

Proposition d'une méthode graphique pour le filtrage des flux d'informations

Samuel Szoniecky

► **To cite this version:**

Samuel Szoniecky. Proposition d'une méthode graphique pour le filtrage des flux d'informations. PROBLÉMATISATION ET MÉTHODOLOGIE DE RECHERCHE, Mar 2011, Bordeaux, France. hal-01098438v2

HAL Id: hal-01098438

<https://hal-univ-paris8.archives-ouvertes.fr/hal-01098438v2>

Submitted on 29 Nov 2015

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.



Samuel Szoniecky

Proposition d'une méthode graphique pour le filtrage des flux d'informations

<samszo@free.fr> - Laboratoire Paragraphe – Université Paris 8 – Université de Cergy

Face à la complexité des écosystèmes d'information, nous proposons une méthode graphique pour le filtrage des flux d'information qui utilise les analogies du « tissu de l'âme », des nanostructures de carbone et du crible, pour développer un outil d'ingénierie des connaissances qui s'adresse au plus grand nombre. Nous montrons comment de telles analogies permettent de gérer graphiquement des requêtes booléennes complexes en suivant les principes des *treemaps*.

Introduction

Confronté depuis toujours à une multitude d'informations dont il ne sait ni ne peut maîtriser totalement les conséquences sur son être et son environnement, l'humain questionne ces moyens d'agir pour répondre à l'« inquiétude » qui naît de cette incertitude. Du satellite mis en orbite, au geste du semeur, l'humain a produit une foule d'outils et de méthodes lui permettant d'adapter au mieux son existence face à l'imprévisibilité du monde qui l'entoure. De la simple imitation d'un rituel au respect strict d'un protocole scientifique, les problèmes et les réponses associées ont pris de multiples formes. Pour gérer la multiplication des paramètres à prendre en compte, des formalisations de plus en plus complexes ont été mises en place pour poser les problèmes, les transcrire en processus d'expérimentation, les valider par la reproduction et finalement diffuser leur efficacité.

Aujourd'hui, la modélisation ontologique (Berners-Lee, 2009), (Ferber, 2009), (Limpens, 2010) semble être un des formalismes sur lesquels on porte les plus grands espoirs comme en témoigne le nombre de recherches effectuées dans ce domaine¹. Pour gérer la complexité du monde, les ontologies s'imposent de plus en plus comme l'interface prépondérante pour les échanges entre individus et les machines. Avec ces technologies, les informations sont devenues des « êtres » qui se lient avec d'autres « êtres » pour former des « écosystèmes d'informations ». Désormais, on peut concevoir la gestion de l'information comme une problématique existentielle à l'intérieur d'une niche écologique. Toutefois, il reste de nombreuses questions à résoudre quant à l'utilisation des ontologies et des écosystèmes d'informations. Par exemple, pour concevoir un « web sémantique » allant au delà des « *linked data* » et de la simple mise en relation formelle, mais permettant une mise en relation du sens (Rastier, 2008). Surtout, il reste encore beaucoup de travail à faire pour populariser ces outils très complexes.

Le succès grandissant des initiatives autour des folksonomies et de l'Open Data, laisse croire que se développe un processus de démocratisation de l'accès aux informations et aux moyens de les traiter. Grâce à l'évolution des capacités graphiques des ordinateurs et l'apparition de librairies informatiques pour la représentation algorithmique des données (flare, protovis...),

¹ cf. <http://scholar.google.fr/scholar?hl=fr&q=ontologie+OR+ontology>

nous pouvons facilement représenter les écosystèmes d'informations sous la forme de cartes, de diagrammes ou de plans (Juanals, 2010). Malheureusement, la plupart de ces technologies sont, soit trop complexes pour être utilisées par le « grand public », soit soumises à des droits qui les rendent obscures et difficilement contrôlables par les utilisateurs. Or, un des enjeux majeur des prochaines années consiste justement à rendre accessible ces technologies pour permettre à chacun d'exprimer un « point de vue » qui puisse interagir avec les écosystèmes d'informations et ainsi « piloter » en toute conscience son « pouvoir d'agir » (Rabardel, 2005) dans la complexité de nos sociétés.

Nous proposons dans cet article un modèle pour le filtrage des flux d'informations par la construction graphique d'un réseau de questionnement pouvant servir à la modélisation d'un point de vue. L'ambition de ce travail vise à fournir les bases pour une ingénierie des connaissances (Bachimont, 2007) utilisant des interfaces graphiques pour la conception et l'exploration d'une économie de l'interprétation (Citton, 2010) sous la forme d'un « serious game » (Boudier, 2010). Nous ne présenterons pas dans cet article la totalité de ce projet au cœur de nos recherches sur les langages symboliques pour l'intelligence collective. Nous nous concentrons, sur la pertinence graphique et la cohérence informationnelle du concept de « tissu de l'âme » face au problème du filtrage de l'information.

Le tissu de l'âme comme réseau de questionnements

Dans son cours sur Leibnitz, Deleuze traite la question du choix. Il pose comme prémisse à la réflexion l'alternative suivante : « Vais-je rester travailler, vais-je rester faire cours, ou bien vais-je aller à la taverne »². Il montre que dans ce type de questionnement, ce qui est en jeu, est « un fourmillement de petites inclinaisons » qui forme ce que Leibnitz appelle le « tissu de l'âme ». Arrêtons-nous un instant sur l'idée d'un tissu composé d'une multitude d'inclinaisons, d'une foule de questionnements pour voir en quoi elle peut-être pertinente pour l'ingénierie des connaissances.

Du point de vue des sciences de l'information et de la communication, il nous semble cohérent de voir l'inclinaison à la manière d'une porte logique qui oscille entre ouverture et fermeture, entre 1 et 0. Les travaux de Claude Bantz sur Shannon insistent sur le fait que cette formulation binaire vise en fait à quantifier le nombre de questions nécessaires pour connaître une information : « comprendre systématiquement la mesure de l'information apportée par un événement comme un nombre de questions nécessaire pour identifier cet événement » (Bantz, 2010, 4). Or cet ensemble de questions forme un réseau, ou un « crible » pour reprendre les mots de Deleuze : « L'événement se produit dans un chaos, dans une multiplicité chaotique, à condition qu'une sorte de crible intervienne. » (Deleuze, 1988, 103). Notre hypothèse est que ce « crible » est assimilable à un réseau de questionnement qui, à la manière d'un tissu, « fourmille » au contact des flux d'information (RSS, Twitter, fil Facebook...) et, de par sa dimension matérielle, nous offre une méthode graphique pour l'ingénierie des connaissances tant pour son tissage (Mugur-Schächter, 2006) que pour son usage (Tellez, 2010).

Notre démarche de recherche consiste à utiliser l'analogie (Descola, 2005), (Serres, 2009), (Sanders, 2000) entre les flux d'information et ce « tissu crible », non seulement pour questionner la dimension théorique des sciences de l'information et de la communication, mais aussi et surtout pour construire, grâce à ces théories, des interfaces simples pouvant servir au plus grand nombre afin de comprendre et maîtriser la complexité des flux d'information : « Chaque chose est particulière, certes, mais l'on peut trouver en chaque chose une propriété qui la reliera à une autre, et cette autre à une autre encore, de sorte que des pans entiers de l'expérience du monde se retrouvent ainsi tissés par la chaîne de l'analogie. » (Descola, 2010, 165). Notre travail sur l'analogie va consister à définir quelles correspondances il faut mettre

² <http://www.webdeleuze.com/php/texte.php?cle=139&groupe=Leibniz&langue=1>

en pratique pour que la cohérence entre « tissu crible » et réseau de questionnement puisse opérer efficacement dans une interface graphique : « Rappelons que l'identification analogique repose sur la reconnaissance d'une discontinuité générale des interiorités et des physicalités aboutissant à un monde peuplé de singularités, un monde qui serait donc difficile à habiter et à penser en raison du foisonnement des différences qui le composent, si l'on ne s'efforçait de trouver entre les existants, comme entre les parties dont ils sont faits, des réseaux de correspondance permettant un cheminement interprétatif. » (Descola, 2006, 182). Dans cette optique, examinons comment représenter graphiquement ce tissu de questionnement pouvant servir pour la gestion des flux d'information.

Des nanostructures comme crible

Pour nous aider dans notre conception, prenons l'exemple des nanostructures de carbone (graphène) qui offre les avantages d'une représentation paramétrique facile, d'une grande diversité de forme et dont les ancêtres créés par Buckminster Fuller ont montré leur efficacité en termes d'architecture. De plus, elles correspondent à une des visions théoriques de la communication dont le schéma « simplifié » – le plus courant – comporte explicitement un canal à propos duquel il faut remarquer que sa première caractéristique, matériellement confortée d'ailleurs par le jeu de communication ci-dessus, est qu'il possède des parois. [...] traitée par ailleurs à travers le concept de « tubulation » » (Baltz, 2010, 45). Ainsi l'utilisation de cette analogie nous permet d'ajouter à la dimension linéaire (2D) de la théorie de l'information qui a pour but de conserver une cohérence entre un émetteur et un récepteur ; une dimension volumique (3D) qui permettrait par une mise en réseau des cohérences, la conception de nouveaux outils : « Cette linéarisation peut se comprendre comme un effet de zoom moyen mais on ne peut pas s'empêcher de penser que, selon l'échelle choisie, elle écrase conceptuellement l'accès à des phénomènes buissonnants de communications locales » (Baltz, 2010, 31). En tissant la ligne d'information afin d'obtenir un tissu de connaissance, nous nous trouvons dans la position des nano-scientifiques qui ont transformé le fil de carbone en nanotube. Ainsi, ils ont ouvert la voie à une foule de nouvelles applications, mais aussi à une multitude de nouveaux questionnements³.

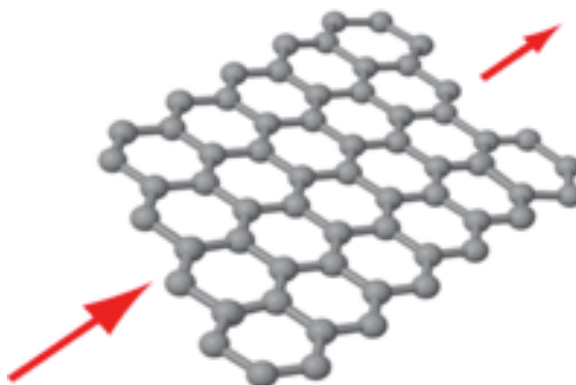


Illustration 1 : Graphène - plan (<http://neel.cnrs.fr/spip.php?article1185>)

³ A ce propos, voir les rencontres 2010 du nouveau monde industriel : http://amateur.iri.centrepompidou.fr/nouveaumonde/enmi/conf/program/2010_2

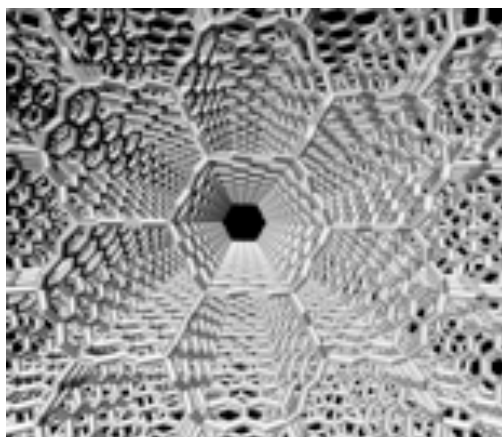


Illustration 2 : Graphène - volume (<http://spie.org/x13545.xml?ArticleID=x13545>)

Basiquement, les nanostructures sont composées d'hexagones mis en relation les uns avec les autres pour former des tubes dont chaque couche prend la forme d'un plan qui rappelle la trame d'un tissu et plus encore à la notion de crible. Nous avons donc une structure formelle qui correspond plutôt bien avec le concept de tube dans lequel peut circuler un flux d'information. De plus cette structure possède aussi les caractères d'un tissu dont la trame compose le crible permettant à l'événement de se produire. Enfin, l'hexagone qui compose la forme élémentaire du graphène peut très facilement servir de structure fractale qui rappelons-le fait partie des caractéristiques fondamentales de l'information-communication (Baltz, 2010).

Nous proposons de représenter la question par un tag au centre des hexagones. À l'instar d'un trou noir au centre du tableau blanc⁴ limité par les flux d'informations, le tag aura pour fonction d'attirer les informations pour filtrer celles qui auront un sens correspondants. Lorsqu'aucune question n'est définie, l'information est négative ce qui correspond à une inclinaison « -1 ». Une question est définie mais aucune information dans le flux ne correspond à cette question, l'inclinaison est « 0 ». Quand il existe une relation entre question et flux d'information, un halo entour le tag et remplit le centre de l'hexagone proportionnellement au nombre d'information (Balpe, 1996).

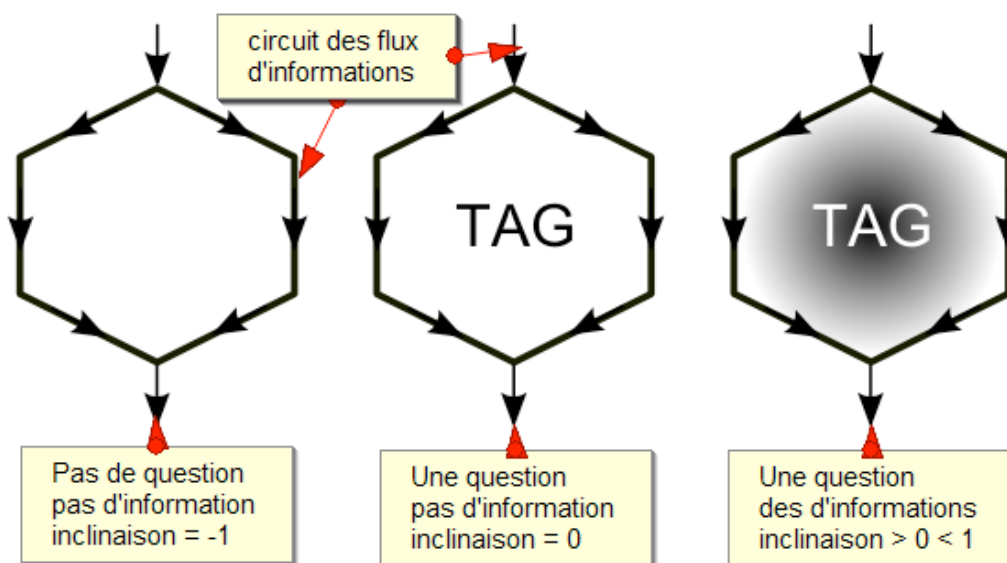


Illustration 3: Structure élémentaire du crible

⁴ Concernant l'interprétation du tag comme trou noir se détachant d'un tableau blanc cf. le concept de visagété dans (D.&G., 1980, 205-234)

Lorsqu'on multiplie les questions, il faut alors définir les relations entre elles. Ces relations s'expriment traditionnellement avec les opérateurs booléens « ET » et « OU ». Les recherches de John Venn (Venn, 1866) sur la représentation graphique de la logique booléenne a montré certaines limites sous la forme des diagrammes qui portent son nom : « Le fait que le nombre de régions d'un diagramme augmente de manière exponentielle par rapport à son nombre de contours est un problème inhérent au diagramme de Venn. On ne peut pas envisager d'utiliser un diagramme de Venn pour représenter plus de quatre ou cinq ensembles. » (Thièvre, 2006, 38). Pour éviter ce problème et pouvoir multiplier à souhait les relations booléennes entre les termes, nous proposons les règles suivantes. Par défaut les tags présents dans un crible sont mis en relation par un opérateur « OU ». Pour obtenir un opérateur « ET », il faut supprimer le canal de circulation des flux entre les éléments.

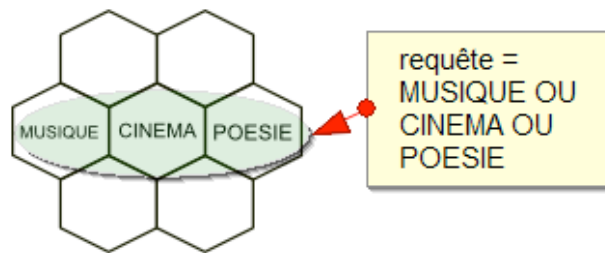


Illustration 4 : Relations OU entre tags

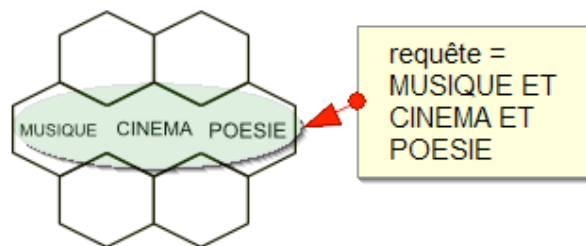


Illustration 5 : Relations ET entre tags

Les représentations ci-dessus montrent comment les flux d'informations « irriguant » le « tissu crible » sont filtrés suivant une logique booléenne. Ces exemples relativement simples peuvent être complexifiés en utilisant la dimension fractale de la structure élémentaire. Un changement d'échelle permettra de définir plus précisément la question que l'on souhaite confronter aux flux d'information. Il sera possible en un coup d'œil de savoir si une source d'information correspond en partie ou totalement à une requête complexe.

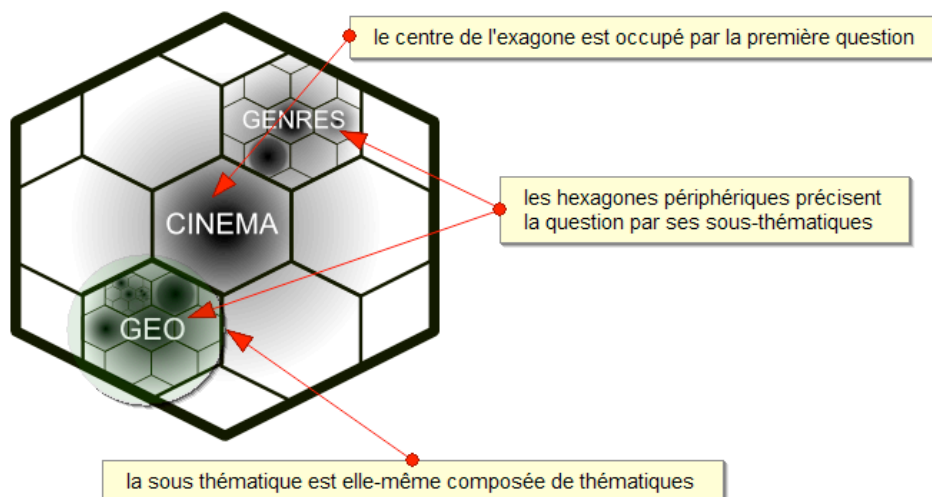


Illustration 6 : Question complexe cinéma

Ce principe graphique s'inspire grandement du modèle de Circle Packing⁵ issu des recherches de Ben Shneiderman⁶ sur les treemaps. Toutefois, nous utilisons ce procédé, non pas uniquement pour visualiser les résultats d'une recherche mais aussi pour paramétrer de façon complexe cette recherche. En fait chaque grille de questionnement peut être utilisée de façon autonome pour filtrer l'information. Ainsi, les sous-thématiques « GENRE » et « GEO » que nous avons présentées ci-dessus dans le contexte du cinéma, peuvent aussi être utilisées dans d'autres contextes par un simple déplacer-lâcher (drag&drop) des hexagones les représentants dans les hexagones « POESIE » ou « MUSIQUE ».

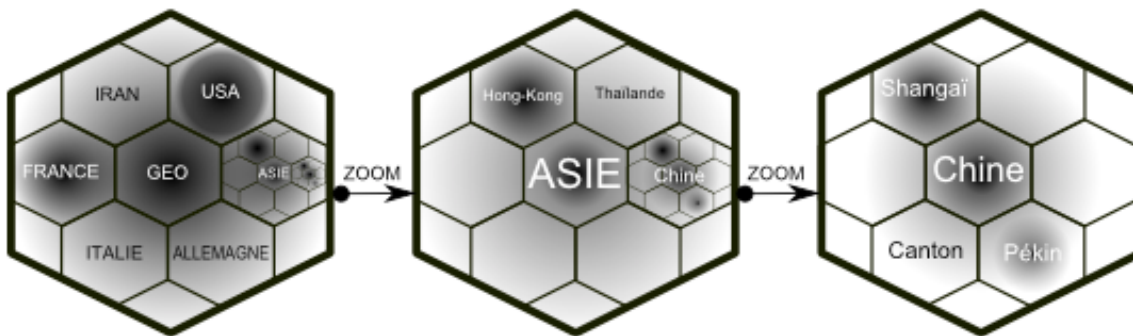


Illustration 7: Question complexe Cinéma -> géo

Dans cette optique, l'outil que nous proposons a pour vocation de mettre en place une boucle de rétroaction continue entre les données et leur visualisations, à l'inverse des moteurs de recherche traditionnels qui, eux, fonctionnent dans un processus itératif : une requête, puis une autre et ainsi de suite. Il semble très important de fournir des outils qui permettent de gérer sur le long terme les interactions avec les écosystèmes d'information. Pensons par exemple aux écoliers qui pourraient disposer ainsi d'une vision évolutive de leurs connaissances au fur et à mesure de l'enrichissement de leur réseau de questionnement et ainsi observer, diffuser, comparer leurs étendues et ce qu'ils leur restent encore à découvrir.

Nous n'avons fait qu'effleurer la problématique du « tissu de l'âme » et de sa pertinence pour la gestion des flux d'informations. Nous n'avons pas traité par exemple la question de la forme que peut prendre ce tissu. Est-il clos sur lui-même pour former une bulle informationnelle dont la multiplication constitue une écume (Rieder, 2010) ? Chaque bulle sera-t-elle la formalisation des rapports entre intériorité et physicalité à l'instar des matrices ontologiques (Descola, 2005) ?

Conclusion

Une interface graphique s'inspirant de la notion de « tissu de l'âme » pour créer des grilles de questionnement, permettrait de filtrer les flux d'informations afin d'en extraire les informations répondant à une interrogation complexe. Remarquons toutefois que nous n'avons pas appréhendé le problème de la relation entre les flux d'information et les tags composants le réseau de questionnement. Ce problème sera abordé dans notre thèse pour montrer comment il est possible de traiter l'interopérabilité et de calculabilité sémantique grâce aux propositions complémentaires de la méthode de conceptualisation relativisée (MCR) (Mugur-Schächter, 2006) et du métalangage pour une économie de l'information (IEML) (Lévy, 2010).

De même nous n'avons pas traité ici la question cruciale des limites des langages symboliques dont on pourrait dépasser la dimension fixe d'une « connaissance faite », au profit d'un langage allégorique dynamique pour une connaissance « en train de se faire ». Point que nous

⁵ <http://vis.stanford.edu/protovis/ex/pack.html>

⁶ <http://www.cs.umd.edu/hcil/treemap-history/>

avons abordé dans un article où nous présentons nos recherches dans ce domaine en vue de la réalisation d'un *serious game* (Szoniecky, 2011). Jeu sérieux qui met en pratique l'allégorie du jardinage pour rendre accessible au plus grand nombre les méthodes d'ingénierie des connaissances et d'écologie de l'information en les associant aux gestes nécessaires pour cultiver un jardin de connaissances. Par exemple, après avoir composé un composte-sémantique à partir des cribles que nous avons décrit dans cet article, le jardinier pourra planter des graines-questions qui alimentées par les flux d'informations des nuages-sources feront pousser automatiquement des végétaux-documents qu'il faudra tailler pour sélectionner ceux qui semblent les plus adéquates avec des racines-concepts...

Enfin, il conviendra d'aborder l'aspect pratique des outils que nous proposons. Cela sera fait suivant deux perspectives complémentaires d'intelligence collective à savoir la génération automatique de texte et l'évaluation globale d'un territoire urbain.

Bibliographie

Bachimont B., 2007 *Ingénierie des connaissances et des contenus : Le numérique entre ontologies et documents*, Paris: Hermes science publications.

Balpe J. & Collectif, 1996, *Techniques avancées pour l'hypertexte*, Hermès.

Baltz C., 2010, *Shannon, critiques, dépliages, reconstruction*, Europa Production, Paris.

Berners-Lee T., 2009, *Putting Government Data online - Design Issues*.

Boudier V.L. & Dambach Y., 2010, *Serious Game : Révolution pédagogique*, Hermes Science Publications.

Chauvin S., 2005, *Visualisations heuristiques pour la recherche et l'exploration de données dynamiques : l'art informationnel en tant que révélateur de sens*. Paris 8.

Citton Y., 2010, *L'avenir des humanités : Economie de la connaissance ou cultures de l'interprétation ?*, Editions La Découverte.

Deleuze G., 1988, *Le pli*, Editions de Minuit.

Deleuze G. & Guattari F., 1980, *Mille plateaux*, Paris: Éditions de minuit.

Descola P., 2005, *Par-delà nature et culture*, Paris: NRF : Gallimard.

Descola P., 2006, *La fabrique des images. Anthropologie et Sociétés*, p.167-182.

Descola, P. & Collectif, 2010, *La Fabrique des images : Visions du monde et formes de la représentation*, Somogy éditions d'art.

Ferber J., 2009, « Vers un modèle multi-agent de construction d'ontologie », *XVIèmes rencontres de Rochebrune*

Juanals, B. & Noyer, J., 2010, « De l'émergence de nouvelles technologies intellectuelles. », *Technologies de l'information et intelligences collectives*. Hermes Science Publications.

Limpens, F., 2010. *Multi-points of view semantic enrichment of folksonomies*. Université de Nice - Sophia Antipolis - Ecole doctorale STIC.

Lévy P., 2010, *From social computing to reflexive collective intelligence: The IEML research program*. *Information Sciences*, p.71–94.

Mugur-Schächter M., 2006, *Sur le tissage des connaissances*, Hermes Science Publications.

Rabardel P., 2005, « Instrument subjectif et développement du pouvoir d'agir. », *Modèles du sujet pour la conception : dialectiques, activités, développement*. Toulouse: Octarès éditions.

Rieder B., 2010, *Membranes numériques : des réseaux aux écumes*.

Rastier F., 2008, « Sémantique du web vs semantic web ? », *Texte !*

- Sander E., 2000. *L'analogie, du naïf au créatif : analogie et catégorisation*, L'Harmattan.
- Serres M., 2009. *Ecrivains, savants et philosophes font le tour du monde*, Paris: Pommier.
- Szoniecky, S., 2011. *Le langage du Web du symbolique à l'allégorique, vers une représentation de la connaissance en train de se faire*. Dans ISKO - Magreb 2011. Concepts and Tools for Knowledge Management (KM). Hammaet, Tunisie.
- Thièvre J., 2006, *Cartographies pour la Recherche et l'Exploration de données Documentaires*. MONTPELLIER II.
- Venn J., 1866, *The logic of chance: An essay on the foundations and province of the theory of probability, with especial reference to its application to moral and social science*, Macmillan.