



HAL
open science

Évolution de la relation hypertextuelle partenariale entre sites institutionnels et commerciaux

Peggy Cadel, Gabriel Gallezot, David Reymond

► **To cite this version:**

Peggy Cadel, Gabriel Gallezot, David Reymond. Évolution de la relation hypertextuelle partenariale entre sites institutionnels et commerciaux : De la citation primaire à la citation construite. *Balisages*, 2023, 6, 10.35562/balisages.1109 . hal-04230000

HAL Id: hal-04230000

<https://hal.science/hal-04230000>

Submitted on 10 Oct 2023

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Balisages

Coordonné par
Ghislaine Chartron,
Benoît Epron et Pascal Robert

NUMÉRO
**30 ANS APRÈS
LE WEB... BIG
BANG DANS LES
INSTITUTIONS...**

6

ÉVOLUTION DE LA RELATION HYPERTEXTUELLE PARTENARIALE ENTRE SITES INSTITUTIONNELS ET COMMERCIAUX

De la citation primaire à la citation construite

Peggy Cadel

TransitionS, Université Côte d'Azur, France
Peggy.CADEL@univ-cotedazur.fr

Gabriel Gallezot

TransitionS, Université Côte d'Azur, France
Gabriel.GALLEZOT@univ-cotedazur.fr

David Reymond

IMSIC, EA 7492, Université de Toulon, Aix-Marseille Université, France
david.reymond@univ-tln.fr

Dans une perspective historique et évolutive, nous proposons une étude théorique et méthodologique de la relation hypertextuelle. Dans un premier temps, nous décrivons le passage d'une citation primaire à une citation construite en lien avec les transformations du Web en termes de technologies, de supports et d'usages. Nous proposons ensuite une méthodologie fondée sur une analyse des marqueurs lexicaux en capacité d'appréhender les formes de citations. L'objectif de cette méthodologie est d'identifier et de caractériser les traces de relations sémantiques sous-jacentes dans les stratégies de communication entre sites web organisationnels (SWO) pour une future étude de cas.

Mots-clés: relation hypertextuelle, lien hypertexte, cocon sémantique, marqueurs lexicaux, moteurs de recherche, SEO, sites web organisationnels

From a historical and evolutionary perspective, we propose a theoretical and methodological study of the hypertextual relationship. First, we describe the transition from primary to constructed quotation in relation to the Web's transformations in terms of technologies, media and uses. We then propose a methodology based on an analysis of lexical markers capable of apprehending citation forms. The aim of this methodology is to identify and characterize traces of underlying semantic relations in

communication strategies between organizational websites for a future case study.

Keywords: hypertextual relationship, hyperlink, semantic cocoon, lexical markers, search engines, SEO, organizational websites

Dans cet article, nous proposons d'étudier l'évolution de la relation hypertextuelle de la citation primaire à la citation construite. Nous commençons par définir le lien et ancrer les relations hypertextuelles dans une perspective historique et évolutive. Cette description nous permet d'appréhender les transformations en termes d'échanges et de rédaction soumises à l'évolution des technologies, des supports, des usages et des mutations de la Search Engine Page Result (SERP)¹ (Andrieu, 2022).

Nous proposons par la suite une méthodologie pour interroger la relation hypertextuelle partenariale entre sites institutionnels et sites commerciaux, catégories de sites web organisationnels (SWO), sur la base des marqueurs lexicaux, afin d'identifier les traces et caractéristiques de relations sémantiques sous-jacentes témoignant de la mise en place de stratégie de communication orchestrée entre les différents partenaires. Cet article ne présente donc pas une étude de cas (en cours de réalisation par ailleurs) mais une approche théorique et méthodologique de la relation hypertextuelle.

LES RELATIONS HYPERTEXTUELLES, CONTEXTUALISATION HISTORIQUE

Dans cette première partie, nous nous attachons à qualifier le lien hypertexte en le comparant aux différents processus de mise en relation et d'aide à la navigation textuels ou non textuels et en le définissant à travers ses caractéristiques techniques et leurs évolutions.

Définition du lien hypertexte

Dans son article de 2003, Olivier Ertzscheid (2003) étudie le rapport entre la note de bas de page et le lien hypertexte en revenant sur les origines du lien hypertexte décrit comme un processus d'annotation de contenu :

L'hypertexte est donc, a priori, un système informatique permettant d'annoter tout type de contenu sur un mode associatif, afin de renvoyer

1. Page de résultats des moteurs de recherche.

vers des contenus similaires et/ou de pouvoir conserver en mémoire (informatique) la trace du cheminement de la pensée (parcours de lecture ou parcours interprétatif).

Le lien hypertexte ainsi ancré dans la citation (et / ou la citation de recherche), l'association sémantique (Charlet et Kembellec, 2016), l'inscription de trace (Ertzscheid, Gallezot et Simonnot, 2013) et le parcours de lecture (Saleh, Papy et Bouhai, 2001) associent un ou des éléments, textuel(s) ou non textuel(s) désigné(s) par un auteur à une destination elle aussi textuelle ou non textuelle. Nous postulons que les unités lexicales associées aux hyperliens clefs (les «textes passeurs», Jeanneret, 2000) portent une sémantique significative par rapport au site web cité. En d'autres termes, ces unités lexicales d'hypertexte (ULH), en tant que termes génériques d'accès à des contenus sous-jacents constituent de fait des métadonnées informatives vis-à-vis des contenus sous-jacents (Reymond *et al.*, 2011; Pinède et Reymond, 2010).

À ces objets textuels s'ajoutent des objets non textuels permettant l'illustration des mots cités et/ou la mise en mot des éléments non textuels par le biais d'un lien hypertexte renvoyant vers un nouveau contenu.

Si le lien hypertexte est textuel, il s'inscrit dans un contexte lexical marqué par les ULH, le nom de fichier de la page de destination, le titre du lien permettant de proposer un texte au survol, texte à la fois utilisé par le lecteur, les moteurs de recherche et les solutions relevant de l'accessibilité et enfin un contexte linguistique étendu au paragraphe dans lequel se situe le lien. Ces trois zones guident le visiteur au travers d'un parcours de lecture destiné plus encore au moteur de recherche qu'au lecteur, devenant un maillage sémantique.

La typologie simplifiée ci-dessous nous permet par la suite de mieux nous focaliser sur les marqueurs lexicaux :

- l'URL, signe par sa racine protocole et l'adresse principale du site web/du serveur (adresse IP, DNS, etc.). L'absence de cette racine désigne les pages relatives. Le changement de cette racine renvoie à tout le Web, le plus souvent d'autres serveurs (mais des zones Domain Name System différentes peuvent cependant pointer le même contenu/site/organisation);
- le chemin permet la navigation dans les pages du serveur, dans l'arborescence des fichiers dispensés;
- l'ULH parfois associée d'une balise (#ancree) ou appliquée à une image (bouton), permet la navigation dans la page affichée;
- les pages calculées (par exemple avec des Common Gateway Interface) produisent des contenus générés de type variés. Ce peut être des adresses (par exemple les résolveurs de lien, les raccourcisseurs de lien), ou des

contenus issus par traitement et calculs à partir des paramètres issus des requêtes des usagers utilisés pour composer un contenu à partir d'une base de données (on pensera aux liens des Content Management System [CMS], réseau sociaux, etc.).

Les temps du lien hypertextuel

Si on peut faire remonter la notion d'hypertexte très loin dans l'histoire (par exemple la pierre de Rosette en ce qu'elle crée du lien entre trois textes, chaque partie, chaque signe du texte), nous nous concentrons ici sur la notion d'hypertexte «réduite» aux simples relations entre éléments, sans traiter la cardinalité, ou la qualification des liens.

Nous faisons démarrer la notion de lien hypertextuel avec Paul Otlet et Henri La Fontaine et leur Répertoire bibliographique universel (RBU) et la Classification décimale universelle (CDU) à la fin du XIX^e siècle (Fayet-Scribe, 2000). Ils jettent les bases d'un système de relations entre documents via une indexation normée créant ainsi un «réseau de connaissances». L'objectif était de collecter et d'organiser toutes les connaissances humaines dans une bibliothèque universelle permettant d'accéder au savoir. Le RBU pourrait s'apparenter au moteur de recherche par son indexation et la restitution de fiche, mais aussi à un wikimédia puisque la navigation entre les fiches est possible.

Le deuxième temps est traditionnellement marqué par l'article "As We May Think" de Vannevar Bush (1945) dans lequel il propose l'idée d'une machine appelée «memex» qui permettrait aux utilisateurs de stocker, d'organiser et de naviguer dans des documents. Il y a ici non seulement l'idée de naviguer au gré de ses pensées, mais également de générer automatiquement des liens hypertextes.

Un troisième temps est marqué par l'invention de la souris par Douglas Engelbart dans les années 1960 et un premier ordinateur équipé en 1973. Sans ce périphérique, les notions de «clic», de lien cliquable, de dénombrement et calcul du «clic» n'auraient jamais émergées. On considéra que le «tactile», même s'il change le rapport à l'écran, n'est qu'une excroissance de la souris filaire en ce qui concerne les ULH.

Le quatrième temps est constitué de l'invention du terme hypertexte (1965) par Ted Nelson et son projet Xanadu (Nelson, 1987), une version numérisée du Mondotheque² d'Otlet et une ébauche informatisée du Memex et une préfiguration du Web. Ce projet très ambitieux a fait l'objet de développements

2. Pour se faire une idée de la Mondotheque d'Otlet: <https://artsandculture.google.com/asset/mondoth%C3%A8que/fQF395S-gTN2Mg?hl=fr>.

successifs, de scissions et rachats divers pour finalement inspirer le World Wide Web.

Le cinquième temps est celui du Web créé par Tim Berners-Lee notamment à ses débuts via les URI (Uniform Resource Identifier), le HTTP (Hypertext Transfer Protocol) et le HTML (HyperText Markup Language) [Berners-Lee, 1990] mais aussi sans oublier celui qu'il a défini «le Web de données» et qui prend forme sous nos yeux via, entre autres, le RDF (Resource Description Framework), SPARQL (Simple Protocol and RDF Query Language), et OWL (Web Ontology Language). Quelle que soit sa version, le principe du Web est de créer un «réseau de connaissance» (pour reprendre ici les termes d'Otlet), un hypertexte mondial (selon les termes de Tim Berners-Lee) qui permet aux utilisateurs de trouver, partager et combiner des informations provenant de différentes sources.

Le Web de données de Tim Berners-Lee inspire à son tour des projets comme celui de Liquid information (Frode Hegland)³. Un projet qui est resté au stade de démonstration, mais est intéressant dans sa volonté de rendre tous les mots de la page web cliquables (hypermots) pour opérer une action: lancer une recherche sur le mot, définir le mot, traduire le mot, etc. Si l'image du flux de l'eau comparable au flux informationnel fonctionne assez bien, le principe de multiplicité de contextualisation des actions à partir de tous les mots de la page semble tellement vaste, qu'il renvoie à des paramétrages de cardinalité du lien effectués par le lecteur, ce qui le noie ou le submerge de possibilités et d'information.

La redocumentarisation (Pédaque et Salaün, 2006) liée au numérique change la relation hypertextuelle non seulement dans sa forme, c'est-à-dire l'ensemble des possibilités informatiques en capacité de générer différents types de liens (voir la partie Évolution et usages), mais également dans sa nature, c'est-à-dire la transition entre des liens associés aux savoirs académiques à ceux liés au monde marchand (voir la partie Des institutions de savoirs aux organisations commerciales).

ÉVOLUTION ET USAGES

Au premier temps du Web, le maillage s'effectue «à la main» entre page d'un même site ou vers un site externe au gré des rencontres, des découvertes de sites ou de partenariats à afficher. L'essor du web inscriptible (nommé aussi Web 2) a accru ce maillage par oblitération du code informatique opéré par des interfaces WYSIWYG (What you see is what you get) facilitant considérablement

3. <http://www.liquidinformation.org/>.

la publication. La création et le développement de contenus ont favorisé une concurrence accrue entre les mots-clés sous-tendue par l'importance accordée à l'indice de popularité des pages dans les critères de pertinence (relevance)⁴ des moteurs de recherche. Les sites web organisationnels (SWO) à vocation commerciale se sont emparés de ce levier pour améliorer leur visibilité. Au travers d'un nouveau design (au sens de conception) d'une page web conçu pour offrir, via un maillage de liens « internes » ou partenaires, un espace dans lequel le contenu peut se développer, puis s'exposer efficacement à l'indexation des moteurs de recherche la relation hypertextuelle est passée du cocon sémantique à la prison commerciale des applications (notamment sur smartphone): on garde les principes d'hyperlien, d'hypertexte en étant encapsulé, enfermé dans un environnement contrôlé par le marchand.

Évolution des marqueurs linguistiques

Le texte associé au lien, dit autrement l'UHL ou l'ancre cliquable, propose un parcours de lecture initié par des mots mis en exergue. Les « cliquez ici » ou « en savoir plus » permettant de guider le visiteur ont été remplacés par des marqueurs conventionnels où la couleur (bleu par défaut) et le souligné habillent des éléments linguistiques à forte valeur sémantique. Le poids associé au texte des liens (aux UHL) pèse à double titre dans les algorithmes de recherche en valorisant la thématique de l'UHL sur la page portant le lien et en attribuant la sémantique de la page de destination du lien. Le développement de maillage sémantique tel que travaillé en Search Engine Optimization (SEO) s'éloigne de l'accompagnement naïf du lecteur dans un parcours de lecture ou un parcours interprétatif pour se construire autour d'un parcours sémantique artificiel destiné aux moteurs de recherche: le cocon sémantique.

Le cocon sémantique (concept, technique ou stratégie) selon Laurent Bourelly⁵ consiste à organiser les pages web autour d'un sujet principal, en créant un ensemble de pages connexes qui se concentrent sur des aspects spécifiques de ce sujet. Les UHL des liens contiennent des mots-clés qui renforcent le contexte sémantique des pages liées et orientent les visiteurs. Cette technique vise à organiser un maillage interne sur des UHL composées de plusieurs mots permettant de décrire sémantiquement un champ. L'association de ces mots forme des syntagmes nominaux de plus en plus longs correspondant à l'allongement du nombre de mots utilisés dans les requêtes

4. Nous distinguons ici ce qui est nommé « pertinence » par le moteur de recherche, qui s'appuie en réalité sur des critères de « relevance » : le document (lien) renvoyé relève bien des termes utilisés (avec tous les paramètres et la pondération des résultats), est-ce pour autant pertinent pour le lecteur? Seul un clic lèvera un peu le doute.

5. <https://www.laurentbourrelly.com/blog/categorie/cocon-semantique>.

envoyées aux moteurs de recherche. Le maillage de ces syntagmes repose à la fois sur une organisation terminologique réelle et sur une prolifération de pages artificiellement construites. On assiste ainsi à la création de réseaux terminologiques dans lesquels l'énonciation réticulaire s'efface devant des constructions en silo. La confiance est ainsi distribuée sous forme de « jus » de liens porté par le poids du texte associé au lien améliorant la popularité d'un site ou d'une partie de site sur une thématique donnée.

Ces constructions en silo s'appuient sur deux marqueurs : les marqueurs lexicaux et les marqueurs de confiance. La thématique large des pages citantes vient renforcer la thématique spécifique des pages citées qui elles-mêmes deviennent des pages citantes. Les relations ne sont pas hiérarchisées mais récursives.

Évolution des critères de « pertinence »

L'enjeu pour les professionnels du SEO est désormais de développer des pratiques et des tactiques (Alloing, 2016) destinées à anticiper les nouveaux critères de pertinence et les mutations de la SERP.

À cela s'ajoute l'avènement de l'auctorisation (Zacklad, 2019) comme nouveau critère d'affichage fondé sur l'Expertise, Authoritativeness and Trustworthiness (EAT) et plus récemment sur l'EEAT⁶ (Experience – à prendre comme l'expérience de l'auteur en lien avec le contenu rédigé – Expertise, Authoritativeness and Trustworthiness). Ainsi, les deux représentations distinctes du Web telles que décrites par Cardon (2013a), le Web comme espace documentaire pour le PageRank de Google et le Edgerank de Facebook comme Web de personnes et mesure d'autorité tendent à se rapprocher tout en restant séparés.

Afin d'exister sur la toile, les organisations commerciales se soumettent alors à des contraintes rédactionnelles fortes, avec un contenu formaté pour les moteurs de recherche : un vocabulaire diversifié (pour tenir compte des synonymes en évitant les pronoms), répétitif (pour augmenter le taux d'occurrences), tout en s'appuyant sur l'*UX writing*⁷ avec des paragraphes courts, des phrases courtes et des fractures visuelles.

De leur côté, les organisations institutionnelles bénéficient d'un allègement de ces contraintes en termes de visibilité et d'existence, liées à une forte notoriété hors ligne (Alloing, 2016) dans un domaine non concurrentiel avec

6. La documentation officielle du « General Guidelines de Google » n'étant plus mise à jour depuis le 15 décembre 2022, voir : <https://developers.google.com/search/blog/2022/12/google-raters-guidelines-e-e-a-t?hl=fr>.

7. Écriture centrée sur l'expérience utilisateur.

une appellation et / ou un nom de domaine non ambigu qui permet au visiteur d'accéder avec une requête simple à la page d'accueil de l'organisation.

C'est pour cette raison que nous avons choisi d'interroger la constitution des maillages entre SWO.

DES INSTITUTIONS DE SAVOIRS AUX ORGANISATIONS COMMERCIALES

Dans cette partie, nous proposons une méthodologie pour étudier les hyperliens entre sites institutionnels et non institutionnels à partir de l'analyse des pages (contenu et UHL) de sites partenaires tels que définis par Reymond (2022), deux sites étant partenaires s'ils opèrent une réciprocité de liens.

De la conception du lien hypertexte envisagé comme un moyen de naviguer dans un «réseau de connaissance» de documents papier (à l'origine) ou numériques (avec le web, voir la partie Les relations hypertextuelles, contextualisation historique), la relation hypertextuelle est progressivement devenue un lien marchand, qu'il s'agisse du bouton «Acheter» ou de la monétisation publicitaire du «clic» ou du «like». Les «Autoroutes de l'information» (Du Castel, 1995) vantées comme un formidable moyen de s'instruire, de connaître le monde ont surtout promu un excellent moyen de commercer bien plus que de l'information, jusqu'à instituer une économie de l'attention capable d'enfermer l'Homme dans un supermarché réticulaire.

De même, si le principe de pertinence des moteurs de recherche s'appuie sur le principe de «qui cite qui» bien connu dans le monde académique avec la bibliométrie, la pondération avec des centaines de critères est opaque, versatile et orientée publicité. Sur le terrain du Web, les activités académiques et mercantiles se mêlent dans un graphe dont des lectures sont proposées par des outils de recherche qui restituent des listes, des catégorisations, des cartes, des synthèses, etc.

Relation hypertextuelle partenariale

Notre objectif est de dépasser le cadre procédural du lien hypertexte en y apportant un éclairage linguistique sur les intentions données aux liens entrants et sur la confiance accordée grâce aux liens sortants. Avec les sites partenaires on ne se limite pas à la notion de confiance. Les sites se citant doivent le faire en accentuant la cohésion thématique bi-directionnelle des liens (hyponymie, synonymie, hyponymie).

Aussi, nous proposons une méthodologie qui s'appuie sur des cahiers exécutables modulables (Reymond *et al.*, 2022). Ces instruments permettent

de disposer d'une extraction morpho-syntaxique permettant de représenter distinctement verbes et substantifs et de déterminer les écarts de contenu entre pages partenaires. Dans une approche de linguistique informatique (Silberztein, 2019), cette analyse est à compléter en exploitant l'ensemble des marqueurs linguistiques, qu'il s'agisse d'éléments d'écriture destinés aux moteurs de recherche ou d'éléments d'écriture destinés au lecteur. L'extraction de lemmes regroupés en catégories grammaticales et couplés au repérage de marqueurs rédactionnels est en capacité d'affiner le traitement réalisé par les instruments exécutables en tenant compte du contexte d'écriture contraint par l'utilisation des liens hypertextes.

Le site web organisationnel

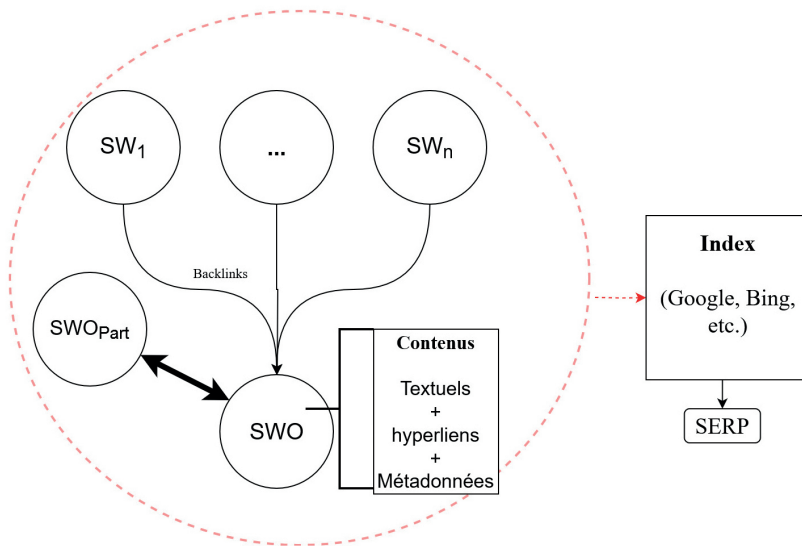
Pour une organisation, la communication sur le web passe par l'élaboration d'un site. Nous reprenons l'expression "Site web organisationnel" (SWO) (Pinède & Reymond, 2013) pour caractériser ces sites associés à une instance énonciatrice de type organisationnel. Le SWO se définit comme un produit hypertextuel d'informations et de services placé sous une seule et même responsabilité. Associé à une image (Rouquette, 2009), le SWO est censé véhiculer un ou des messages en phase avec les objectifs communicationnels de l'organisation. Récemment, Pinède (2018) propose un modèle global de cette projection numérique de l'image et de l'identité d'une organisation, en la considérant au carrefour de logiques complexes. Il s'agit de situer le site web organisationnel en tant qu'unité cohérente de texture (Brügger, 2012) afin de discuter de l'identité numérique d'une organisation. Par essence, le site web couvre en communication un objectif d'image: un message en phase avec la stratégie et s'inscrit dans une cohérence formelle (design, architecture, charte, etc.) et une cohérence sémantique de contenus (Zacklad, 2019). L'ensemble est relié par des hyperliens, qui améliorent la qualité de l'indexation et la représentation du site dans les index et qui en relient cohérence sémantique et cohérence formelle pour constituer une « unité cohérente de texture » des actions physiquement performatives (Pinède, 2018; Brügger, 2012).

Ainsi, dans ce modèle d'analyse des identités numériques organisationnelles, Pinède (2018) étudie comment les organisations se donnent à voir, sont perçues et sont évaluées.

La figure page suivante (Couao-Zotti et Reymond, 2022) représente l'extension du modèle SWO proposé sur le plan des contenus et de leur représentation informationnelle sur les index. Le SWO est, en interne, un ensemble de contenus textuels et de métadonnées, renforcé par des hyperliens structurant une topologie. Cette dernière s'inscrit dans le Web par des hyperliens externes: des citations d'autres SWO que nous considérons comme partenaires (SWO_Part)

si réciproque, et cité par d'autres sites web (SWO ou pas): les backlinks (de 1 à n). Notons que dans ce modèle sont exclus les liens vers et depuis les réseaux sociaux (RSN) qui, bien qu'ayant leur intérêt pour la communication web, relèvent de stratégies différentes dans leur mise en place et ne participent pas au référencement hors de ces réseaux. Nous considérons comme « organisations partenaires » les sites visés par l'ensemble des liens sortants du SWO dont émane un lien entrant. Nous nous attendons à observer une certaine cohérence sémantique dans les liens mis en place entre ces organisations partenaires, qui se placent par citation mutuelle en tant qu'extension de la structure endogène de SWO (Pinède, 2018). Notre objectif est d'étudier l'unité cohérente de structure que constituent ces sites au plan linguistique.

Figure. Représentation extensionnelle du modèle SWO



Légende : SWOPart sont les partenaires institutionnels considérés comme participant à la communication web, cités par et citant le SWO d'origine. Avec les autres liens entrants, ils participent tous à la représentation sémantique du SWO original sur les index.

Avec ce modèle organisationnel, nous retrouverons le cocon sémantique par les étapes suivantes :

- reconstitution de la topologie relationnelle entre les sites partenaires ;
- extraction des contenus contextuels des hyperliens et segmentation selon les niveaux offerts par le langage HTML (balise alt, titre, ULH, paragraphe) ;
- analyse des relations hypertextuelles.

Dans un projet en cours de réalisation, nous appliquerons cette méthode au cas du site web du parc national de Port-Cros pour lequel nous avons déjà analysé l'inscription hypertexte globale à partir des liens entrant sur le site (Reymond *et al.*, 2022).

PERSPECTIVES / DISCUSSIONS

Afin de mesurer l'évolution des liens hypertextes au travers des marqueurs lexicaux associés, nous proposons d'étudier les échanges numériques entre sites institutionnels et sites commerciaux sur la base de l'analyse des ULH et du contenu textuel des pages contenant les ULH.

La méthodologie proposée permettra d'appréhender le lien comme une trace sémantique, un indicateur lexical, un marqueur de confiance, un marqueur d'accessibilité, un indicateur électif, une énonciation réticulaire (Ertzscheid *et al.*, 2013).

Nous analyserons la réciprocité au sein des échanges entre SWO. Dans ces échanges émanant d'une zone éditoriale contrôlée ou non, le lien peut être considéré comme un vote de confiance (Cardon, 2013b) pour lequel l'attribution des voix s'inscrit dans le cadre d'une stratégie de communication où les sites institutionnels pourraient bénéficier algorithmiquement du statut de grands électeurs.

L'analyse des liens hypertextes devra tenir compte de l'évolution constante des algorithmes avec l'essor des marqueurs d'accessibilité (Pinède, 2022), l'atténuation des sanctions pour les achats de liens, l'absence de sanction pour les contenus émanant de générateurs automatiques, la proposition d'une SERP continue et le maintien des utilisateurs dans cette SERP avec les moteurs de réponses et l'enfermement dans un Web constitués d'applications propriétaires.

En appliquant cette méthodologie sur différents projets, plusieurs enseignements sont attendus.

La notion de cocon sémantique est-elle utilisée, du moins consciemment travaillée dans tous les domaines d'activité? Le SEO s'empare-t-il de cette

notion ou bien d'autres optimisations sont-elles privilégiées (plus rapides, moins coûteuses, plus efficaces, etc.) ?

L'analyse des relations hypertextuelles entre sites marchands et académiques montre-t-elle la prédominance de liens sortants vers le monde académique ou inversement ? Est-il possible de détecter aisément le transfert de technologie, l'application de brevet, l'open innovation, le spin-off, ou plus simplement les partenariats public-privé ?

Assiste-t-on à une désémantisation de la relation hypertextuelle ou au contraire à une « sursémantisation » ? Deux développements (évolutions ? transformations ?) sont en effet à l'œuvre sous nos yeux : celui de la mise en application du Web (au sens des applications de nos smartphones), voire de la mise en objet (Internet of Things, IoT) du Web et celui des intelligences artificielles (IA) génératrices de réponses (qui utilisent des outils et méthodes à caractère sémantique). Ces développements transforment les relations hypertextuelles soit en les vidant de sens, du moins en les cantonnant à une simple branche (lien primaire) d'une arborescence minimaliste soit en les construisant « à la volée ». Les relations hypertextuelles ainsi construites s'appuient sur des termes renseignés par un utilisateur et un référentiel sémantiquement traité au préalable (sursémantisation). Le résultat devient une réponse dont l'utilisateur ne voit pas les traces de construction (l'arborescence des liens/liaisons de l'algorithme).

Ainsi, la citation primaire tend à disparaître sous l'effet d'enfermement dans les applications (ou alors s'entend comme auto-citation) et la citation construite tend à être cachée par oblitération des sources du texte exposé. On peut également s'interroger sur une artificialisation du lien et du sens : le SEO utilisant des IA, et les générateurs de textes également des IA. Comment ne pas percevoir des boucles dangereusement autoréférentielles ?

BIBLIOGRAPHIE

Andrieu, O. (2022). *Réussir son référencement, Stratégies et SEO*. Édition 2022-2023. 11^e éd. Paris : Eyrolles.

Alloing, C. (2016). La fabrique des réputations selon Google. *Communication & Langages*, 188, 101-122. DOI : <https://doi.org/10.4074/S0336150016012060>.

Berners-Lee, T. (1990). « Proposal for a HyperText Project ».

Brügger, N. (2012). L'historiographie de sites web : quelques enjeux fondamentaux. *Le Temps des médias*, 18 (1), 159-169. DOI : <https://doi.org/10.3917/tmd.018.0159>.

Bush, V. (1945). As We May Think. *The Atlantic*. 1^{er} juillet 1945. Repéré à : <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>.

- Cardon, D. (2013a). Du lien au like sur Internet : deux mesures de la réputation. *Communications*, 93 (2), 173. DOI : <https://doi.org/10.3917/commu.093.0173>.
- (2013b). Dans l'esprit du PageRank : une enquête sur l'algorithme de Google. *Réseaux*, n° 177 (1), 63-95. DOI : <https://doi.org/10.3917/res.177.0063>.
- Charlet, J. et Kembellec, G. (2016). Du web sémantique au web des données, quels enjeux professionnels ? *I2D - Information, données & documents*, 53 (2), 54-55. DOI : <https://doi.org/10.3917/i2d.162.0054>.
- Couao-Zotti, K. et Reymond, D. (2022). « Distorsions de la communication web : analyse des contenus reliés par hypertexte », Séminaire IMSIC axe 1 : brevets, captas et Science ouverte, Toulon.
- Du Castel, F. (1995). Autoroutes de l'information et société de communication. *Réseaux. Communication - Technologie - Société*, 13 (71), 107-115. DOI : <https://doi.org/10.3406/reso.1995.2694>.
- Ertzscheid, O. (2003). « De la note de bas de page au lien hypertexte : philosophie de l'identique et stylistique de l'écart ».
- Ertzscheid, O., Gallezot G. et Simonnot, B. (2017). « À la recherche de la "mémoire" du web : sédiments, traces et temporalités des documents en ligne ». Dans *Manuel d'analyse du Web*. 2^e éd. Paris : Armand Colin, 60-76. DOI : https://archivesic.ccsd.cnrs.fr/sic_01687801.
- Fayet-Scribe, S. (2000). *Histoire de la documentation en France : culture, science, et technologie de l'information, 1895-1937*. Paris : CNRS.
- Jeanneret, Y. (2000). *Y a-t-il (vraiment) des technologies de l'information ?* Villeneuve d'Ascq : Presses universitaires du Septentrion.
- Nelson, T. H. (1987). *Literary machines: the report on, and of, Project Xanadu concerning word Processing, electronic publishing, hypertext, thinkertoys, tomorrow's intellectual revolution, and certain other topics including knowledge, education and freedom*. Ed. 87.1. Swarthmore, Pa: Theodor H. Nelson. Repéré à : <http://books.google.com/books?id=QsDaAAAAMAAJ>.
- Pédaque, R.-T, et Salaün J.-M. (2006). *Le document à la lumière du numérique*. Caen : C&F Éditions.
- Pinède, N. (2022). Handicap et numérique : entre besoins, opportunités et paradoxes. *Servir*, vol. 516, n° 6, 37-40.
- Pinède, N. (2018). Du site web aux identités numériques organisationnelles. *Questions de communication*, n° 34, 75-94. DOI : <https://doi.org/10.4000/questionsdecommunication.15587>.
- Pinède, N. et Reymond D. (2010). « De la diversité au lissage informationnel : création d'une taxonomie inductive pour les sites web universitaires ». Repéré à : https://archivesic.ccsd.cnrs.fr/sic_01279674.
- . (2013). Classer les sites web organisationnels. Une approche taxonomique des liens hypertextes. *Hermès, La Revue*, 66 (2), 181-88. DOI : <https://doi.org/10.4267/2042/51573>.

Reymond, D., Couao-Zotti, K., Tabariès A., Lebourgeois A., et Campos L. (2022). Analyse lexicale d'une page web. *Revue française des sciences de l'information et de la communication*, n° 24. DOI: <https://doi.org/10.4000/rfsic.12365>.

Reymond, D., Pinède, N., Lespinet-Najib, V. et Le Blanc, B. (2011). « Une étude terminologique de la communication hypertexte web. Caractéristique du domaine universitaire », 9th International Conference on Terminology and Artificial Intelligence, novembre 2011. Paris, pp. 139-142. Repéré à : <https://hal.science/hal-00674113>.

Rouquette, S. (2009). Sociologie des sites web d'entreprise. *Esprit critique: revue internationale de sociologie et de sciences sociales*, 12 (1). Repéré à : <http://espritcritique.uiz.ac.ma/pdfs/1201/esp1201article04.pdf>.

Saleh, I., Papy, F. et Bouhai, N. (2001). « HyWebMap: structurer l'information sur le Web ». In *29th Annual Conference of the Canadian Association of Information Science (CAIS/ACSI)*, Québec : Laval, Mai.

Silberztein, M. (2019). Les outils informatiques au service des linguistes : présentation. *Langue française*, n° 203 (3), 7-14. DOI: <https://doi.org/10.3917/lf.203.0007>.

Zacklad, M. (2019). Le design de l'information : textualisation, documentarisation, auctorialisation. *Communication & Langages*, n° 199 (1), 37-64. DOI: <https://doi.org/10.3917/compla1.199.0037>.