

Sciences humaines combinées

ISSN : 1961-9936

: Université de Bourgogne, Université de Franche-Comté, COMUE Université
Bourgogne Franche-Comté

15 | 2015

Actes du colloque 2015

Outils sémantiques au service du livre numérique, modélisation et visualisation des liens transtextuels

01 August 2015.

Clément Borel

DOI : 10.58335/shc.391

🔗 <http://preo.ube.fr/shc/index.php?id=391>

Clément Borel, « Outils sémantiques au service du livre numérique, modélisation et visualisation des liens transtextuels », *Sciences humaines combinées* [], 15 | 2015, 01 August 2015 and connection on 15 June 2026. DOI : 10.58335/shc.391. URL : <http://preo.ube.fr/shc/index.php?id=391>

PREO

Outils sémantiques au service du livre numérique, modélisation et visualisation des liens transtextuels

Sciences humaines combinées

01 August 2015.

15 | 2015

Actes du colloque 2015

Clément Borel

DOI : 10.58335/shc.391

 <http://preo.ube.fr/shc/index.php?id=391>

Lecture critique : limites du support papier

 L'exemple du Commentaire au Discours de la Méthode

 Une solution : le livre numérique

Outils et langages sémantiques

 RDF, langage de description

 OWL, langage ontologique

 SPARQL, langage de requête

Vers un livre numérique sémantique

 Des liens hypertextes aux liens profonds

 TROW, une ontologie de la transtextualité

 Architecture technique et interface de lecture parallèle

Démonstrateur et évaluation

 Déroulement de l'évaluation

 Résultats

Conclusions et perspectives

- 1 L'histoire des technologies de l'information a montré, à de nombreuses reprises, que l'invention d'un nouveau médium ne supprime pas l'ancien mais que les deux cohabitent, au moins pour un temps, chacun se spécialisant. Il ne s'agit donc pas d'opposer les deux conceptions du livre : papier/numérique, mais de s'interroger sur les apports des technologies numériques pour le livre et d'imaginer com-

ment son homologue numérique peut venir compléter l'expérience de lecture traditionnelle. Nous nous intéressons en particulier à l'usage des technologies et langages sémantiques du Web pour la réalisation d'un livre numérique qui pourrait accompagner le lecteur et faciliter sa tâche dans le cadre d'une lecture critique.

Lecture critique : limites du support papier

- 2 Parmi les méthodes traditionnellement employées lors de l'étude d'un texte, l'une des plus courante consiste à repérer les sources et les reprises des concepts développés par l'auteur. Cela nécessite la consultation, si possible simultanée, de nombreux livres. Il faut ainsi pouvoir repérer les différentes liaisons entre les textes d'un corpus et pouvoir accéder à l'ensemble des ressources convoquées par ces liaisons.
- 3 Nous parlerons d'un « texte source », pour qualifier le texte principal dont on souhaite faire l'étude. Les « textes cibles », quant à eux, désigneront les textes utiles à l'étude du texte source, convoqués par des liaisons.

L'exemple du Commentaire au Discours de la Méthode

- 4 Les auteurs et les éditeurs ont, depuis longtemps, imaginé des solutions pour faciliter cet exercice et ainsi favoriser la compréhension d'un texte dont on souhaite faire l'étude.
- 5 Parmi ces solutions, on peut citer, à titre d'exemple, le Discours de la Méthode accompagné des commentaires d'Etienne Gilson, dont la première édition fut publiée chez Vrin en 1925 (Descartes et Gilson 1925). Ce livre regroupe sur un même support, un texte - le Discours de la Méthode de René Descartes - et l'ensemble des commentaires argumentés utiles à son étude. Ces commentaires référencent, passage après passage, les ouvrages qui ont pu inspirer Descartes dans son écriture du Discours ainsi que tous ceux qui permettent d'expliquer, commenter, critiquer... en un mot comprendre le texte de Descartes. Gilson espère ainsi, d'après ses propres termes, « faciliter

l'exercice de la réflexion personnelle sans pour autant s'y substituer » (Descartes et Gilson, 1925, préface).

- 6 Si le commentaire de Gilson offre un formidable outil d'analyse du Discours de la Méthode en facilitant l'exercice de la réflexion personnelle, sa version papier limite ses aspirations. Gilson s'excuse lui-même d'un appareil d'érudition qu'il aurait voulu « moins encombrant et moins pédantesque » (ibid.). Concernant l'encombrement, précisons que sur les 480 pages du livre, 400 sont réservées au Commentaire de Gilson.
- 7 Dans cet ouvrage, les passages du Discours qui constituent une entrée du commentaire ne sont pas directement repérés dans le texte. Il faut alors effectuer des va-et-vient entre les deux textes - Discours et Commentaire Historique - pour repérer ces entrées. De plus, les sources citées par le commentateur ne sont pas directement accessibles ; elles sont uniquement référencées. Enfin, lorsqu'une entrée de commentaire fait référence à un passage particulier d'un texte cible, l'intégralité du passage concerné est cité par le commentateur. Cependant, puisque le texte cible n'est pas accessible dans son intégralité, il n'est pas possible de le replacer dans son contexte d'origine.

Une solution : le livre numérique

- 8 Nous pensons que les technologies numériques peuvent apporter une solution aux limites que nous venons de présenter. En particulier, le mécanisme des liens hypertextes permet de repérer, à l'intérieur même du Discours de la Méthode, les passages constituant une entrée de commentaire. Ces hyperliens permettent, en outre, d'accéder directement à l'entrée du commentaire concernée ou aux textes cibles dans leur intégralité. Enfin, le numérique autorise l'affichage de tous ces textes sur le même support : l'écran.
- 9 Ainsi, nous pensons que le livre numérique est bien adapté à l'exercice de la réflexion personnelle lors de l'étude critique d'un texte.

Outils et langages sémantiques

- 10 Depuis sa création, le Web, bien que s'appuyant toujours sur le même principe, a fait évoluer nos usages. Pour certains, ces évolutions se

traduisent par des versions successives du Web, qu'ils notent Web 1.0, Web 2.0 et Web 3.0. Sous ces différentes appellations ne se cachent pas des versions différentes du Web, mais bien des nouveaux usages rendus possibles par l'émergence de nouvelles caractéristiques, de nouvelles technologies et de nouveaux langages.

- 11 Le Web de Données ambitionne d'ajouter une couche sémantique au-dessus du Web existant (Berners-Lee, Hendler, et Lassila, 2001). Cet ajout de sens est rendu possible par l'interconnexion et l'interopérabilité des données. On parle du passage d'un Web de « documents connectés » à un Web de « données connectées ». Le Web de Données ne remet pas en cause le Web existant ; il se compose d'une succession de couches technologiques qui viennent s'appuyer sur les technologies déjà présentes. Contrairement à la logique de documentation dans laquelle on construit le savoir à partir de l'observation du monde (Otlet, 1934), les couches successives du Web de Données permettent de reconstruire le savoir à partir des ressources disponibles.
- 12 Sans entrer dans les détails techniques, nous souhaitons tout de même présenter succinctement les 3 principaux langages sémantiques que nous utilisons dans notre étude ; à savoir RDF, OWL et SPARQL.

RDF, langage de description

- 13 La première brique sémantique du Web de Données est RDF, pour *Resource Description Framework* (Klyne, Carroll, et McBride, 2004). Comme son nom l'indique, il s'agit d'un *framework*¹ permettant de décrire les ressources sur le Web. C'est un standard qui permet de décrire des métadonnées, permettant d'obtenir des descriptions sur les ressources disponibles sur le Web (Manola et Miller, 2004). Ce langage structure l'information sous la forme d'un graphe orienté et valué.
- 14 RDF fonctionne par assertions qui prennent la forme d'un triplet {sujet, prédicat, objet }
- 15 En guise d'exemple, l'expression de la phrase « René Descartes a écrit le Discours de la Méthode » peut se faire par le biais de l'écriture d'un triplet RDF. Ici, « René Descartes » serait le sujet, « Discours de la

Méthode » l'objet, et ces deux entités sont reliés par le prédicat « a écrit ». L'ensemble des triplets sur un ensemble d'entités forme un graphe.

OWL, langage ontologique

- 16 La couche ontologique permet d'explicitier les concepts grâce à un langage formel. L'ontologie, qui doit son nom à une branche de la philosophie, est définie en informatique comme une spécification explicite et formelle d'une conceptualisation partagée (Studer, Benjamins, et Fensel, 1998) ; elle est explicite car elle est définie de manière claire, précise ; formelle car elle est modélisée grâce à un langage formel qui permet à la machine de l'interpréter. Il s'agit d'une spécification d'un modèle abstrait qui a été pensée et acceptée par une communauté. Le W3C propose le langage OWL, pour *Web Ontology Language*, afin de modéliser les ontologies informatiques (McGuinness et van Harmelen, 2004).
- 17 Ce langage nous permet de définir des classes, auxquelles pourront appartenir des individus, et les propriétés qui relient ces différentes classes. Dans notre exemple, il faudrait définir, grâce à une ontologie, les concepts « Auteur » et « Livre » ainsi que la relation « aEcrit ». En pratique, il s'agit alors de 2 classes OWL distinctes – « Auteur » et « Livre » - reliées par une propriété « aEcrit ». Il est également possible de définir une nouvelle propriété « écritPar » qui serait la propriété inverse de « aEcrit ». L'ensemble de ces déclarations constitue la boîte terminologique - TBox - de notre base de connaissance RDF.
- 18 La définition des individus appartenant à ces classes forme quant à elle la boîte assertionnelle - ABox - de notre base de connaissance. Dans notre exemple, « René Descartes » serait une instance de la classe « Auteur » et « Discours de la Méthode » une instance de la classe « Livre ».
- 19 Grâce à cette base de connaissances, si la machine trouve une assertion du type : « René Descartes a écrit le Discours de la Méthode », elle sera capable d'inférer la propriété inverse, à savoir : « le Discours de la Méthode est écrit par René Descartes ». Notons que l'exemple donné dans cette partie n'a qu'une visée explicative et ne reflète pas

les structures réelles des bases de connaissances réalisées dans le cadre de nos travaux.

SPARQL, langage de requête

- 20 Afin d'interroger une base de données ou tout autre système d'information, nous avons besoin d'un langage de requête approprié. Chaque système traitant des données stockées dans une base de connaissance a besoin d'un langage d'interrogation pour les valoriser (Guha et al., 1998). La structure des données interrogées étant différente pour chaque système d'information, la grammaire du langage de requête doit être adaptée.
- 21 Les bases de connaissance de type RDF ont elles aussi besoin d'un langage de requête adapté ; il s'agit de SPARQL - *SPARQL Protocol And RDF Query Language*. Faisant l'objet d'une recommandation du W3C depuis 2008 (Prud'hommeaux et Seaborne, 2008), ce dernier définit la syntaxe et la sémantique nécessaires au requêtage d'une base de connaissances RDF.

Vers un livre numérique sémantique

- 22 L'expression « livre numérique » recouvre différentes acceptions. Elle qualifie tantôt le support numérique de lecture : liseuse, tablette ; tantôt les formats de fichiers numériques : ePub, PDF ; ou plus largement les contenus numériques (principalement textuels) lisibles sur un écran. Dans le cadre de notre travail, nous retenons la dernière définition, et considérons le livre numérique comme un ensemble de ressources numériques, principalement textuelles, constituant des entités autonomes, rendus lisibles à travers une interface numérique de lecture adaptée. Cette adaptation concerne le type de ressources en présence et le support matériel et logiciel utilisé par le lecteur.
- 23 Il a fallu du temps au livre imprimé pour se perfectionner et trouver une forme stable. De la même façon, le livre numérique n'a pas encore de forme bien définie. On trouve ainsi dans l'offre numérique de lecture des livres que l'on peut qualifier d'homothétiques. Ceux-ci reprennent la forme et la composition des livres papiers, comme les

premiers livres imprimés imitaient les codes et les formes des livres manuscrits. Notons cependant que ces livres numériques homothétiques proposent parfois des outils permettant de naviguer à l'intérieur du texte en automatisant les outils tels que la table des matières et les glossaires déjà présentant dans le livre papier.

- 24 On observe une évolution dans l'offre numérique avec l'apparition de livres enrichis ou augmentés qui proposent de tirer profit des possibilités offertes par le numérique. Ceux-ci bénéficient de compléments tels que l'ajout de médias temporels (sons, animations, vidéos) et l'utilisation d'interactions. Ces interactions permettent de naviguer à l'intérieur du livre, d'accéder à des contenus complémentaires (grâce, la plupart du temps à des liens hypertextes) ou encore de moduler son expérience de lecture : les actions du lecteur modifiant le contenu même de l'ouvrage.
- 25 Suivant cette évolution, nous proposons de définir un livre numérique sémantique (LNS) qui tirerait profit des technologies sémantiques et permettrait de modéliser et de visualiser les différentes liaisons existantes entre les textes d'un corpus. Tout texte est au centre d'un réseau de textes qui, l'expliquant, le critiquant, le replaçant dans son contexte, lui donne toute sa signification. Les liens unissant ces textes ne possèdent pas tous la même typologie, ils ne contribuent pas tous de la même manière à la compréhension de l'œuvre principale.
- 26 Si une machine computationnelle est capable de les interpréter et donc de les différencier, elle est également capable de les filtrer et ainsi faciliter la tâche du lecteur soucieux d'étudier un texte en ne lui proposant que les liaisons qu'il juge pertinentes en fonction de ses objectifs de lecture. Elle permet aussi d'inférer des liaisons inédites qui ne seraient pas préalablement enregistrées de façon explicite.

Des liens hypertextes aux liens profonds

- 27 Pour réaliser les liaisons intertextuelles d'un LNS, les liens hypertextes classiques ne suffisent plus. Il faut un lien qui puisse associer plus de deux ressources cibles à une même ancre. En effet, un même passage du texte étudié renvoie parfois vers plusieurs textes ou pas-

sages de textes différents. Il faut également un lien qui soit capable d'enregistrer le type de liaison qu'il caractérise. En effet, comme nous l'avons déjà énoncé, les textes cibles ne contribuent pas toute de la même manière à la compréhension de la ressource principale, et nous souhaitons exploiter cette typologie dans le cadre d'un LNS. Enfin, ce lien doit être capable de fonctionner dans les deux sens. En effet s'il existe une relation entre une ressource X et une ressource Y, il existe alors la relation inverse entre Y et X. Nous souhaitons que la machine puisse inférer cette relation inverse et donc proposer la liaison de Y vers X sans que celle-ci ne soit explicitement enregistrée.

- 28 Pour exploiter ces fonctionnalités nous proposons de définir des « liens profonds » (Borel et Roxin, 2013), que nous définissons comme des hyperliens sémantiques étendus. Ils sont étendus - terme emprunté au langage XLink : « extended » - car ils permettent d'associer à une même ancre plus de deux ressources cibles. Ils sont également sémantiques car ils possèdent des métadonnées précises concernant la liaison intertextuelle. Nous empruntons le terme de lien profond, à Ted Nelson qui parle de « deep links », pour qualifier les liens « abondants et incassables créés par les auteurs » dans l'évolution de son projet Xanadu (Nelson, 1999).
- 29 Pour réaliser ces liens, nous souhaitons utiliser les outils et langages sémantiques du Web de Données pour donner les moyens à une machine computationnelle de les traiter et les analyser.
- 30 Dans le cadre de notre travail, nous utilisons les langages sémantiques pour modéliser les liaisons entre les textes d'un corpus. Pour ceux-ci, les assertions RDF sont du type :
- 31 « le bon sens » qui est un passage du Discours de la Méthode de Descartes, est relié, selon une relation transtextuelle particulière, à « de bona mente » qui est un passage de De Vita Beata de Sénèque. Notre ontologie OWL, quant à elle, définit les différents types de relations transtextuelles.

TROW, une ontologie de la transtextualité

- 32 Les recherches en littérature s'intéressent depuis longtemps aux liaisons qui transcendent le texte. Ainsi dans les années 60, le groupe Tel

Quel développe la théorie de l'intertextualité (Tel Quel, 1968) selon laquelle « tout texte est un intertexte ; d'autres textes sont présents en lui à des niveaux variables, sous des formes plus ou moins reconnaissables : les textes de la culture antérieure et ceux de la culture environnante ; tout texte est un tissu nouveau de citation révolues » (Barthes, 1974).

33 En 1982, Genette intègre le concept d'intertextualité à une notion plus large, celle de la transtextualité qui s'intéresse à « tout ce qui met [le texte] en relation manifeste ou secrète avec d'autres » (Genette, 1982). Ainsi il distingue 5 types de relations transtextuelles :

34 - **l'intertextualité** représente la coprésence de textes, liée à la citation, le plagiat ou l'allusion

35 - la **paratextualité** est utilisée pour tout ce qui est périphérique au corps d'un énoncé (titre, sous-titre, intertitre, préface, postface, avertissement, note, etc.)

36 - la **métatextualité** est identifiée au commentaire qui « unit un texte à un autre texte dont il parle, sans nécessairement le citer (le convoquer), voire à la limite sans le nommer »

37 - **l'hypertextualité** raison de l'analyse de Genette, consiste en un rapport d'imitation (ou de transformation) entre deux textes;

38 - **l'architextualité** permet d'organiser, ou de déterminer le statut générique d'un énoncé par simple relation d' « appartenance taxinomique » à une catégorie.

39 Travaillant sur le Discours de la Méthode, nous nous sommes intéressés à la typologie de l'intertextualité pour un ouvrage de philosophie. Ainsi, pour le corpus cartésien, l'organisation intertextuelle pourrait se faire suivant cinq catégories découlant des types de relations transtextuelles citées précédemment : cotexte, doctritexte, métatexte, extratexte, analotexte (Hufschmitt 2009).

40 - On désigne par **cotexte** les ouvrages et surtout passages d'ouvrages qui ont pu déterminer l'écriture du texte noyau, directement ou indirectement.

41 - Les textes extraits du corpus, antérieurs ou postérieurs, y compris d'autres textes de l'ouvrage lui-même, constituent le **doctritexte**. Ce

corpus inclut aussi les ouvrages qui témoignent ou résument la doctrine.

42 - Le **métatexte** est constitué des textes postérieurs qui ont le texte noyau, ou un de ses passages, comme objet explicite d'étude et d'analyse.

43 . On parle d'**extratexte** lorsque les textes utilisent le texte noyau ou un de ses passages dans une optique qui n'a rien à voir avec la compréhension du texte.

44 - Enfin, l'**analotexte** désigne les différents textes qui présentent tout autre rapport au texte noyau

45 Nous avons restructuré l'ensemble de ces relations afin de créer notre modèle ontologique. L'ontologie que nous avons modélisée, définit l'ensemble des propriétés transtextuelles selon le modèle précédemment établi. Il s'agit d'une ontologie OWL baptisée « TROW », acronyme inverse de « Web Ontology of Transtextuality ». Outre sa fonction d'acronyme, « *trow* » est un ancien terme anglais signifiant « penser », « croire » et évoque selon nous l'utopie du savoir universel dont la réalisation, même partielle, passe par la mise en relation des connaissances disponibles. Notre ontologie participant modestement à tisser des liens sémantiques entre les textes trouve dans sa dénomination une certaine résonance.

46 Le Dublin Core (ISO, 2009), schéma de description bibliographique très largement utilisé, possède une propriété `dc:Relation` qui permet de définir qu'une ressource entretient une relation quelconque avec une autre. Puisque les propriétés que nous définissons dans notre ontologie représentent justement un type de relation particulier entre 2 ressources, nous avons défini la propriété mère de notre hiérarchie comme une sous propriété de `dc:Relation`. Nous souhaitons ainsi replacer notre ontologie dans un cadre plus large et permettre une plus grande interopérabilité des données.

Architecture technique et interface de lecture parallèle

47 Nous pouvons désormais présenter l'architecture d'un Livre numérique sémantique mettant en œuvre les liens profonds que nous

avons proposés et dont les relations sont définies par notre ontologie TROW.

- 48 L'architecture du LNS, se compose de 3 niveaux. Le premier, textuel constitué de l'ensemble de ressources textuelles disponibles enregistrées au format XML ou dérivé. Le second, descriptif, contient les métadonnées RDF qui définissent les liens profonds pour chaque texte ; c'est à dire l'ensemble des métadonnées pour décrire les liens et les ancrs et leur relation. Enfin le troisième niveau, sémantique est composé des ontologies qui définissent le vocabulaire utilisé par les métadonnées du niveau précédent, comme notre ontologie TROW.
- 49 Outre l'architecture, nous proposons également une interface, pour le LNS, qui puisse elle aussi favoriser l'exercice de la réflexion personnelle. Lorsqu'on étudie un texte et ces sources sur support papier, il est courant de disposer l'ensemble des textes devant soi, sur son bureau, ouverts afin de faciliter le passage de l'un à l'autre et de pouvoir remettre rapidement en contexte, dans le texte principal d'étude, les informations glanées dans le second. Nous souhaitons qu'un LNS permette cette disposition parallèle des textes. A l'activation d'un lien nous souhaitons que l'affichage du texte ciblé ne se fasse pas au détriment du texte en train d'être lu. Cela permettant aux lecteurs de garder à l'écran le texte et le passage du texte qui a motivé l'activation du lien.
- 50 Nous imaginons alors une interface de lecture divisée en deux parties. À gauche de l'écran s'affiche le texte source paginé. Les textes appelés par l'activation d'un lien s'affichent dans la partie droite de l'écran et proposent un mode de défilement vertical. Soucieux de proposer au lecteur les conditions idéales d'appréhension des textes, nous souhaitons replacer les textes qui apparaissent dans leur contexte d'origine, c'est-à-dire dans leur intégralité, et cela même si la liaison ne concerne qu'un passage particulier. Bien entendu, la page de texte est positionnée de manière à afficher le passage concerné dès l'activation du lien.

Démonstrateur et évaluation

51 Afin de tester notre proposition de LNS, nous avons réalisé une évaluation pour laquelle nous avons développé un démonstrateur. Le développement de ce dernier nous a permis de tester la validité de nos choix techniques et de mettre en œuvre l'interface de lecture parallèle que nous imaginions.

Déroulement de l'évaluation

52 L'évaluation s'est déroulée en trois phases :

53 - La première introductive sous forme d'entretien, durant laquelle nous présentons le cadre de notre étude et nous invitons le lecteur à répondre oralement à une série de questions concernant son profil, ses pratiques de lecture, et ses connaissances préliminaires du Discours de la Méthode.

54 - Durant la seconde phase, le lecteur est invité à prendre place devant un ordinateur sur lequel est installé notre démonstrateur. En autonomie, le lecteur doit alors interagir avec le démonstrateur afin de répondre à une série de questions concernant la compréhension du Discours de la Méthode. Lorsque le lecteur est satisfait de ses réponses, l'expérimentation s'arrête. Un fichier numérique sur lequel les questions sont formulées sert de formulaire de réponse.

55 - Dans la troisième phase, nous avons interrogé les lecteurs sur leur expérience quant à l'utilisation du démonstrateur et sur leur connaissance des typologies de liaisons transtextuelles.

56 Chacune de ces phases a été enregistrée à l'aide d'un dispositif audio-vidéo permettant d'enregistrer les réponses du lecteur lors des phases d'entretien et d'enregistrer les commentaires et les actions à l'écran lors de la phase procédurale.

57 Pour cette dernière, chaque lecteur est installé devant un ordinateur présentant notre démonstrateur dans son état initial. Pour l'évaluation, les lecteurs n'avaient accès qu'à un extrait du Discours de la Méthode - soit les 6 premières pages de l'édition Vrin de 1939 - dans lequel 29 ancres étaient disponibles proposant 48 liens et donnant accès à 22 textes cibles.

58 À l'issu de l'évaluation, nous avons confronté les séquences d'interaction effectives de chaque lecteur aux séquences d'interactions attendues pour répondre à chaque question. Pour cela, nous avons analysé les actions effectuées par chaque utilisateur afin de répondre à une question et en avons extrait des séquences d'interactions récurrentes que nous avons alors comparées aux séquences d'interaction minimales attendues.

Résultats

59 Notre analyse montre que les séquences d'actions que nous avons observées sont relativement conformes aux séquences d'interaction attendues. Les réponses aux questions données par les lecteurs sont également conformes aux réponses attendues.

60 Ainsi, du point de vue théorique, les résultats soutiennent nos propositions : l'affichage parallèle des textes favorise la remise en contexte des informations, la compréhension du texte noyau ; l'utilisation des liens profonds a permis aux lecteurs de trouver les informations nécessaires afin de répondre aux questions posées et même de les compléter.

61 Du point de vue applicatif, les résultats donnent des pistes pour l'amélioration du dispositif :

62 - l'ajout d'un champ de recherche ;

63 - l'ajout de filtres pour masquer et/ou réorganiser les liens ;

64 - la possibilité d'activer/désactiver le surlignement du texte cible ;

65 - l'explicitation des catégories transtextuelles : les entretiens ont montré que l'utilisation d'une typologie de lien semble pertinente, ceux-ci doivent être explicités de façon plus claire pour les utilisateurs. On peut imaginer que les lecteurs choisissent eux même le vocabulaire qu'ils souhaitent utiliser - comme pour une folksonomie - et qu'ils puissent associer chaque terme à un type de liaison que nous avons défini.

66 Bien entendu, les résultats obtenus doivent être modérés compte tenu des limites de notre évaluation. D'une part, notre échantillon relativement réduit n'est pas très représentatif. Afin de conforter nos propositions, il faudrait mener une évaluation quantitative auprès

d'un plus large panel de lecteurs. D'autre part, l'utilisation d'un extrait du Discours de la Méthode limite le nombre de liens et de textes cibles. Il faudrait éprouver notre démonstrateur avec un plus grand ensemble de données, observer comment le prototype réagit et, grâce à une nouvelle mise en situation, évaluer la concordance des résultats. Enfin, la comparaison entre l'exercice demandé à l'aide de notre démonstrateur et le même exercice avec les versions papier des ouvrages permettrait de mesurer l'efficacité de notre LNS par rapport aux livres papier.

Conclusions et perspectives

- 67 Nous avons exploité les outils et technologies du Web de Données pour proposer un modèle réutilisable et interopérable de livre numérique sémantique qui, grâce aux liens profonds définis par notre ontologie de la transtextualité et à une interface permettant l'affichage parallèle des textes, favorise l'exercice de la réflexion personnelle en recréant les conditions idéales d'appropriation d'un texte : le replacer au centre d'un réseau de textes qui, l'expliquant, le critiquant, le remplaçant dans son contexte, lui donne toute sa signification.
- 68 Contrairement aux liens hypertextes, les liens profonds que nous proposons permettent de lier plusieurs ressources à une même ancre, de définir le type de relation qu'elles entretiennent, et d'enregistrer l'ensemble des informations nécessaires concernant chaque liaison. Afin de définir ces différentes relations nous avons modélisé une ontologie de la transtextualité. Cette ontologie permet au livre numérique, non seulement d'assister le lecteur dans le repérage et le tri des liaisons, mais également de rendre le modèle interopérable.
- 69 Le livre numérique sémantique possède une interface permettant la lecture parallèle du texte source et des textes cibles. Cet affichage en vis à vis permet de replacer le passage ciblé dans son contexte d'origine tout en gardant à l'écran le passage du texte source qui a motivé l'activation du lien, et ainsi permettre la lecture simultanée des 2 textes.
- 70 Quant aux perspectives de cette recherche, nous envisageons d'éprouver notre ontologie de la transtextualité sur d'autres types de corpus. De cette manière, nous pourrions évaluer si les catégories

transtextuelles que nous avons modélisées sont pertinentes et suffisantes à l'annotation d'autres types de textes. Et, le cas échéant, compléter notre ontologie pour l'adapter à tous les genres de corpus.

- 71 Notre dispositif de lecture n'affiche, pour l'instant, que des ressources textuelles. Nous souhaitons l'ouvrir à d'autres types de médias – images, sons, vidéos – et définir, grâce aux technologies sémantiques, des catégories médiatiques. Grâce à celles-ci, il sera possible de spécifier si les liaisons proposées pointent vers un texte, une image, un son, une vidéo ou un contenu mixte.
- 72 Jusqu'à présent, les métadonnées sont entrées « à la main », ou plus précisément, elles sont extraites des données XML que nous avons récoltées dans le cadre du projet Descartes. Nous souhaitons proposer une interface de saisie de ces métadonnées qui permettrait d'enregistrer directement les liaisons aux formats RDF. Nous pourrions également imaginer une interface commune de lecture et d'annotation permettant à chaque lecteur de proposer ses propres découvertes transtextuelles.
- 73 Enfin, nous aimerions réaliser une interface pour le LNS adaptable à chaque type de dispositif numérique de lecture. En effet, la disposition parallèle des textes semble efficace sur un écran d'ordinateur mais doit être repensée pour les systèmes d'affichage plus petits. Le challenge est grand car nous désirons tout de même conserver, en l'adaptant au terminal de lecture, la juxtaposition des textes.

Barthes R. (1974). Texte (théorie du). Dans Encyclopaedia Universalis.

Berners-Lee T., Hendler J. et Lassila O. (2001). The Semantic Web, A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 34-43.

Studer R., Benjamins R. V. et Fensel D. (1998). Knowledge engineering: principles and methods. *Data & Knowledge Engineering*, 25(1-2), 161-197.

Borel C. et Roxin I., (2013). Deep Links Implementation for Semantic Digital Books. The 12th International Conference on Informatics in Economy (IE 2013) Education, Research & Business Technologies, Bucarest, 25 - 28 April 2013.

Genette G. (1982). *Palimpsestes, la littérature au second degré* (Le Seuil).

Guha R. V., Lassila O., Miller E. et Brickley D. (1998). Enabling Inferencing. Ex-

trait de <http://www.w3.org/TandS/QL/QL98/pp/enabling.html>

Hufschmitt B. (2009). Génération automatique de liens dans un texte de philosophie? Relation d'expérience: L'exemple du « Discours de la Méthode » de Descartes. Dans H2PTM'09 Rétrospective et Perspective 1989 - 2009. Présenté au Hypertextes et hypermédias. Produits, Outils et Méthodes, Université Paris 8.

ISO. (2009). ISO 15836:2009 Information et documentation -- L'ensemble des éléments de métadonnées Dublin Core. Norme ISO. Extrait de http://www.iso.org/iso/fr/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=52142

Klyne G., Carroll J. J. et McBride B. (2004, octobre 2). Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation.

Manola F. et Miller E. (2004, octobre 2). RDF Primer. W3C Recommendation.

McGuinness D. L. et van Harmelen F. (2004, octobre 2). OWL Web Ontology Language Overview. W3C Recommendation. Extrait de <http://www.w3.org/TR/owl-features/>

Nelson T. H. (1999). Xanalogical structure, needed now more than ever: parallel documents, deep links to content, deep versioning, and deep re-use. *ACM Computing Surveys*, 31(4), 33. doi:10.1145/345966.346033

Otlet P. (1934). *Traité de documentation: le livre sur le livre, théorie et pratique* (Editiones Mundaneum.). Palais mondial, Bruxelles: Editiones Mundaneum.

Prud'hommeaux E. et Seaborne A. (2008, janvier 15). SPARQL Query Language for RDF. W3C Recommendation.

Tel Quel. (1968). *Théorie d'Ensemble* (Seuil.). Paris.

1 L'Office québécois de la langue française (OQLF) propose de traduire *framework* par « cadre d'applications ». On trouve également les termes de « canevas », « cadriciel » ou « socle d'applications », comme traductions possibles. Nous utilisons, quant à nous, le terme anglais.

Français

Tout texte est au centre d'un réseau de textes qui, le remplaçant dans son contexte, l'expliquant, le critiquant, lui donne toute sa signification. Le livre numérique offre l'avantage de pouvoir recréer, à l'intérieur même du texte, les liaisons qui existent entre ces textes, ou passages de textes, tout en permettant l'accès et l'affichage immédiat de tous ces textes. Mais il faut préalablement créer et modéliser le tissage informationnel pour chaque contenu qui s'offre à lire. Ainsi, nous proposons l'exploitation des langages et outils sémantiques afin de définir un modèle réutilisable et interopérable pour les livres numériques.

Le recours aux outils sémantiques et aux « liens profonds » va bien au-delà de la simple liaison hypertextuelle. Grâce à une architecture technique adaptée, il est possible d'implémenter de manière efficiente ces liaisons pour un livre numérique. Pour ce faire, nous introduisons le concept de livre numérique sémantique avec une architecture sur trois niveaux. Le premier niveau, textuel, est constitué de l'ensemble des ressources numériques textuelles. Les métadonnées sur ces ressources se trouvent au deuxième niveau, descriptif, et sont enregistrées dans une base de connaissances assertionnelle modélisée en triplets RDF. Le troisième niveau, sémantique, contient les connaissances théoriques du corpus formalisées en ontologies (RDFS, OWL) qui définissent le vocabulaire utilisé par les métadonnées du niveau précédent.

L'implémentation de cette architecture pour l'édition d'un corpus philosophique autour du Discours de la Méthode de Descartes nous a conduit au développement d'une ontologie de la transtextualité baptisée TROW (pour Web Ontology of the TRanstextuality). La typologie retenue pour notre ontologie est issue de la théorie de l'intertextualité (initiée par le groupe Tel Quel puis développée par Genette) à laquelle s'ajoute une typologie plus spécifique de l'intertextualité pour les textes philosophiques, typologie proposée par notre laboratoire dans le cadre du projet Descartes.

Dans le prototype fonctionnel d'un livre numérique sémantique que nous avons réalisé pour le Discours de la Méthode, nous avons également créé une interface adaptée à la lecture parallèle des textes. L'expérimentation de ce dispositif auprès d'un panel de lecteurs a, par ailleurs, permis de valider nos hypothèses de recherche.

Clément Borel

Docteur en Sciences de l'Information et de la, ELLIADD - EA 4661 - UFC

IDREF : <https://www.idref.fr/228238730>