papakonstantinou Yannis, Garcia-Molina Héctor and Widom Jennifer, 1995, Object exchange across heterogeneous information sources, *11th International Conference on Data Engineering (ICDE'95)*, Pages: 251-260

(Parente et Bishop, 2003)  
Parente, Peter and Bishop, Gary, 2003, Enrichments. Unpublished paper. UNC Computer Science. URL: <http://www.cs.unc.edu/~parente/et/Enrichments.pdf> (04/10/2006)

(Paul, 2005)  
Paul Jones, 2005, ITEE seminar: Ambient Information Displays, The School of Information Technology & Electrical Engineering, University of Queensland, Australia Date: 08/11/2005

(Phelps, 1998)  
Phelps Thomas A., 1998, *Multivalent Documents: Anytime, Anywhere, Any Type*,

*Every Way User-Improvable Digital Documents and Systems*, Ph.D. Dissertation, University of California, Berkeley, UC Berkeley Division of Computer Science Technical Report No. UCB/CSD-98-1026, December 1998  
<http://citeseer.ist.psu.edu/context/1225810/0 04/09/06>

(Phelps et Wilensky, 2001)

Phelps Thomas A. and Wilensky Robert, 2001, *The Multivalent Browser: A Platform for New Ideas*, Editor: Munson, Ethan V., *Proceedings of Document Engineering*. Atlanta, Georgia,

(Phelps et Wilensky, 2004)

Phelps Thomas A. and Wilensky Robert, 2004, *Robust Hyperlinks: Cheap, Everywhere, Now*, *Digital Documents: Systems and Principles*, 28-43

(Porter, 1986)

M. Porter, 1986, *Choix stratégique et concurrence. Technique d'analyse des secteurs et de la concurrence dans l'industrie*. *Economica*, Paris.

(Prescott et Gibbons, 1993)

Prescott, J.E. and Gibbons, P.T., 1993, *Global Competitive Intelligence: An Overview*. In J.E. Prescott, & P.T. Gibbons (Eds.), *Global Perspectives on Competitive Intelligence*. Society of Competitive Intelligence Professionals, Alexandria, VA:

(Price et al, 1998)

Price Morgan N., Schilit Bill N., Golovchinsky Gene, 1998, *XLibris: The Active Reading Machine*, *Proceedings of CHI98*,  
<http://www.fxpal.com/publicationsIFXPAL-PR-981111.pdf>

(Prie, 1999)

Prie Y., *Modélisation de documents audiovisuels en Strates Interconnectées par les Annotations pour l'exploitation contextuelle*, Thèse de Doctorat à l'Université Claude Bernard Lyon I, France, 1999

(Pyle, 2003)

Pyle Dorian, 2003, *Business Modeling and Data Mining*, Morgan Kaufmann Publishers, San Francisco

(Reed et John, 2003)

Reed Dale and John Sam, 2003, *Web Annotator*, *Proceeding of ACM Special Interest Group on Computer Science Education (SIGCSE), SIGCSE'03*, February 19-23, 2003, Reno, Nevada, USA.  
<http://logos.cs.uic.eduireed/Papers/SIGCSEAnnotatorPaper.pdf> 10/02/2006

(Reix,1969)

Reix Robert, 1969, *Techniques Quantitatives de Gestion: Traitement des Informations*, Les Editions Foucher, Paris

(Reix,2004)

Reix Robert, 2004, Systèmes d'information et management des organisations, Vuibert, Paris, 5ème Edition - 486 pages

(Revelli, 1998)

Revelli C., 1998, Intelligence stratégique sur internet, Editions Dunod, Paris.

(Reimer et Staudt)

Reimer Ullrich, A., Staudt Margelisch, M., 2000, EULE: A Knowledge-Based System to Support Business Processes, *Knowledge Based Systems*, Volume: 13, Issue: 5, Pages: 261-269 (28/04/2005)

(Rijsbergen, 1979)

Rijsbergen, Cornelis Joost van, 1979, Information Retrieval, Butterworth, London

(Robert, 2003)

Robert, A.B.C., *Représentation des activités du veilleur en contexte de l'intelligence économique*, DEA en Sciences de l'information et de la Communication, Université Nancy 2, Université de Metz, Octobre 2003, Pp 15

(Röscheisen et al, 1994)

Röscheisen, Martin, Mogensen Christian, and Winograd Terry, 1994, Shared Web Annotations As A Platform for Third-Party Value-Added Information Providers: Architecture, Protocols, and Usage Examples,  
URL: <http://www-diglib.stanford.edu/diglib/pub/reports/commentor.html>

(Röscheisen et al, 1995)

Röscheisen, M., Mogensen, C., and Winograd, T., 1995, Interaction Design for Shared World-Wide Web Annotations, *Proceedings of CHI 1995*

(Roscheisen et al, 2001)

Roscheisen Martin, Winograd Terry and Paepcke Andreas, 2001, Content Ratings, and Other Third-Party Value-Added Information: Defining an Enabling Platform  
URL: <http://dbpubs.stanford.edu:8090/pub/showDoc.Fulltext?lang=en&doc=199556&format=pdf&compression=&name=1995-56.pdf>

(Rostaing, 1993)

Rostaing Hervé, 1993, *Veille Technologique et Bibliométrie: Concepts, Outils, Applications*, Thèse de docteur en Sciences, Spécialité sciences de l'Information et de la Communication, Université de droit et des sciences d'Aix-Marseille, Faculté des Sciences et Techniques de Saint Jérôme, Aix-Marseille

(Rumeau, 2006)

Rumeau Y., 2006, Historique des découvertes  
URL: [http://grottesdegargas.free.fr/PAGE\\_5.html](http://grottesdegargas.free.fr/PAGE_5.html) (25/06/2006)

(Saget et al, 1993)

Saget Pierre, Brunet Jean and Sirinelli Jean, 1993, Les sciences de l'information et de la communication, Comité National d'Evaluation, Rapport d'évaluation Mars 1993, Pierre Saget, chargé de mission pour les évaluations,

[http://www.cne-evaluation.fr/WCNE\\_pdf/INFO\\_COM.pdf](http://www.cne-evaluation.fr/WCNE_pdf/INFO_COM.pdf) 31/08/06

(Schickler et al, 1996)

Schickler Matthew A., Mazer Murray S., and Brooks Charles, 1996, Pan-Browser Support for Annotations and Other Meta-Information on the World Wide Web, *Fifth International World Wide Web Conference* May 6-10, 1996, Paris, France

(Schneider, 1996)

Schneider Daniel, 1996, *Modélisation de la démarche du décideur politique dans la perspective de l'intelligence économique*, Faculté des sciences économiques et social département de sciences politiques, Université de Genève, Genève, Pages: 414

(Schölkopf et al, 1998)

Schölkopf B., Burges C. and Smola A., 1998, *Advances in Kernel Methods: Support Vector Machines*, MIT Press, Cambridge, <http://citeseer.ist.psu.edu/33565.html>

(SCIP, 2002)

SCIP, 2002, What is CI?, Society of Competitive Intelligence Professionals URL: <http://www.scip.org/ci>

(Shannon, 1948)

Shannon C.E., 1948, The Mathematical Theory of Communication, *The Bell System Technical Journal*, Vol. 27, pp. 379-423,623-656, July, October, 1948

(Shipman et al, 2003)

Shipman Frank, Price Morgan, Marshall Catherine C., Golovchinsky Gene, 2003, Identifying useful passages in documents based on annotation patterns, *7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003)* Trondheim, Norway, August 17-22,2003

(Shirky, 2005)

Shirky Clay, 2005, Clay Shirky's Writings About the Internet: Ontology is Overrated: Categories, Links, and Tags, URL: <http://www.shirky.com/writings/ontology-overrated.html> 14/05/2006

(Sidhom, 2002)

Sidhom Sahbi, 2002, *Plate-forme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'information: de l'écrit vers la gestion des connaissances*, Thèse de Doctorat à l'Université Claude Bernard Lyon 1, France, Mars 2002.

(Smith, 1999)

Smith Tony, 1999, Third Voice slammed for defacing Web sites URL: [http://www.theregister.co.uk/2004/06/1/five3ears\\_ago/](http://www.theregister.co.uk/2004/06/1/five3ears_ago/) 11/10/2006

(Spais, 2005)

Spais George S., 2005, Marketing Analytics: Managing Incomplete Information in Consumer Markets and the Contribution of Mathematics to the Accountability of Marketing Decisions, 2005 *Hercma Conference*, Athens, <http://www.aueb.gr/pympelhercmalproceedings2005/H05-FULL-P APERS-I/SP AISVELOUTSOU-1.pdf>

(Sreenath et al, 2002)

Sreenath D.V., Grosky William, and Andres Frédéric, 2002, Intelligent Virtual Worlds: Technologies and Applications in Distributed Virtual Environments, chapter Metadata-Mediated Browsing and Retrieval in a Cultural Heritage Image Collection. World Scientific Publishing Company, Singapore

(Stacia, 2004)

Stacia Wyman, Robert K. Jansen and Jerrey L. Boore, 2004, Automatic annotation of organellar genomes with DOGMA, Lawrence Berkeley National Laboratory, University of California, Los Angeles

(Strassner et al, 2004)

Strassner John, Fleck Joel, Huang Jenny, Faurer Cliff and Richardson Tony, 2004, TMF White Paper on NGOSS and MDA: Version 1.0, TeleManagement Forum

(The Institute, 2006)

Strategy, Science of Competitive Intelligence Gathering Techniques ITom "The Art of War"  
URL: [http://www.clearbridge.com/competitive\\_intelligence.htm](http://www.clearbridge.com/competitive_intelligence.htm) 09/08/2006

(Shortreed et al, 2005)

Shortreed Susan, Handcock Mark S. and Hoff Peter, 2005, Positional Estimation within the Latent Space Model for Networks, *Methodology*, vol. 2 no. 1, 24-33. <http://www.stat.washington.edu/susanms/meth06.pdf>04/06/06

(Sudhir et al, 2005)

Sudhir A., Siegfried H., and Staab S., Annotation, Composition and Invocation of Semantic Web Services, *Journal of Web Semantics*, 2005, <http://www.websemanticsjournal.org/ps/pub/2005-5,21/03/05>

(Thagard, 2001)

Thagard Paul, 2001, How to make decisions: Coherence, emotion, and practical inference, MIT Press, Cambridge, Pages: 355-371.

(Thiery et David, 2002)

Thiery Odile and David Amos, 2002, Modélisation de l'utilisateur, Systèmes d'Informations Stratégiques et Intelligence Economique, *Revue Association pour le Développement du Logiciel (ADELI)*, Volume : 47, Pp: 12 p

(Thiery et al, 2002)

Thiery O., Ducreau A., Bouaka N. and David A., 2002, Piloter une organisation : de l'information stratégique à la modélisation de l'utilisateur; application au domaine de la GRH, *Congrès Métamorphoses des organisations* Nancy Novembre 2004.

- (Treurniet, 2004)  
 Treurniet J., 2004, An Overview of Passive Information Gathering Techniques for Network Security, Defence R&D Canada Ottawa Technical Memorandum, DRDC Ottawa TM 2004-073, Ottawa, Pg 50, May 2004
- (Troncy, 2004)  
 Troncy Raphaël, 2004, *Formalisation des connaissances documentaires et des connaissances conceptuelles à l'aide d'ontologies: application à la description de documents audiovisuels*, Thèse de doctorat, Université Joseph Fourier (Grenoble)
- (Truitt et Rogers, 1960)  
 Truitt T.D. and Rogers A.E., 1960, Basics of analog computers, John F. Rider Publishing, Inc New York
- (Tumbull, 2005)  
 Tumbull Douglas, 2005, Automatic Music Annotation, Department of Computer Science and Engineering, University of California, San Diego, pp 19
- (Vannevar, 1945)  
 Vannevar Bush, 1945, As We May Think, *The Atlantic Monthly*, July 1945
- (Vatton et al, 1999)  
 Vatton L, Guétari, R., Kahan, J., and Quint, V., 1999, Amaya Overview, Technical report, W3C. <http://www.w3.org/Amaya>. 12/04/2006
- (Vernon, 2005)  
 Vernon Prior, 2005, The language of business intelligence URL: [http://www.scip.org/2\\_langbi.php](http://www.scip.org/2_langbi.php)
- (Veryard, 1992)  
 Veryard Richard, 1992, Information Modelling: Practical guide, Welland, Ray Prentice Hall, Hertfordshire, Pp: 12-15
- (Vriens,204)  
 Vriens Dirk, 2004, Information and communication technology for competitive intelligence, IRM Press, Hershey, P.A.
- (Watzlawick,1999)  
 Watzlawick Paul, 1999, Rencontre de Paul Watzlawick, Technologie et Communication L'Harmattan, Paris, Pages: 133-139
- (Weber et al, 1997)  
 Weber G., Robert A.B.C., Carsky R.J, 1997, Handbook of use of LEXSYS (Legume Expert System): Decision Support for integrating herbaceous legumes into farming systems. International Institute of Tropical Agriculture, Ibadan, Nigeria. 31pp

(Wheeler, 2005)

Wheeler A. David, 2005, Why open source software / Free software (OSS/FS, FLOSS, or FOSS)? Look at the numbers,  
[http://www.dwheeler.com/lossJs\\_why.html](http://www.dwheeler.com/lossJs_why.html)(27 /06/2006)

(Whittington, 1996)

Whittington C. D., Annotation systems: A review. 1996,  
<ftp://ftp.cs.strath.ac.uk/researchreports/rr-96-202.ps.Z>,

(Wyman et al, 2004)

Wyman, S.K., Jansen, R.K. and Boore, J.L., 2004, Automatic annotation of organellar genomes with DOGMA, *Bioinformatics.*; 20( 17):3252-5

(Yee, 2002)

Yee K., 2002, CritLink: Advanced Hyperlinks Enable Public Annotation on the Web, Demo to the *CSCW 2002 conference*, New Orleans, Dec 2002,  
<http://zesty.ca/pubs/yee/critcsw2002-demo.pdf> 30/03/2005