

Lindexation documentaire n'est pas une pratique nouvelle. Mais le développement d'algorithmes toujours plus performants, avec la puissance augmentée de l'intelligence artificielle, a conduit cette pratique au delà de la sphère des professionnels de l'information et a investi tous les domaines de nos vies privées et professionnelles en matière de recherche d'information : sauvegardes de l'historique de recherche, réponses à nos requêtes même les plus hasardeuses, recommandations et bulles informationnelles sont autant de procédés qui tendent à lénifier la frustration des internautes vis-à-vis du silence documentaire.

Face à ce postulat, l'Ecole est face à un enjeu sociétal et éducatif conséquent. Ces nouvelles technologies deviennent un objet d'étude et pédagogique auxquels les élèves doivent se confronter. Pour les professeurs documentalistes, cet enjeu constitue un axe d'enseignement fondamental, qu'il leur est nécessaire d'apprivoiser.

Par conséquent, nous avons choisi de consacrer deux numéros à cette thématique. Dans ce premier opus, il s'agit de mettre en exergue une approche technique afin de comprendre le fonctionnement des algorithmes et ce qu'ils induisent dans nos pratiques de recherche d'informations.

C'est le point de vue retenu par Eric Bruillard, enseignant-chercheur à Université de Paris : dans sa contribution, il se propose de mettre la focale sur l'évolution, les caractéristiques de l'indexation des moteurs de recherche par le prisme de l'intelligence artificielle.

Dans la même lignée, Elodie Renvoisé, professeure documentaliste dans l'académie de Nantes s'intéresse aux encarts informationnels, ou *info-boxes*, comme objet didactique pour les professeurs documentalistes. C'est ainsi qu'elle met en exergue quelques pistes pédagogiques pour aborder ces condensés d'informations.

Pour clore ce dossier, Florian Reynaud, professeur documentaliste dans l'académie de Lyon, soumet comme matériaux pédagogiques, deux simulateurs, *WebFinder* et *WebLink*, afin de transmettre aux élèves une culture numérique des algorithmes corrélée à la recherche d'information.

Dans la rubrique *Varia* et dans le prolongement de la thématique du dossier, Sophie Jehel, Maîtresse de conférences HDR en Sciences de l'information



et de la communication, Université Paris 8, eu égard à ses derniers travaux et étude, interroge la relation des adolescents aux plateformes socio-numériques, mettant en exergue des interactions complexes entre lesdits adolescents et l'économie numérique.

Pour conclure cette rubrique, le Bureau national de l'APDEN, via la plume de Florian Reynaud, présente les atouts et les nouveautés relatives à la réfection du *Wikinotions Info-doc*, espace numérique de ressources pour l'information-documentation et l'Éducation aux médias et à l'information.

Bien loin de ces réflexions, Hans-Petter Storemyr, bibliothécaire scolaire, en Norvège, dans un article figurant dans la rubrique « *Ailleurs* », rend compte du développement des bibliothèques, de la situation des bibliothécaires scolaires et de leurs balbutiements.

Face aux enjeux sociétaux, culturels et éducatifs que représentent les algorithmes dans le domaine de la recherche d'information, cette notion et celles qui lui sont connexes constituent pour les professeurs documentalistes des objets d'enseignement dont ils ne peuvent faire l'économie. En comprendre les rouages et leurs conséquences sur les pratiques informationnelles des élèves constituent un point d'entrée important pour transmettre une culture du numérique et de l'information. Afin de prolonger la réflexion, il importera dans un second temps de réfléchir à des pistes pédagogiques concrètes et à des alternatives pour composer avec ces technologies et les juguler autant que faire ce peut et ainsi doter nos élèves d'un esprit critique face à la recherche et au traitement de l'information.

Les bibliothèques et bibliothécaires scolaires en Norvège

Ailleurs

◆ **Hans-Petter STOREMYR**

Bibliothécaire scolaire, Lycée Saint Olav, Norvège

La Norvège est un des pays scandinaves du Nord de l'Europe et il a des frontières avec la Suède, la Finlande et la Russie (dans la partie nord).

La population norvégienne est actuellement de 5 391 369 habitants (chiffres de janvier 2021) dont 11,2% de population immigrée (données de 2021).

Il y a 2799 écoles primaires et 415 écoles secondaires en Norvège (chiffres de 2019). La plupart des écoles sont des établissements publics mais 23% des écoles secondaires sont privées¹.

Le système éducatif en Norvège

Sur une population d'environ 5,3 millions d'habitants, plus de 900 000 personnes sont

actuellement en formation initiale ou en apprentissage. De plus, approximativement un million de personnes participe à des cours de formation continue pour adultes chaque année.

La Direction Générale Norvégienne pour l'Éducation et la Formation est responsable du développement de l'éducation de la maternelle, du primaire et du secondaire. Elle constitue l'agence exécutive pour le Ministère de l'Éducation et de la Recherche².

Responsabilités

La Direction Générale a la responsabilité générale de superviser la maternelle, l'éducation et la gouvernance du secteur éducatif, ainsi que la mise en œuvre des Actes du Parlement et des règlements.

¹ Chiffres et données à propos de l'éducation Norvégienne disponibles sur : <https://www.udir.no/in-english/education-mirror-2019/>
Chiffres et données à propos de l'état de l'éducation en Norvège disponibles sur : https://read.oecd-ilibrary.org/education/education-at-a-glance-2016/norway_eag-2016-73-en#page1

² <https://www.udir.no/in-english/Quality-in-ECEC-Schools-and-Vocational-Education-and-training/>
<https://www.udir.no/laring-og-trivsel/mobbing/guide-til-skoleforeldre-med-barn-som-ikke-har-det-bra-pa-skolen/guide-for-parents-if-your-child-does-not-thrive-at-school/>
https://www.udir.no/contentassets/977da52955c447bca5fc419d5be5e4bf/the-norwegian-assessment-for-learning-programme_final-report-2018.pdf
https://read.oecd-ilibrary.org/education/education-at-a-glance-2016/norway_eag-2016-73-en#page1 (L'Éducation en Norvège rapport de l'OCDE)

Elle a également en charge l'exploitation de toutes les statistiques nationales qui concernent l'école maternelle, le primaire et le secondaire. Sur la base de ces statistiques, elle initie, développe et contrôle la recherche et le développement.

Son objectif est de garantir que tous les enfants, élèves et apprentis reçoivent l'éducation de haute qualité à laquelle ils ont droit .

Le système éducatif présente les caractéristiques suivantes :

- Tous les enfants entre 6 et 16 ans doivent aller à l'Ecole gratuitement ;
- L'Etat a la responsabilité générale de la Loi sur l'Education avec les règlements, les contenus et le financement des écoles primaire et secondaire et de l'apprentissage ;
- L'Ecole primaire : niveaux 1 à 7 ;
- le collège : niveaux 8 à 10 ;
- le lycée : niveaux 11 à 13 ;
- l'année scolaire est d'une durée de 38 semaines et est divisée en 2 semestres ;
- le lycée n'est pas obligatoire. Chaque élève peut choisir, parmi 12 programmes, de candidater pour ceux qu'il veut suivre en fonction des résultats scolaires obtenus. Le système de notation pour le collège et pour le lycée va de 1 à 6 (6 est la meilleure note). Les universités et l'enseignement supérieur utilisent le système A à F (système international) ;
- les programmes actuels incluent les 13 années de l'école primaire à la dernière année du lycée. Les attendus de fin de scolarité obligatoire sont d'apprendre à lire, calculer, écrire, s'exprimer et maîtriser les compétences numériques standards pour chaque niveau atteint. Un ensemble d'objectifs pour être "actif" dans chaque discipline est mis en œuvre chaque année et est évalué aux niveaux 2, 5, 7 et 10 avec des tests nationaux. Au lycée c'est le même plan, mais il y a des examens terminaux chaque année dans certaines matières.

Les bibliothèques

L'organisation des bibliothèques en Norvège inclut les bibliothèques de lecture publique, les bibliothèques municipales, les bibliothèques scolaires, les bibliothèques universitaires d'établissements publics et privés, les bibliothèques de recherche, la bibliothèque de livres audio et en braille, et la Bibliothèque Nationale.

Bref historique du Programme pour la Bibliothèque Scolaire en Norvège

Le Programme pour la Bibliothèque Scolaire en Norvège est un projet national de quatre années (de 2009 à 2013) conçu pour renforcer la bibliothèque scolaire en tant qu'outil éducatif. L'objectif du programme est que les bibliothèques scolaires soient activement utilisées dans l'éducation à la lecture et la promotion de la culture de l'information. L'Université d'Adger (UiA) a piloté le programme au nom de la Direction Générale norvégienne pour l'Education et la Formation, et a défini plusieurs domaines prioritaires : développement des compétences à travers les formations, les stages et la création d'une base de données rassemblant toutes les idées qui ont émergé de ce programme ainsi que des ressources pédagogiques sur la culture de l'information.

Au moment de l'achèvement du programme en août 2013, 210 projets de développement ont été réalisés sous leurs auspices : 105 municipalités norvégiennes ont participé et tous les comtés ont été représentés. Les cours à l'université en "bibliothéconomie scolaire" ont été développés et deux sites web ont été créés. Le programme a nécessité du chef d'établissement d'être un gestionnaire de projet et le groupe entier participant au projet, à savoir le chef d'établissement, le bibliothécaire scolaire et les professeurs investis, a pu suivre des formations continues. L'implication de la part de la collectivité territoriale, par le biais de la municipalité, a été perçue comme également importante.

Les projets ont mis l'accent sur l'instauration de coopérations, à la fois au sein des établissements

mais aussi avec les bibliothèques publiques et les communautés locales. L'Institut Nordique pour les Etudes en Innovation, Recherche et Education⁴ (INEIRE) a évalué le programme et a publié le rapport en mars de cette année 2021. Ce rapport présente quelques initiatives conduisant à l'établissement du Programme pour la Bibliothèque Scolaire en Norvège et résume la gestion des expériences du programme à la lumière de l'évaluation de l'INEIRE.

Le Programme pour la Bibliothèque Scolaire en Norvège est le résultat de beaucoup de travail, d'un nombre important de revendications et d'initiatives professionnelles, incluant des campagnes menées par des organisations de promotion de lecture et des bibliothèques, des rapports au niveau étatique et une enquête sur les bibliothèques scolaires norvégiennes. Le programme a été conçu pour veiller à ce que la bibliothèque scolaire soit incluse dans l'éducation à la lecture et la promotion de la culture de l'information. L'INEIRE a complété cette évaluation du programme en mars 2014.

L'évaluation de l'INEIRE montre que les établissements participant au programme ont un bilan positif et la mise en œuvre par l'Université d'Agder a été considérée comme une réussite. Les responsables territoriaux, les chefs d'établissements, les professeurs et les bibliothécaires scolaires ont signalé que le suivi par la gestion de projet a été très utile. Au niveau national, il y a toujours des défis. Le nombre total d'établissements norvégiens utilisant la bibliothèque comme un outil pédagogique n'a pas augmenté, et toujours moins de la moitié des écoles primaires emploie un bibliothécaire scolaire qualifié. Le rapport de l'INEIRE suggère comme explications, des stratégies nationales floues, une législation insuffisante et un manque de priorisations politiques. L'INEIRE souligne également le besoin de transmettre les ressources des écoles plus systématiquement à travers une infrastructure établie de communication. L'évaluation du rapport défend que les

collectivités territoriales, à savoir les municipalités, doivent avoir une responsabilité claire pour diffuser l'information à propos de leurs projets.

Les Associations professionnelles de bibliothécaires scolaires en Norvège

L'Association des Bibliothèques Norvégiennes¹ (ABN) est une organisation indépendante, à but non lucratif, qui a pour objectif la promotion et le développement des services d'information et des bibliothèques. L'ABN a été fondée en 1913. Elle se veut être un groupe d'intérêt uni et force de proposition pour le secteur des services d'information et des bibliothèques norvégiennes. Elle doit être le garant de principes et le défenseur des bibliothèques à tous les niveaux gouvernementaux. L'Association Norvégienne des Bibliothèques Scolaires² (Norwegian School Library Association) est un groupe spécial de l'ABN.

Les lois et réglementations des bibliothèques scolaires en Norvège³

- Chaque élève à l'école primaire et au collège doit avoir accès à une bibliothèque scolaire ou à une bibliothèque publique durant la journée ;
- En pratique beaucoup de lycées sont dotés d'une bibliothèque scolaire avec des bibliothécaires qui détiennent une licence ou un master en sciences de l'information et des bibliothèques de l'Université de la Métropole d'Oslo.
- Dans les écoles primaires, un professeur est généralement responsable de la gestion du fonds documentaire dans l'école. La plupart des professeurs qui font fonction dans les écoles primaires possèdent un diplôme en "bibliothéconomie scolaire" de l'Université d'Agder.

³ Loi d'Education de 1998 : <https://www.regjeringen.no/en/dokumenter/education-act/id213315/> ainsi que https://lovdata.no/dokument/NLE/lov/1998-07-17-61#KAPITTEL_10

Formation à une maîtrise de l'information

Les récents travaux de recherche sur le sujet et les nouveaux programmes impliquent, entre autres choses, de mettre l'accent sur l'esprit critique. Notamment, il est énoncé que les élèves « doivent être capables d'évaluer différentes sources de savoir et d'avoir une réflexion critique sur la manière dont le savoir est développé. » Si nous lisons le Kunnskapsløftet 2020⁴, nous constatons un nombre d'éléments de compétences informationnelles dans le tronc commun, dans les thèmes interdisciplinaires, les éléments fondamentaux et les compétences basiques, aussi bien que les objectifs spécifiques des compétences.

Deux exemples issus des programmes de sciences peuvent illustrer cela. Les compétences de lecture de documents en science sont décrites, entre autres choses, comme « explorer, identifier, interpréter et utiliser de l'information de différents types de textes et évaluer de manière critique comment l'information scientifique est produite et utilisée dans les arguments ». Par ailleurs, les compétences numériques dans les sciences impliquent « l'utilisation des outils de recherche, la maîtrise des stratégies de recherche, l'évaluation critique des sources et la sélection d'information pertinente sur les sujets scientifiques » :

- Il y a des connaissances et des compétences en culture numérique et en culture de l'information dans toutes les disciplines du niveau 1 à 13 en Norvège ;
- dans beaucoup de lycées, les bibliothécaires scolaires enseignent l'information documentaire, l'évaluation des sources et de l'information lorsque les élèves travaillent sur des projets particuliers ;
- l'information documentaire peut être enseignée dans toutes les matières, mais il est plus

courant de s'appuyer sur des disciplines comme le Norvégien, les sciences sociales, ou les cours d'anglais.

Les élèves doivent :

- être capables de cerner leur besoin d'information ;
- savoir trouver, localiser l'information ;
- être capables d'évaluer l'information (évaluation des sources) ;
- savoir exploiter l'information dans les apprentissages ;
- savoir citer leurs sources (technique de citations, préparation d'une liste de références) ;
- connaître les principes clés du droit d'auteur (ce qui peut être utilisé dans différents contextes) ;
- être capables de communiquer en ligne (produire et partager du contenu en ligne).

Au lycée St. Olav (2^{nde} à Terminale), nous utilisons :

- l'outil "Citation Compass – Kildekompasset"⁵ qui apprend aux élèves à citer les sources correctement, et également d'avoir un regard critique sur celles-ci ;
- le standard bibliographique "APA 6th"⁶ dans le cadre d'un cours spécial pour enseigner l'évaluation des sources avec les élèves.

Par ailleurs, les élèves doivent savoir ce qu'est le référencement, à savoir la méthode que nous utilisons afin de citer les sources et documents utilisés pour leur travail.

Les deux parties du référencement :

1. les citations dans le texte, qui sont courtes, par des mentions entre parenthèses. Elles doivent être directement écrites après les citations du travail d'autres personnes.

⁴ Nouveaux programmes scolaires de 2020

⁵ <https://kildekompasset.no/en/>

⁶ <https://kildekompasset.no/en/referencing-styles/apa-6th/>

2. les références bibliographiques, dans une liste classée par ordre alphabétique à la fin du devoir, où sont indiquées toutes les sources d'information citées dans le texte.

Les élèves doivent mesurer l'importance du référencement et ses objectifs tels que :

- donner au lecteur la possibilité de retrouver les sources d'information citées et de voir eux mêmes si elles sont citées correctement ;
- éviter le plagiat.

Les dernières recommandations de l'IFLA pour les bibliothèques scolaires, révisées sous l'égide de Barbara Schultz-Jones et Dianne Oberg, ont été traduites en Norvégien en 2015. Certaines de ces recommandations sont pertinentes pour

les bibliothèques scolaires norvégiennes : *"La bibliothèque scolaire est une composante essentielle des enseignements et des apprentissages à l'école ; elle contribue également aux objectifs de l'école en matière de sociabilité, avec l'engagement des élèves, la prise en compte de la diversité des profils d'apprenants, le développement des relations avec une communauté élargie. Les objectifs de la bibliothèque scolaire devraient être en conformité avec les objectifs de l'école tels que l'apprentissage de la lecture, l'apprentissage basé sur les programmes d'enseignement et la formation du citoyen. La contribution de la bibliothèque scolaire à la réalisation des objectifs de l'école est subordonnée aux ressources et aux moyens humains qui lui sont alloués."*⁷

⁷ Point 5.2 du "Programme et activités" disponible sur : <https://www.skolebibliotek.no/wp-content/uploads/2020/03/ifla-school-library-guidelines-nr-1.pdf>
(Traduction en norvégien des Recommandations IFLA)

Références bibliographiques

NORWEGIAN LIBRARY ASSOCIATION, N. L. (2021). Hentet fra <https://norskbi-bliotekforening.no/in-english/>

COMMITTEE, F. P. (2015, 06). *IFLA school library guidelines*. Hentet fra <https://www.ifla.org/wp-content/uploads/2019/05/assets/school-libraries-resource-centers/publications/ifla-school-library-guidelines.pdf>

DAMGAARD, N. (2009). *How can we raise the level of competence among school librarians? 12 years of experience with net-based courses in Norway* . Hentet fra <https://journals.library.ualberta.ca/slw/index.php/iasl/article/view/7684/4548>

LOVDATA. (1998, 07 17). Hentet fra Act relating to Primary and Secondary Education and Training (the Education Act): https://lovdata.no/dokument/NLE/lov/1998-07-17-61#KAPITTEL_10

NIFU. (2014, 06). *Evaluation of program for school library*. Hentet fra <https://www.nifu.no/projects/evaluering-av-program-for-skolebibliotek/>

OECD.(2016).*ducationataGlance2016:OECDIndicators*.Hentetfrahttps://read.oecd-ilibrary.org/education/education-at-a-glance-2016/norway_eag-2016-73-en#page1

TALLAKSEN, E. R. (2008). *Opplevelse, oppdagelse, opplysning : fagbok om skolebibliotek* /. Oslo: Biblioteksentralen.

UDIR. (2020, 10 17). *The Norwegian Directorate for Education*. Hentet fra <https://www.udir.no/in-english/>

Moteurs

de recherche :

des algorithmes sans
contrôle en quête
de compréhension ?

Dossier :
Enseigner
les
algorithmes

◆ Eric Bruillard

Université de Paris, Education, Discours, Apprentissages (EDA),
eric.bruillard@u-paris.fr

Des algorithmes dans notre vie quotidienne

Le mot « algorithme » fait figure d'épouvantail depuis quelques années, sorte de monstre malveillant tirant les ficelles derrière les applications offertes sur les smartphones. Pourtant, la notion d'algorithme est très ancienne (voir par exemple, « Histoire d'algorithmes »¹) et correspond simplement à une suite finie et non ambiguë d'instructions et d'opérations permettant de résoudre une classe de problèmes².

Elle est d'abord associée aux mathématiques et le mot lui-même viendrait d'un mathématicien persan du IX^e siècle nommé Al-Khwārizmī³. Les élèves de maternelle y sont familiarisés. Ainsi, ils construisent des objets selon une règle, par exemple en élaborant un collier avec des billes de différentes couleurs et des successions à

respecter, ce qui les prépare notamment à la notion de numération. Les techniques opératoires qu'ils apprennent à l'école primaire, comme l'addition et la soustraction, sont des algorithmes qui transforment les nombres dans leur écriture décimale. Notons également un algorithme *littéraire*, celui que Georges Perec a écrit pour la revue *L'Enseignement programmé*, paru en décembre 1968 : « L'art et la manière d'aborder son chef de service pour lui demander une augmentation », adapté ensuite pour le théâtre sous le titre *L'augmentation*. Cet algorithme est brillamment développé, décrivant les différentes actions à mener pour obtenir l'augmentation convoitée⁴.

Un peu plus tard, au moment de l'enseignement de la programmation, divers auteurs cherchaient à montrer que, comme Monsieur Jourdain faisait de la prose sans le savoir, nous suivons tous des algorithmes sans en avoir conscience : en effet

¹ Chabert Jean-Luc, Barbin Evelyne, Guillemot Michel, Djebbar Ahmed, ... (2010). Histoire d'algorithmes. Du caillou à la puce. Belin.

² <https://fr.wikipedia.org/wiki/Algorithme>

³ <https://interstices.info/famille-algorithmes-programmation/>

⁴ <http://lecture.cafeduweb.com/lire/13001-art-maniere-aborder-son-chef-service-pour-lui-demander-augmentation---georges-perec.html>

nous effectuons des séquences d'actions avec quelques formes de contrôles avec « tant que » ou « jusqu'à ce que » pour aboutir à un résultat final. Les recettes de cuisine constituent depuis fort longtemps un exemple très commenté⁵. Il y a d'un côté les ingrédients qu'il va falloir utiliser et de l'autre une description de la suite d'actions à réaliser, avec différents tests et contrôles : attendre l'ébullition de l'eau, que le beurre soit fondu mais pas noirci, que la pâte soit onctueuse... Les descriptions des recettes montrent d'ailleurs certaines limites : comment juger que le mélange est bien réalisé, que la poêle est assez chaude, qu'il y a suffisamment de sel, etc. Certains ajoutent des « trucs », d'autres les gardent pour eux. Les recettes sont plus ou moins complexes, des tours de main peuvent s'avérer nécessaires, et le plus souvent le résultat n'est pas garanti, même si on suit scrupuleusement l'algorithme décrit dans la recette.

L'intelligence artificielle est maintenant convoquée, avec des machines pouvant exécuter des recettes de grands chefs cuisiniers : le *Gastronomy Flagship Project* de Sony⁶ est dédié à la gastronomie, une intelligence artificielle pour la création des recettes, un robot pour les réaliser et un réseau social de cuisiniers pour une création communautaire⁷ ; *Moley*, la première cuisine entièrement robotisée au monde⁸ ou le *Bot Chef* de Samsung⁹ : « Non seulement le robot cuisine des repas complets, mais il vous indique quand les ingrédients doivent être remplacés, suggère des plats en fonction des articles que vous avez en stock, apprend ce que vous aimez et nettoie même les surfaces après lui-même ».

En fait, les techniques d'intelligence artificielle peuvent donner aux appareils une certaine capacité d'agir par eux-mêmes et, au lieu de toucher un

écran ou manipuler des boutons, nous parlerons bientôt à nos appareils, constamment à l'écoute, dans nos maisons et nos bureaux. La promesse ultime est dans leur capacité à prédire ce que nous voulons avant même qu'on le demande. L'inquiétude que l'on peut ressentir ne vient pas d'activités menées par des humains suivant des algorithmes, mais naît lorsque ces derniers nourrissent des systèmes ou des machines automatiques qui prennent des décisions ou orientent nos choix et nos visions du monde. Nous allons tenter d'en montrer les enjeux autour des questions de recherche d'information.

Les moteurs de recherche : des algorithmes au pouvoir grandissant

Au démarrage de l'Internet grand public, avec le développement du *Web*, il a fallu inventer des techniques pour aider les humains à trouver les pages qui pouvaient les intéresser.

Pour cela, il fallait indexer un grand nombre de pages *web* pour pouvoir les retrouver et déterminer un lien entre la demande d'un utilisateur et ces pages. Les techniques documentaires classiques, à partir des descripteurs des pages, s'avéraient peu performantes, notamment parce que l'indexation des pages, assurée par les auteurs eux-mêmes, ne suivait pas les règles classiquement recommandées en documentation. En outre, le fait que le *Web* est un hypertexte, c'est-à-dire que les pages sont liées entre elles, n'était pas pris en compte. Cela a été la grande force de *Google*, d'une part de constituer des parcs de machines pour indexer les pages *web* et, d'autre part d'élaborer un algorithme reposant sur le contenu des pages et les liens entre elles¹⁰. En effet, les

⁵ Voir par exemple <https://interstices.info/les-ingredients-des-algorithmes/#1> et <https://interstices.info/les-ingredients-des-algorithmes/#2> ou interroger un moteur de recherche avec l'expression « algorithme recette de cuisine »

⁶ <https://ai.sony/projects/>

⁷ <https://www.futura-sciences.com/tech/actualites/intelligence-artificielle-ia-sony-veut-defier-top-chef-84723/>

⁸ <https://moley.com/>

⁹ <https://www.cnetfrance.fr/news/ifa-2019-on-a-cuisine-avec-le-robot-chef-de-samsung-39890207.htm>

¹⁰ Sur Page Rank, l'algorithme de tri utilisé par Google, voir Delahaye (2007) ou Guerraoui (2014)

moteurs doivent palier un défaut structurel du Web : les liens hypertextes ne sont pas bidirectionnels et si vous consultez une page, vous ne pouvez pas savoir quels liens pointent sur cette page. Les moteurs de recherche ont récupéré ces informations et ils s'en servent pour déterminer l'intérêt d'une page. Ainsi, plus il y a de liens pointant sur une page, plus il estime qu'une page est *intéressante* (surtout si ces liens proviennent de pages également jugées *intéressantes* par le moteur !).

Notons que cette caractéristique a orienté des techniques courantes pour améliorer le référencement d'une page et a également été détournée via ce que l'on nomme le *Google bombing* ou le *bombardement Google*. Cela consiste à multiplier les liens vers une page ou un site avec la même expression dans l'origine du lien. Cette expression, souvent négative voire dénonciatrice ou diffamatoire, devient un mot clé pour conduire au site (voir quelques exemples de détournements¹¹).

En outre, contrairement aux algorithmes habituels en documentation qui fournissent de manière exhaustive les réponses à une demande, le très grand nombre de pages web susceptibles de répondre à une requête¹² conduit à ordonner les réponses à afficher. C'est ce tri, le choix effectué automatiquement par la machine et l'ordre de présentation, qui va s'avérer déterminant.

La question qui s'ensuit est de trouver les meilleurs algorithmes, c'est-à-dire ceux qui satisfont le mieux les utilisateurs. Pour cela, si les algorithmes des moteurs de recherche disposent d'autres informations et pas uniquement du ou des quelques mots écrits dans la barre de recherche, ils seront certainement mieux armés. La notion de « profil » va s'imposer : si on sait que quelqu'un qui vous ressemble a consulté une

page pour répondre à une requête identique à la votre, on vous présentera la même page. Bien évidemment, plus on connaît l'utilisateur, son pays, sa langue, ses opinions, ses goûts, ses préférences, etc., plus on peut avoir de chances de le satisfaire. Mais que veut dire « satisfaire » : pour un usager, qu'il soit content afin de le faire rester puis revenir ? Pour un annonceur, pousser les internautes vers son site ? Ne s'agit-il pas alors, dans les choix opérés, d'exploiter des biais cognitifs : donner à l'utilisateur ce qui confirme ce qu'il croit déjà, par exemple ce que d'autres qui lui ressemblent trouvent plutôt intéressant. Et au lieu d'attendre qu'il formule une requête, pourquoi ne pas directement lui proposer ce qui est susceptible de l'intéresser. On passe alors de la recherche d'informations aux systèmes de recommandations, qui « poussent » l'information vers les utilisateurs. On arrive aux bulles de filtres : ce qui est rendu visible à chacun dépend des différentes données collectées sur lui. Il est installé dans une « bulle » unique, construite à la fois par les algorithmes et par ses propres choix (« amis » sur les réseaux sociaux, sources d'informations, etc.), optimisée pour sa personnalité supposée¹³ (voir aussi Doctorow, 2007, pour une vision dystopique).

Une autre évolution des moteurs de recherche est leur transformation en moteurs dits de réponse : les moteurs de recherche ne répondaient pas directement aux questions posées par les utilisateurs mais proposaient des documents, à charge pour les demandeurs de les lire et d'en extraire les informations recherchées. Il s'agit maintenant de leur fournir le plus directement possible des éléments en lien avec leur question, voire directement une réponse¹⁴. Dans les travaux de recherche actuels, il s'agit d'aller encore plus loin : des données à l'information, aux connaissances¹⁵

¹¹ Voir https://fr.wikipedia.org/wiki/Bombardement_Google ou <https://optimiz.me/google-bombing/> ou <https://smartkeyword.io/seo-netlinking-google-bombing/>.

¹² Parfois, la requête est un simple mot et toutes les pages contenant ce mot sont supposées répondre à cette requête (recherche plein texte)

¹³ Selon Eli Pariser, voir https://fr.wikipedia.org/wiki/Bulle_de_filtres

¹⁴ https://fr.wikipedia.org/wiki/Moteur_de_réponse

¹⁵ Au sens donné par les informaticiens à la notion de connaissance : un ensemble structuré d'informations sur lequel on peut effectuer des traitements

(Abiteboul, 2019), ce que *Google* offre déjà dans des encarts à droite de la page de résultats.

Détourner pour mieux comprendre les fonctionnements des algorithmes

Pour résumer, côté documentation, on est passé :

- des bases de données de documents : un corpus bien délimité avec une liste de champs (auteur, titre, éditeur, mots clés, résumé...), des langages documentaires et une réponse exhaustive et sûre à une requête (aux erreurs de saisie près) ;
- au *Web* : un corpus mal délimité (bien que d'une taille considérable, seule une partie du *Web* est indexée), le contenu des pages (chaînes de caractères), des liens entre les pages, une réponse non exhaustive, un algorithme de tri des réponses.

Avec le *Web*, on dispose de trois manières de trouver de l'information : (1) utiliser un moteur de recherche ou un outil de recherche, il faut alors décrire ce que l'on cherche, en tapant une requête ou en fournissant un fragment (image, son, texte) ; (2) naviguer en cliquant sur des liens proposés sur la page *web* que l'on consulte, mais on dépend de ce qui est présent sur cette page ; (3) consulter ce que notre réseau humain nous envoie ou ce que le système, selon ses paramètres, nous envoie automatiquement.

Dans tous les cas, un processus a permis de filtrer l'information et de plus ou moins la personnaliser, en fonction de ce que le système de recherche ou de préconisation peut connaître de nos souhaits et de nos besoins. Mais pour satisfaire un humain, n'est-il pas finalement plus facile de le rendre prédictif, de le façonner petit à petit de façon à ce qu'il en arrive à souhaiter ce que l'on prévoit pour lui ? Heureusement, ce n'est pas aussi simple que cela.

Les professeurs documentalistes ont un rôle important à jouer, pour aider les élèves à mieux appréhender les différents instruments de recherche à leur disposition et à exercer leur jugement sur leur fonctionnement et sur ce qu'ils proposent.

D'abord, ne pas oublier que la « pertinence » est avant tout un jugement sur l'adéquation entre des besoins de connaissance¹⁶ et des documents qui « contiennent » des informations censées répondre à ce besoin (une capacité que les machines sont encore loin d'avoir) et qu'il y a différentes voies pour les trouver : taper une requête, choisir *via* une image, cliquer sur un lien visible, etc. Ainsi, choisir une page web à partir d'une image est une technique qui peut être performante : avec *Google images*, on peut survoler rapidement de très nombreux exemples et les images elles-mêmes reflètent souvent un niveau de technicité. Par exemple, avec la requête « cycle de l'eau », les images proposées, leur complexité et le lexique visible, donnent des indices sur le niveau de classe auquel elles peuvent correspondre (plus facile à choisir qu'à partir de la lecture des brefs extraits fournis). En conséquence, il faut développer des techniques d'enquêtes et faire preuve de réflexivité : utiliser plusieurs méthodes, comparer, etc. Même si on ne peut pas répéter trop souvent ces activités qui prennent du temps, il est important de saisir des occasions afin de les mener et de conduire quelques séances « marquantes », au cours desquelles des interrogations et des situations inhabituelles devraient surprendre les élèves et les amener à considérer les instruments de recherche différemment¹⁷.

Il est certes difficile d'enseigner le fonctionnement des machines informatiques qui nous environnent, mais on peut se donner l'objectif d'essayer de mieux le comprendre, et pour comprendre, il est important de détourner (Bruillard, 2020). Ainsi, affiner le jugement de pertinence est un objectif important, mais il est plus facile de détecter des

¹⁶ Les modèles, comme celui de Mizzaro (1997) utilisent l'expression « besoin d'informations », mais, dans le contexte scolaire, la finalité est celle de l'acquisition de connaissances par les élèves.

¹⁷ Penser aux générations d'élèves qui apprennent à se servir du compas en exécutant des rosaces !

choses non pertinentes que des choses pertinentes. Il faut apprendre à rejeter, puis apprendre à classer des documents meilleurs que d'autres relativement à une requête dans un contexte particulier. Voici quelques exemples d'activités à mener :

- aller chercher des choses inattendues : ne pas prendre uniquement les recherches dans une vision de service mais en conduire avec des demandes inhabituelles :

- faire des jeux (voir par exemple Simonnot, 2008), essayer de faire taire le moteur (taper plusieurs mots qui ensemble ne donnent aucun résultat),

- rechercher un poème *via* une image, un avocat marron qui ne soit pas un fruit *via Google images*, etc.

- refuser la similarité : à l'instar des banques qui vous demandent un profil (prudent, risqué...) et gèrent selon ce profil, pourquoi n'y aurait-il pas un profil de *chercheur* (curieux, aventureux, scolaire...) ? Pourquoi ne pas demander des choses éloignées de ce que l'on connaît ? Quelles poésies recommander à un adolescent amateur de science-fiction ? Se méfier des *nudges* (Boissière et Bruillard, 2021)¹⁸.

Que la formation aux instruments de recherche offre de multiples perspectives, qu'elle soit le lieu d'ouvertures et de rencontres inattendues, au-delà de la satisfaction de critères simplistes de performance, n'est-ce pas ce que l'on pourrait souhaiter ?

¹⁸ Développée par l'économiste comportemental Richard Thaler et le juriste Cass Sunstein, la théorie du nudge désigne une méthode d'influence cherchant à modifier des comportements humains, notamment des choix à faire, sans contrainte, ni obligation, ni sanction.
[https://fr.wikipedia.org/wiki/Nudge_\(livre\)](https://fr.wikipedia.org/wiki/Nudge_(livre))

Références bibliographiques

ABITEBOUL SERGE (2019). Des données à l'information, aux connaissances. *Youtube* [en ligne]. 13 décembre 2019. [Consulté le 20 octobre 2021]. Disponible à l'adresse : <https://www.youtube.com/watch?v=9Sk42Fy6IMo>

BOISSIÈRE JOËL ET BRUILLARD ÉRIC (2021). *L'école digitale. Une éducation à vivre et à construire*. Collection Sociologie. Armand Colin, 368 p.

BRUILLARD ÉRIC (2020). L'écriture inclusive ouvre des liens surprenants. Réflexions en didactique de l'informatique. *STICEF, Volume 27, n°1*, 2020. Disponible à l'adresse : <http://sticef.univ-lemans.fr/num/vol2020/27.1.4.bruillard/27.1.4.bruillard.html>

CHABERT JEAN-LUC, BARBIN EVELYNE, GUILLEMOT MICHEL, DJEBBAR AHMED ET AL. (2010). *Histoire d'algorithmes. Du caillou à la puce*. Belin.

DELAHAYE JEAN-PAUL (2007). Un moteur de recherche, pour le meilleur et pour le pire. *Interstices, INRIA*. [en ligne]. Disponible à l'adresse : <https://interstices.info/un-moteur-de-recherche-pour-le-meilleur-et-pour-le-pire/>

DOCTOROW CORY (2007). *enGooglés*, Traduction de Scroogled (Septembre 2007 ? Magazine Radar) par Valérie Peugeot, Hervé Le Crosnier et Nicolas Taffin pour C & F éditions. <https://cfeditions.com/scroogled/scroogled.html>

GUERRAOUI RACHID (2014). Un algorithme : PageRank de Google. *Blog Binaire, Le Monde*. [en ligne]. Disponible à l'adresse : <https://www.lemonde.fr/blog/binaire/2014/12/01/un-algorithme-pagerank-de-google/>

MIZZARO STEFANO (1997). Relevance: the whole history. *In Journal of the American Society for Information Science*, vol. 48, n°9, p. 810-832.

SIMONNOT BRIGITTE (2008). Quand les moteurs de recherche appellent au jeu : usages ou détournements ? *Questions de communication*, 14 | 2008, p. 95-114. [en ligne]. Disponible à l'adresse : <http://journals.openedition.org/questionsdecommunication/752>

Didactisation

de l'indexation automatique

à travers les encarts informationnels (Infoboxes)

◆ Élodie Renvoisé

Professeure Documentaliste - Lycée des Métiers Claude Chappe - Rue des Collèges, Arnage
elodie.renvoise@ac-nantes.fr

L'indexation automatique est un procédé documentaire qui a évolué avec les nouvelles technologies du *Web* utilisées par les moteurs de recherche, telles que le *machine learning* et le *deep learning*. Parmi ces nouveautés utilisant les algorithmes d'intelligence artificielle, les encarts informationnels, ou infoboxes, proposent aux internautes des condensés d'information (e.g. *Knowledge Graph*). Notre recherche montre que les professeur.e.s documentalistes enseignent pour une grande majorité l'indexation automatique. Toutefois, même s'il.elle.s ont quasiment tou.te.s observé ces infoboxes, il.elle.s les étudient peu avec les élèves. Notre étude propose une progression originale de son enseignement à travers ces *infoboxes* pour tout le secondaire.

Mots-Clés : Indexation automatique, encart informationnel, didactique, moteur de recherche, algorithme, intelligence artificielle.

Les moteurs de recherche et les algorithmes ont beaucoup évolué depuis leur création au début des années 1990. Aujourd'hui, des algorithmes tentant de mimer le cerveau humain y sont intégrés afin de faciliter l'accès à l'information en ligne.

Au cœur des moteurs de recherche, l'indexation automatique permet d'analyser le contenu des documents disponibles sur le *Web* afin que les moteurs d'indexation puissent retrouver l'information la plus pertinente par rapport à la recherche effectuée par les utilisateurs. À partir du moment où les moteurs de recherche se sont emparés de la technologie de l'indexation automatique dans les années 1990, elle n'a cessé d'être améliorée jusqu'au *Web* sémantique actuel.

À la fin des années 1990, le moteur de recherche *Google* est créé avec son algorithme principal (*PageRank*) basé sur la popularité. À cette époque, les écoles entrent progressivement dans la culture numérique en étant dotées d'ordinateurs connectés à Internet (Rousseau, 2015). À la fin des années 2000, les professeur.e.s documentalistes commencent à étudier le fonctionnement des moteurs de recherche avec les élèves, dont le processus d'indexation automatique. La notion de moteur de recherche devient une entrée dans l'enseignement de l'information documentation (et principalement celui de *Google* ; Rabat, 2008 ; Sogliuzzo, 2013). C'est en 2013 que la notion de « moteur de recherche » entre dans le Wikinotions¹

¹ Wikinotions : plateforme collaborative de notions info-documentaires développée par l'association des professeurs documentalistes de l'éducation nationale (APDEN)

(voir historique des versions : Reynaud, s. d.). L'indexation automatique est alors enseignée principalement à travers l'étude de la page de résultats (Sogliuzzo, 2013).

Toutefois, depuis plusieurs années, nous voyons apparaître sur la page de résultats des encarts informationnels, ou infoboxes, avec souvent, des références aux pages de l'encyclopédie en ligne Wikipédia, des cartes géographiques ou, plus récemment, des définitions. Leur relative hégémonie en haut de la page de résultats montre que les procédés d'indexation des pages *Web* se sont perfectionnés depuis la reconnaissance de caractères de Hans Peter Luhn dans les années 1950 (El Hachani, 1997). C'est pourquoi, nous interrogeons l'impact de l'enseignement de l'indexation automatique (à travers les cartouches informationnelles) sur l'acquisition d'une culture de l'information numérique chez les élèves. Est-ce un objet pertinent pour comprendre le fonctionnement des moteurs de recherche, et notamment l'indexation automatique ? Les professeur.e.s documentalistes se sont-ils.elles réellement emparé.e.s de la question de l'enseignement de l'indexation automatique, et si oui, le font-ils.elles à partir des cartouches ? Est-il nécessaire de repenser l'enseignement et la didactisation de l'indexation automatique avec les dernières avancées du *Web* ?

Qu'est-ce que l'indexation automatique ?

La notion d'indexation, et son ancêtre l'index, sont des procédés documentaires anciens qui visent à faciliter l'accès à une information d'un ou plusieurs documents (Kovacs & Dibie, 2006 ; Cacaly et al., 2004). L'indexation automatique est, comme l'indexation, un procédé facilitateur de l'accès au contenu d'un document (Boulogne, 2004). Elle est à différencier de l'indexation manuelle, ou humaine, par l'automatisation du processus. Aujourd'hui, la numérisation des

documents a entraîné l'apparition de méta-données² supplémentaires aux composantes classiques des documents analogiques (Titres, auteurs, année de publication, etc., Cacaly et al., 2004 ; Broudoux 2012). L'indexation automatique est donc un processus d'analyse du contenu d'un document informatisé, pour lequel, l'accès au contenu du document est réalisé par la création et l'utilisation d'un index. Cet index est issu de données et de métadonnées, et l'efficacité de la recherche est améliorée par des algorithmes toujours plus perfectionnés facilitant l'accès au sens des informations contenues dans les documents (*i.e. Web sémantique*).

Quatre étapes régissent le fonctionnement de l'indexation automatique au sein d'un moteur de recherche, depuis la segmentation du document en unités sémantiques, jusqu'à la structuration de l'information, où nous retrouvons le rôle de l'index, en passant par la normalisation des mots et l'analyse syntaxique. Cette dernière se complexifie aujourd'hui à l'heure des algorithmes qui favorisent l'apprentissage des ordinateurs par eux-mêmes (*machine learning*) (Figure 1 ; Delestre & Malandin, 2017).

Un cas concret : les « infoboxes » et autres cartouches informationnels

Depuis plusieurs années, la page de résultats de certains moteurs de recherche propose comme résultats d'indexation automatique des encarts informationnels sous forme textuelle, iconographique et/ou vidéo. Aujourd'hui, ces encarts informationnels, ou *infoboxes*, deviennent des encarts définitionnels dont l'origine de l'information pose parfois question (Renvoisé, 2020).

Le terme « *infobox* » a d'abord été associé à la structuration des formulaires Wikipédia. Une *infobox* Wikipédia est un cartouche, ou un encadré, placé en général en haut à droite de la page

² Métadonnées : données qui caractérisent une autre donnée. Les métadonnées peuvent être manuelles ou générées automatiquement par les algorithmes de « lecture » des documents (Delestre & Malandin, 2017)

FIGURES

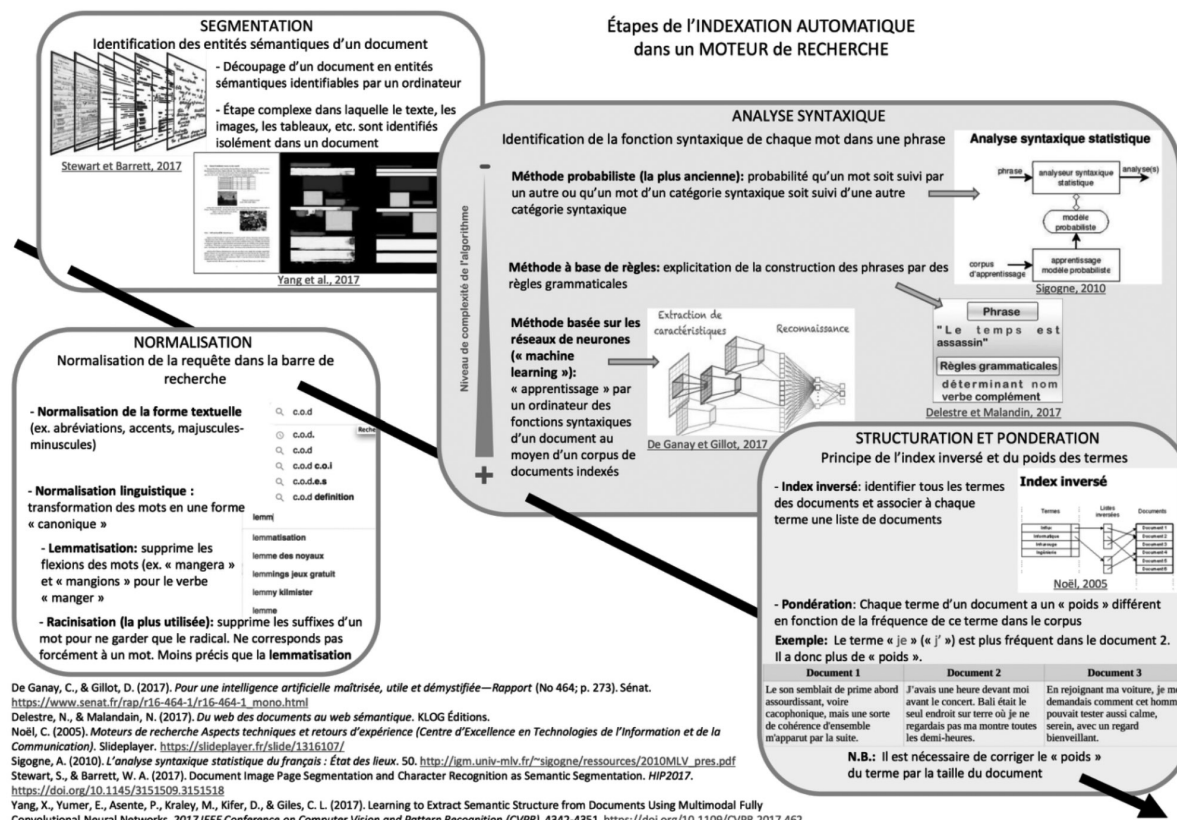


Figure 1 : Étapes de l'indexation automatique dans un moteur de recherche.

Wikipédia présentant une synthèse d'informations importantes sur un sujet. Elle complète un article de la page Wikipédia en fournissant des informations générales sur le sujet (Aide : *Infobox in Wikipédia*, 2019). L'origine de l'information est issue d'une indexation manuelle, c'est-à-dire que l'auteur.trice renseigne les informations importantes dans un gabarit pré-formaté de son choix.

Lorsque l'*infobox*, contenant des informations issues de la page Wikipédia, se situe sur la page de résultats, des algorithmes supplémentaires de ceux de l'indexation automatique (Figure 1) vont aussi pointer et indexer d'autres pages du

Web à partir d'une source centralisante et ajouter des informations complémentaires en-dessous des informations Wikipédia. Nous sommes alors en présence d'« *infoboxes d'ontologie* », aussi appelés « *Knowledge graph* » par Google, qui sont issues des données liées du Web de données³ (Sire, 2018). Ces nouveaux algorithmes d'intelligence artificielle utilisés par les moteurs de recherche, par exemple Google, Bing et Qwant, ont pour objectifs de proposer une réponse à une requête, souvent une question, la plus complète possible en se basant sur le Web sémantique (*Knowledge panel help de Google*). Pour ce faire, les algorithmes des moteurs de recherche

³ Web de données (ou *Web of linked data*) : Données liées du Web basées sur le déréférencement (accès à toutes les données) et l'interopérabilité des données (lisibles par toutes les interfaces)(Delestre&Malandain, 2017)

utilisent des modèles en forme de graphe où les entrées de l'ontologie forment le graphe et le moteur de recherche tente de faire correspondre sémantiquement le graphe requête avec le graphe résultats (Delestre et Malandin, 2017). Autrement dit, l'ontologie, sur laquelle se base le *Knowledge graph*, permet de coder du lien sémantique entre des objets, des mots, des images, afin d'en faire du sens, comme en biologie entre les protéines et les gènes qui les codent (Renvoisé, 2020), ou comme la grammaire pour le langage (Wikipédia, 2019). Ces mêmes relations sémantiques, ou

graphe, sont utilisées pour faire correspondre la réponse à la requête initiale. Le résultat de ce processus est situé dans un cartouche en haut à droite de la page de résultats (Figure 2).

Aujourd'hui nous voyons apparaître un troisième type d'*infobox* qui s'apparente à un dictionnaire contenant principalement des définitions. Les deux seuls moteurs de recherche, à notre connaissance, à utiliser ce type d'*infobox* définitionnelle sont *Bing* et *Google*. Entre les deux moteurs de recherche, les définitions et leur structuration diffèrent comme dans les dictionnaires imprimés.

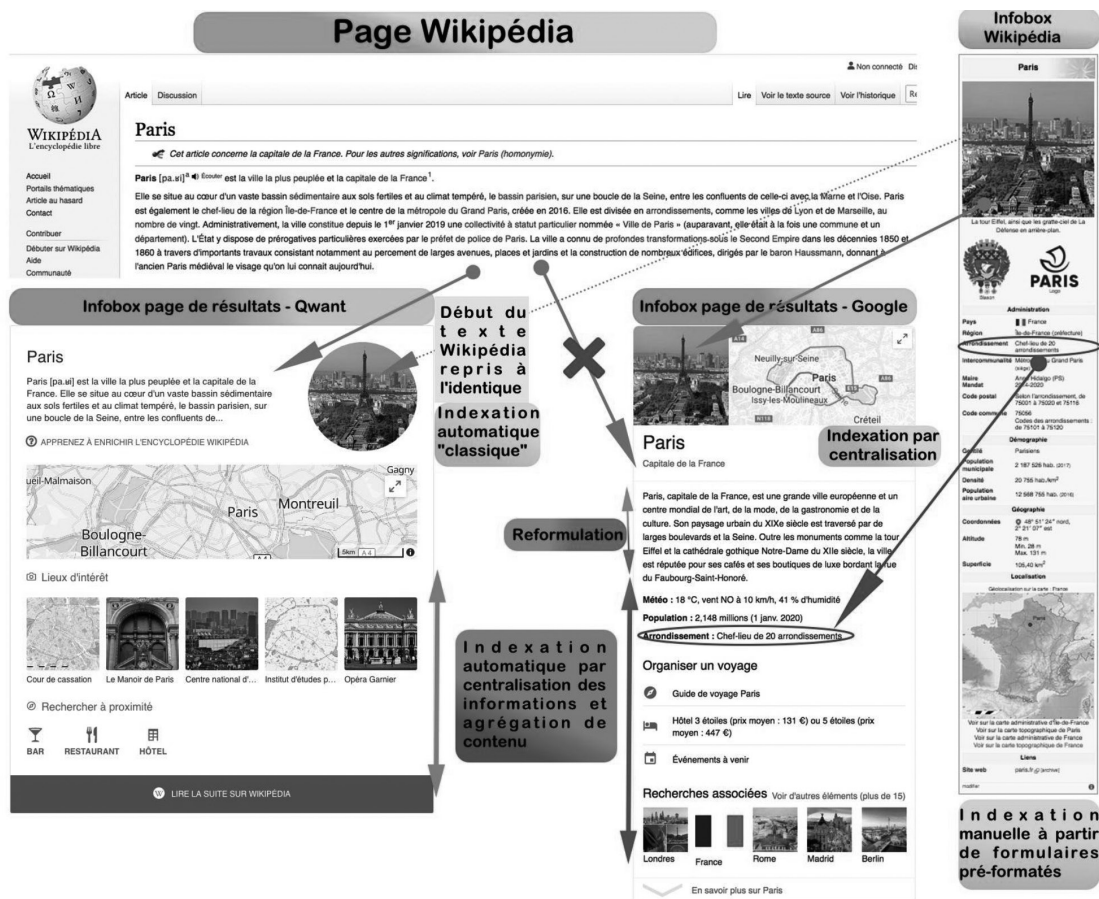


Figure 2 : Exemples de deux types d'infoboxes : infobox Wikipédia et infobox d'ontologie de la page de résultats à partir de la requête « Paris » dans les moteurs de recherche *Qwant* et *Google*. Alors que l'*infobox* Wikipédia est issue d'une indexation manuelle (ou humaine) à partir de formulaires pré-formatés, les infoboxes des pages de résultats de *Qwant* et *Google* sont issues de l'indexation automatique. Il apparaît clairement que les choix éditoriaux ne sont pas les mêmes entre ces deux moteurs de recherche. Le moteur de recherche *Qwant* a fait le choix d'indexer en « plein texte » les premières lignes de la page Wikipédia, alors que le moteur de recherche *Google* a opté pour une « reformulation » issue de différents sites pour réaliser l'introduction de l'*infobox*. Ceci explique pourquoi il n'y a pas de source dans cette partie pour le moteur de recherche *Google*. Le reste des informations pour les deux *infoboxes* de page de résultats pour les moteurs de recherche *Qwant* et *Google* est issu de l'indexation automatique incluant la centralisation des informations et l'agrégation de contenu.

Proposition de progression pédagogique de l'enseignement de l'indexation automatique à partir des *infoboxes*

Un des résultats de l'étude⁴ que nous avons menée auprès d'une cinquantaine de professeur.e.s documentaliste (Renvoisé, 2020) montre que la majorité d'entre eux.elles enseignent l'indexation automatique au collège (en 6^e) et au lycée (en SNT en 2nde). Cependant, peu de professeur.e.s documentalistes se sont emparé.e.s de l'objet d'étude des *infoboxes* dans leurs enseignements (27%). Par conséquent, nous proposons une séquence spiralaire⁵ afin d'encourager d'autres collègues à enseigner l'indexation automatique à travers cet objet d'étude, pour tous les niveaux du secondaire. De plus, une réactualisation de la notion d'indexation automatique est proposée dans la nouvelle version du Wikinotions (APDEN, à paraître) et dans le mémoire de Master 2 d'Élodie Renvoisé (Renvoisé, 2020).

Par exemple, à partir de la requête « Paris » dans deux moteurs de recherche, *Qwant* et *Google*, il est possible d'identifier les deux types d'indexation manuelle/humaine et automatique entre l'*infobox* Wikipédia et les *infoboxes* des *Knowledge graph* (Figure 2). Dans le cas des *infoboxes* Wikipédia, il est possible de déterminer que l'indexation est manuelle/humaine car il existe une page expliquant la réalisation de ces *infoboxes* par complétion d'un formulaire pré-formaté au choix. En ce qui concerne les « *infoboxes* d'ontologie » de *Qwant* et de *Google*, la comparaison des premières lignes introductives permet de déterminer l'origine de l'information (Figure 2). Depuis la page de résultats de *Google*, il est possible de déduire que l'indexation automatique du texte sous les images et la carte résulte d'une « reformulation ». Le texte est différent de celui de la page Wikipédia et ne provient d'aucun autre site sourcé. Par conséquent, le processus d'analyse syntaxique basé sur les réseaux de neurones de l'indexation automatique a dû être utilisé

ici (Figure 1). Plusieurs niveaux de formulation peuvent alors être utilisés pour aborder l'indexation automatique dans les *infoboxes* : un niveau débutant de « réécriture » de l'information, un niveau intermédiaire de « reformulation », jusqu'à un niveau avancé de segmentation et de syntaxe de l'information indexée (Figure 3).

Pour comprendre le processus de centralisation des données en pointant vers d'autres pages web (du type *Knowledge graph*), il est possible d'utiliser l'*infobox* de la page de résultats de *Google* qui reprend une information de l'*infobox* Wikipédia (Figure 2). En effet, le fait que seule une partie de l'*infobox* Wikipédia soit reprise dans celle de la page de résultats de *Google* montre que l'indexation automatique, non seulement est capable d'accéder au contenu d'un document, mais est capable d'en extraire une seule partie à retenir.

En ce qui concerne la carte géographique, elle résulte d'un processus de pointage vers une page en particulier qui répertorie des données géographiques. Le moteur de recherche *Google* « pointe » vers la page *Google maps*, alors que le moteur de recherche *Qwant* « pointe » vers *Qwant maps*.

Conclusion

La didactisation des processus d'indexation automatique à travers les encarts informationnels, ou *infoboxes*, est une entrée originale pour permettre aux élèves de comprendre l'accès à l'information et son origine sur le *Web*, elle est cependant peu pratiquée par les professeur.e.s documentalistes. Il est vrai que traditionnellement, l'indexation automatique est enseignée à travers l'étude de la page de résultats pour comprendre le fonctionnement des moteurs de recherche. Or, l'enseignement des *infoboxes* offre un grand potentiel didactique car, non seulement il permet d'aborder la notion d'indexation automatique à travers le fonctionnement des moteurs de recherche, mais aussi celle de l'évaluation de l'information (crédibilité, source), ou encore

⁴ Le questionnaire distribué aux professeur.e.s documentalistes est situé en Annexe

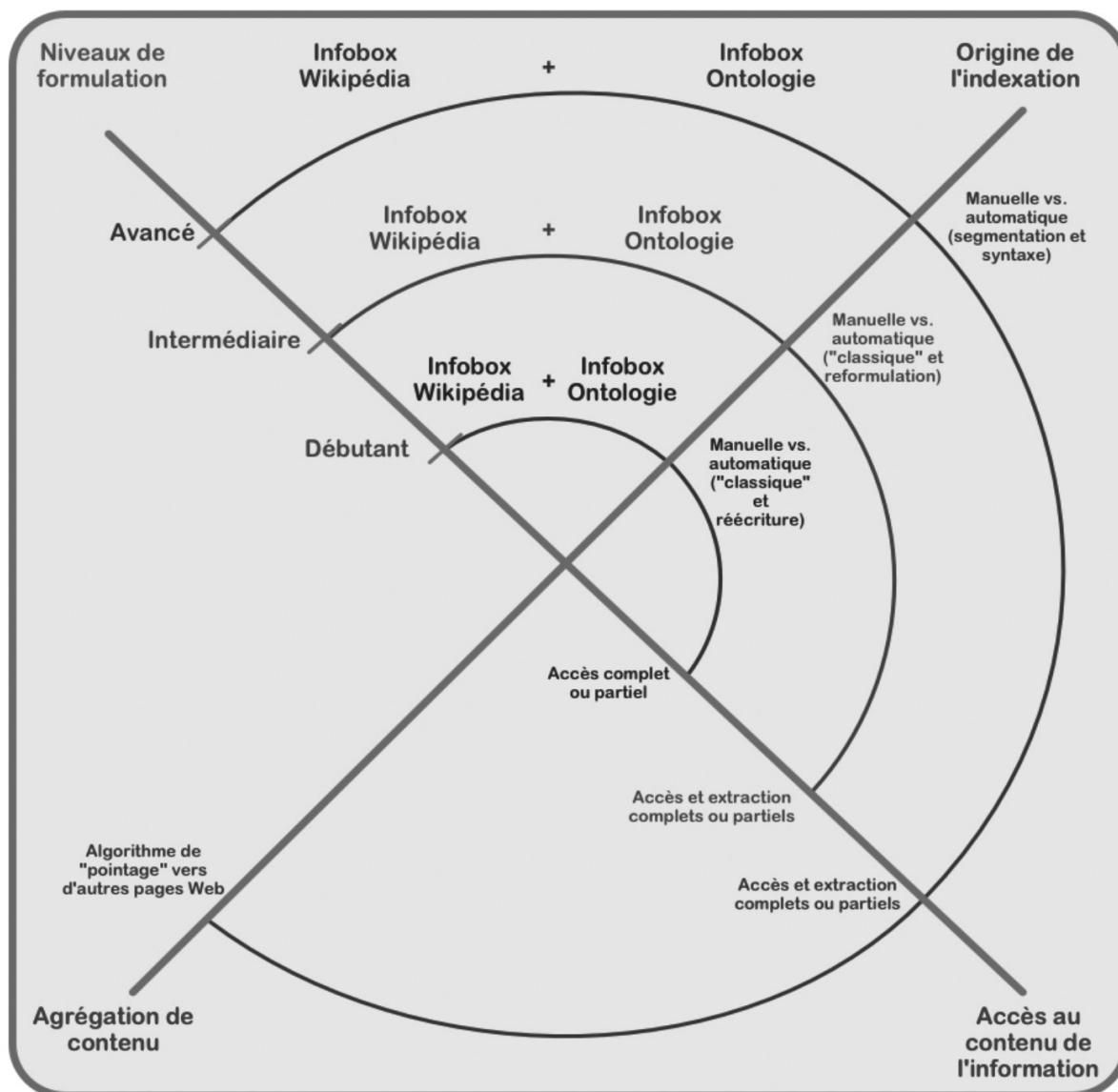


Figure 3 : Progression spiralaire des notions d'indexation manuelle et automatique du niveau débutant à avancé. Les notions sont wwddécomposées en trois caractéristiques : l'origine de l'information, l'accès au contenu de l'information et l'agrégation de contenu. Les notions sont abordées à partir de la comparaison entre l'infobox de la page Wikipédia et l'infobox d'ontologie, présente sur la page de résultats de la plupart des moteurs de recherche.

celles liées à la recherche informationnelle. Il offre une alternative à l'étude du portail documentaire pour enseigner la comparaison entre l'indexation humaine et l'indexation automatique. Par ailleurs, les professeur.e.s documentalistes, et *a fortiori* les élèves, ont remarqué l'existence des différents types d'encarts informationnels, d'autant plus

remarquables qu'ils sont souvent situés en haut de la page de résultats. Que pensent les élèves de ces infoboxes ? Comment les utilisent-ils ? Quels sont leurs représentations ? Telles sont les questions à leur poser, soit dans une autre étude, soit en classe afin de les mettre face à une situation-problème quant à l'origine de l'information.

Références bibliographiques

AIDE:INFOBOX. (2019). In *Wikipedia*. <https://fr.wikipedia.org/wiki/Aide:Infobox>

APDEN (à paraître). Indexation automatique In *Wikinotions*. <http://wikinotions.apden.org/index.php?title=Accueil>

BOULOGNE, A. (2004). *Vocabulaire de la documentation*. Paris: ADBS Éditions.

BROUDOUX, E. (2012). Vers l'objet documentaire (re)contextualisé. Objets documentaires numériques ?. *9e congrès des enseignants documentalistes de l'Éducation nationale, Paris*, 47-51. https://archivesic.ccsd.cnrs.fr/sic_00715868

CACALY, S., LE COADIC, Y.-F., POMART, P.-D. & SUTTER, E. (2004). Indexation. Dans S. Cacaly (dir.), *Dictionnaire de l'information* (p. 114-116). Paris: Armand Colin.

DELESTRE, N. & MALANDAIN, N. (2017). *Du web des documents au web sémantique*. Bois-Guillaume: KLOG Éditions.

EL HACHANI, M. (1997). *L'indexation automatique* (Mémoire de DEA, ENSSIB, Lyon). Récupéré le 10 novembre 2021 de la Bibliothèque de l'ENSSIB : <http://enssibal.enssib.fr/bibliotheque/documents/dea/elhachani.pdf>

KNOWLEDGE PANEL HELP, SANS DATE. How Google's Knowledge Graph works. Consulté le 1 novembre 2021, à l'adresse: <https://support.google.com/knowledgepanel/answer/9787176?hl=en>

KOVACS, S., & DIBIE, G. (2006). L'index du texte littéraire aux XVII^e et XVIII^e siècles, entre outil de consultation et discours critique. In I. Timimi et S. Kovacs (dir.), *Indice, index, indexation : Actes du colloque organisé par les laboratoires CERSATES et GERICO de l'université Lille-3 des 3 et 4 novembre 2005* (p. 63-80). Paris: ADBS Éditions.

RABAT, F. (2008). *Une année avec Google (suite)*. Récupéré le 14 avril 2020 sur le site de la documentation de l'académie de Normandie Rouen : <http://documentation.spip.ac-rouen.fr/?Une-annee-avec-Google-suite>

REYNAUD, F. (S. D.). Accueil : Historique des versions In *WikiNotions*. Consulté le 15 novembre 2020, à l'adresse <http://wikinotions.apden.org/index.php?title=Accueil&action=history>

RENVOISÉ, É. (2020). *Didactisation de l'indexation automatique à travers les moteurs de recherche : Étude des encarts informationnels ou infoboxes*. Education. Récupéré le 15 novembre 2020 sur le site du dépôt universitaire des mémoires après soutenance DUMAS : <https://dumas.ccsd.cnrs.fr/dumas-02893472>

ROUSSEAU, J. (2015). *Le moteur de recherche et les collégiens : Quelles représentations, quels obstacles, quelle didactisation ?* Education, 364. Récupéré le 27 mars 2020 sur le site des archives-ouvertes HAL : <https://dumas.ccsd.cnrs.fr/dumas-01202004>

SIRE, G. (2018). Web sémantique : Les politiques du sens et la rhétorique des données. *Les Enjeux de l'information et de la communication*, N° 19/2(2), 147 160. Consulté le 20 novembre 2021, à l'adresse : <https://www.cairn.info/revue-les-enjeux-de-l-information-et-de-la-communication-2018-2-page-147.htm>

SOGLIUZZO, G. (2013). *Comment fonctionne un moteur de recherche : L'exemple de Google*. Récupéré le 14 avril 2020 sur le blog Culture de l'info et des médias en lycée : <http://beaumont-redon.fr/wp/cultureinfomedias/2013/11/20/comment-fonctionne-un-moteur-de-recherche-lexemple-de-google/>

Annexe

Annexe 1 : Questionnaire – Indexation automatique – Professeur.e.s documentalistes

Q1a – Abordez-vous la notion d'indexation automatique ?

OUI

NON

Q1c - Si oui, en lien avec quelles notions génériques ?

	MOTEUR DE RECHERCHE
	VEILLE
	PORTAIL DOCUMENTAIRE
	AUTRES (précisez)

Précision

Q1d - Dans quel cadre pédagogique ?

	INFORMATION-DOCUMENTATION
	EPI
	AP
	SNT
	INTERDISCIPLINARITÉ
	AUTRES (précisez)

Précision

Q1b - Pour quel(s) niveau(x) d'élèves enseignez-vous l'indexation automatique?

☐ **Si non, pourquoi ?**

Q2 – Abordez-vous la notion d'indexation en amont de la notion d'indexation automatique ?

OUI

NON

☐ **Si oui, pourquoi ?**

☐ **Si non, pourquoi ?**

3 - Les encarts informationnels, aussi appelés « infobox », sont une table de données, sous forme d'encadré ou de cartouche, présentant sommairement des informations importantes sur un sujet. Ils sont présents dans la page de résultats de certains moteurs de recherche sur le Web (Google et Bing).

Q3a – Avez-vous remarqué ces encarts informationnels sur une page de résultats ?

OUI

NON

☐ Si oui, lesquels ?

	ENCART WIKIPÉDIA
	ENCART IMAGES OU VIDÉOS
	ENCART DE DÉFINITION

Q3b - Abordez-vous la question des encarts informationnels avec les élèves ?

OUI

NON

☐ Si oui, que souhaitez-vous que les élèves retiennent de l'étude des encarts informationnels ?

--

Q4 – Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ?

OUI

NON

☐ Si oui, lesquelles ?

--

Q5 – Quelle méthode pédagogique vous paraît être la plus pertinente pour enseigner l'indexation automatique, les moteurs de recherche ou les encarts informationnels (Justifiez) ?

--

Q6 – Depuis combien de temps exercez-vous en tant que professeur.e-documentaliste ?

--

Q7 – Dans quelle Académie enseignez-vous ?

--

WebFinder et WebLink :

les algorithmes

à portée des élèves

◆ Florian Reynaud

Professeur documentaliste dans l'académie de Lyon

Les algorithmes sont l'affaire de toutes et tous, avec des enjeux palpables en matière de maîtrise et de contrôle de l'information, ainsi qu'en matière de souveraineté individuelle dans la pratique du *Web*. Qu'il s'agisse de se déplacer avec son smartphone, de chercher des renseignements, d'échanger et de se divertir sur les réseaux sociaux numériques, les algorithmes « dirigent » nos activités, ils « façonnent » notre quotidien. Cette notion d'algorithme, dans sa relation à l'information, oblige à poser la question de son enseignement dans le cadre de l'information-documentation. Dans une approche technique, au-delà de simples observations du *Web*, le développement d'une simulation de moteur de recherche peut apparaître comme un outil intéressant pour mieux faire découvrir aux élèves cet aspect particulier du *Web*, et ce dans une progression pluriannuelle sur le sujet.

L'information-documentation et les algorithmes, une relation étroite

La documentation est affaire de procédures. La manipulation de l'information, par exemple pour la constitution de documents tertiaires, de catalogues, est affaire de rigueur technique. Au fur et à mesure, les technologies sont venues soutenir ce travail, avec des instructions informatiques normées, développées pour faciliter les opérations

de documentation et de recherche. Ainsi, la documentation devient une affaire d'algorithmes. Parmi les professionnels de ce domaine, les professeurs documentalistes sont de ce fait amenés à composer avec ces instructions plus ou moins complexes, sans avoir à les maîtriser : le renseignement d'une notice revient à remplir les cellules d'une base de données et la recherche au sein de cette base de données, notamment par les élèves, est soumise aux algorithmes.

Une question se pose alors. Si l'on ne peut être professeur documentaliste sans maîtriser les normes de catalogage et les techniques de classement, peut-on l'être sans comprendre les algorithmes ? Voire sans les enseigner ?

Comme tous les utilisateurs du *Web*, nous sommes confrontés au quotidien aux algorithmes. En tant que professionnels de l'information, nous sommes sensibles à leur importance, ainsi qu'à leurs enjeux en matière de choix individuels, de consommation, de politique... Il est une chose de visualiser cette influence de l'algorithme, au quotidien, en observant les affichages sur un mur de réseau social numérique, en constatant des différences de résultats d'un même moteur de recherche entre plusieurs individus. Une autre voie peut être d'ouvrir le capot et de « mettre le nez dans le moteur », afin de comprendre autrement le fonctionnement de ces algorithmes.

Cette proposition découle de bilans successifs, suite à des séances pédagogiques en information-documentation en collège. Ainsi, l'observation par les élèves, en classe, des réseaux sociaux numériques, des moteurs de recherche, permet d'aborder la question des algorithmes, d'en soupçonner l'existence, mais sans les toucher, sans véritablement en comprendre l'essence. De même, si une approche du *Web* par la navigation permet d'envisager l'existence d'un réseau mondial, elle n'est pas complète sans la connaissance des aspects techniques associés au stockage des fichiers du *Web* dans des serveurs et *data centers*. Tout comme l'approche du livre s'accompagne bien de connaissances sur les modalités et procédures d'édition et de fabrication de ce livre.

L'objectif est alors de donner les moyens aux élèves d'observer les algorithmes au plus près, en relation avec la manipulation, au sens technique du terme, de données informationnelles. Il ne s'agit pas pour autant de chercher à développer des compétences avancées en programmation. C'est alors qu'est née l'idée de simulations qui permettent une double visualisation, systématique, avec d'un côté l'interface publique, de l'autre la base de données et certains algorithmes. Le propos pédagogique est alors de demander aux élèves de comprendre les liens, les va-et-vient, entre ce qu'on leur propose publiquement et ce qui est stocké dans la base de données. En rendant visible l'invisible, on ajoute à l'observation simple des sites *web*, une approche du fonctionnement technique.

Dans une progression info-documentaire, ce processus pédagogique en collège accompagne d'autres découvertes des algorithmes, en mathématiques et en technologie, dans lesquelles on manipule surtout des instructions de déplacement dans l'espace. Elle précède par ailleurs des approches plus techniques, notamment en Sciences numériques et technologie (SNT), au lycée, dans lesquelles les élèves vont être amenés à programmer, avec alors la possibilité de créer

un moteur de recherche, ou encore un catalogue documentaire.

Une simulation de moteur de recherche en 5^{ème}

En préalable à cette découverte de la simulation, les élèves ont été amenés, en 6^{ème} et en 5^{ème}, à observer les pages de résultats des moteurs de recherche selon différentes requêtes proposées pour un même sujet. Concernant les aspects techniques du *Web*, ils ont été initiés au langage HTML avec une heure de création d'une page web avec du code simple. Ces deux activités, avec les compétences qui y sont développées, sont invoquées lors du travail sur la simulation développée pour ce projet : *WebFinder*. Le travail spécifique à cette simulation permet de travailler les notions d'algorithme, de base de données, de code, de moteur de recherche. D'autres notions ont trait à la séance et peuvent être envisagées pour des niveaux d'enseignement supérieurs, telles que les notions d'indexation automatisée, de référencement. Par ailleurs, une ouverture de fin de simulation permet d'introduire les notions d'économie de l'information et de publicité, mais sans approfondissement dans la séance elle-même.

La simulation se présente sous la forme d'un site *web* public accessible à toutes et tous, sans conditions d'accès¹. Cette simulation s'appuie sur dix pages web de dix sites web différents, autour d'une seule thématique. Pour être en adéquation avec les programmes de 5^{ème}, c'est la « digestion » qui a été choisie. Trois niveaux sont proposés, et trois modes différents. Les niveaux correspondent à des niveaux scolaires, pour le début du collège, la fin du collège et le lycée. Les modes correspondent au temps disponible en séance, soit une ou deux heures, mais aussi à la possibilité de faire le parcours en autonomie. La simulation comprend différentes étapes, avec des questions à chaque étape. Avec le mode « en autonomie », les réponses sont données au bout d'un certain nombre d'échecs.

¹ La simulation est disponible sur : <https://iddocs.fr/webfinder/> ; le code et l'explication technique de la simulation, sous licence ouverte, sont disponibles sur : <https://profdoc.iddocs.fr/spip.php?article56>

Supposons...

Sur le Web, seulement 10 pages !



Mentions légales - 2016-2021

Niveau :

- ☒ Débutant (10-12 ans)
- ☐ Intermédiaire (13-15 ans)
- ☐ Avancé (16-18 ans)

Mode :

- ☐ En classe (2 heures)
- ☒ En classe (1 heure)
- ☐ En autonomie



Au fur et à mesure de la simulation, les élèves disposent à la fois de l'observation de la base de données, avec des informations qui vont aller en se complexifiant, sur la droite de l'écran, et de l'interface des résultats du moteur de recherche, sur la gauche de l'écran. Selon le niveau choisi, les indices pour répondre seront plus ou moins importants, les informations de la base de données plus ou moins explicites.

Pour chaque étape, une ou plusieurs questions sont posées. Elles nécessitent d'observer les

éléments d'interaction entre base de données et affichage des résultats. Les élèves ont ainsi à relever les critères de classement des résultats en observant la base de données, avec dans un premier temps la pertinence (au nombre de mots-clés), puis la popularité (au nombre de visites), enfin la notoriété (au nombre de liens vers la page, depuis d'autres pages web).

À chaque étape, en classe, on fait le point, soit quand tous ont terminé une étape, soit quand on estime la fin du temps imparti avant de donner

WebFinder
moteur de recherche

Rechercher...

Observe les résultats (dans la partie de gauche).

D'après toi, et en observant la base de données sur quel critère (ou colonne) sont ordonnés les 10 résultats ?

Vérifier

La base de données :

Tu peux trier par colonne en cliquant sur un titre de colonne. Pour rappel, tu as tapé "digestion".

Titre de la page web (< titre >)	Adresse URL	Texte	Nombre de liens, sur le Web, vers cette page	Nombre de visites sur cette page	Nombre de mots du terme "digestion" dans le texte
Comment digérons-nous les aliments ? - EurekaSanté par VIDAL	eurekasante.vidal.fr...	Comment digérons-nous les aliments ? Mis à jour : Lundi 30 Novembre 2009 42 Sommaire Comment digérons-nous les aliments ? La digestion commence dans la bouche La digestion continue dans l'estomac ...	21	24520	24
Digestion humaine - Wikipédia	fr.wikipedia.org/wiki/Digestion_humaine	Cet article traite du développement, de l'organisation et du fonctionnement du système digestif de l'homme. Sommaire 1 Développement embryologique du tube digestif 1.1 Généralités 1.2 Organogénèse 1.2.1 Oesophage 1.2.2 Estomac 1.2.3 R...	10	56100	38
La digestion des aliments - Dictionnaire visuel	www.ikonet.com/fr/la...	La digestion des aliments L'énergie indispensable au fonctionnement de notre corps nous est fournie par l'alimentation. Une dizaine d'organes, composant le système digestif, s'allient pour décomposer...	61	13045	8
Leçon sur notre système digestif	www.larecre.net/fr/sciences/etre_humain/lecon_sur_notre...	Notre système digestif (Leçon) DEFINITIONS ET LOCALISATIONS La langue : Organe du goût. Les dents : Eléments implantés dans les mâchoires servant à couper et broyer les aliments. Le pharynx ...	75	19630	29
Alimentation - digestion - absorption - métabolisme - excrétion	pst.chez-alice.fr/svitu/m/mange.htm	L'alimentation ou prise alimentaire ou encore ingestion est la première étape de la fonction digestive au sens large. Les aliments liquides sont avalés, les aliments solides sont mastiqués (appareil masticateur comprenant les dents) et subissent un début de digestion (modification chimique) par les ...	0	3821	61
Fiche exposé : la digestion des aliments	www.lepetitquotidien...	Ce que nous mangeons donne de l'énergie à nos cellules pour faire fonctionner notre corps. Mais les aliments que nous avalons sont trop gros pour arriver directement aux cellules par le sang. Ils doivent...	21	15497	1
Encyclopédie Larousse en ligne - Digestion, des aliments aux nutriments	www.larousse.fr/ency...	Digestion, des aliments aux nutriments Digestion, des aliments aux nutriments La plupart des aliments que nous consommons sont constitués de substances simples : glucides, lipides, protéines. ea...	78	37524	4

Résultats de la recherche

1. Alimentation - digestion - absorption - métabolisme - excrétion
pst.chez-alice.fr/svitu/m/mange.htm
L'alimentation ou prise alimentaire ou encore ingestion est la première étape de la fonction digestive au sens large. Les aliments liquides sont avalés, les aliments solides sont mastiqués (appareil masticateur comprenant les dents) et subissent un début de digestion (modification chimique) par les ...

2. Digestion humaine - Wikipédia
fr.wikipedia.org/wiki/Digestion_humaine
Cet article traite du développement, de l'organisation et du fonctionnement du système digestif de l'homme. Sommaire 1 Développement embryologique du tube digestif 1.1 Généralités 1.2 Organogénèse 1.2.1 Oesophage 1.2.2 Estomac 1.2.3 R...

3. Leçon sur notre système digestif
www.larecre.net/fr/sciences/etre_humain/lecon_sur_notre...
Notre système digestif (Leçon) DEFINITIONS ET LOCALISATIONS La langue : Organe du goût. Les dents : Eléments implantés dans les mâchoires servant à couper et broyer les aliments. Le pharynx : Partie du tube digestif située entre la bouche et l'oesophage. L'oesophage : Partie du tube diges...

4. Comment digérons-nous les aliments ? - EurekaSanté par VIDAL
eurekasante.vidal.fr/nutrition/corps_aliments/digestion
Comment digérons-nous les aliments ? Mis à jour : Lundi 30 Novembre 2009 42 Sommaire Comment digérons-nous les aliments ? La digestion commence dans la bouche La digestion continue dans l'estomac Le rôle de l'intestin grêle dans la digestion

la solution. Ces restitutions collectives permettent de faire le point sur le fonctionnement du moteur de recherche, notamment sur « la cuisine des algorithmes », comparaison entre l'instruction informatique et la recette de cuisine qui est proposée aux élèves. Cela passe par l'analyse d'une instruction de recherche des résultats.

En fin de simulation, l'affichage d'une publicité, ou « annonce », en tête de résultats, permet d'aborder cette autre notion et de faire le point sur le caractère mercantile des moteurs de recherche (qu'on peut accompagner d'une recherche-observation sur *Google* autour des hôtels, par exemple, recherche typique de mise en valeur d'annonceurs).

Vers le niveau le plus élevé, on propose, au cours de la simulation, des questions sur le droit à l'oubli, sur l'instruction de ne pas indexer une page, sur

le texte affiché dans les résultats... Pour les plus avancés, un travail récapitulatif de construction de définitions, pour chaque notion, est proposé en bilan de l'activité.

Une autre approche avec les réseaux sociaux numériques

Dans la progression suivie en collège, et cette fois pour l'approche relative aux réseaux sociaux numériques, un même type de simulation est proposé pour les élèves de 3ème, sous l'intitulé *WebLink*. Il s'agit alors d'avoir, sur la partie gauche de l'écran, l'interface du réseau, et, sur la partie droite, la base de données, avec là aussi des étapes sous forme de questions.

Cette simulation n'est pas publique, pour l'heure. Mais elle peut être installée par qui le veut sur un hébergement local ou distant².

- l'interface publique -

- la base de données -

Dans la table des **utilisateurs** (table "user"), quand tu t'es inscrit correctement, les informations sont enregistrées dans la base de données :

Id.	Pseudo	Nom	Prénom	Naissance	Sexe	Sport	Musique
30	Batman	Wayne	Bruce	2010-03-02	M		
8	La Calas	Esteban	Maria	2007-03-12	F		
7	Marco	Anrieux	Marc	2007-12-25	M		
6	Flipper	Prescott	Sidney	2007-09-25	F		
5	le grand	Berkovitch	Yvan	2007-10-26	M		
4	Kalina	Allaoua	Khalissa	2007-06-03	F		
3	Johnny	Morin	Franck	2007-06-15	M		
2	marina	Poussin	Marina	2007-11-04	F		
1	Kevin	Durand	Kevin	2007-03-28	M		

Ta ligne :

Id.	Pseudo	Nom	Prénom	Naissance	Sexe	Sport	Musique
30	Batman	Wayne	Bruce	2010-03-02	M		

Un code permet une **mise en relation avec un autre utilisateur** du réseau, de sexe opposé ! avec sa date de naissance s'il l'a donnée. Si tu es parmi les premiers inscrits, recharge cette page pour observe des changements.

Il y a une table de **publicités** pour produits (table "pub") :

Pour l'instant, la base de données ne te connaît pas beaucoup.

Observe les deux parties de l'écran pour répondre aux questions suivantes :

Quelles sont les deux « tables » que tu peux voir dans la base de données ?	La première :
	La deuxième :
Pour un produit à vendre, quelles sont les colonnes utilisées dans la table des publicités pour faire le lien entre le produit et l'utilisateur du réseau ?	La première :
	La deuxième :
Pourquoi te propose-t-on ces deux publicités ?	
Quelle information sur toi, affichée à gauche, n'as-tu pas donnée lors de l'inscription ?	
Sur quel critère te suggère-t-on une autre personne ?	
Quelle information personnelle te donne-t-on sur elle ?	

² Le code et l'explication technique de la simulation, sous licence ouverte, sont disponibles sur : <https://profdoc.iddocs.fr/spip.php?article60>

On retrouve le principe des algorithmes, pour un autre type d'instructions, un autre type de sites *web*. Ce sont globalement les mêmes notions qui sont développées par la manipulation et l'observation, ici dans le contexte des médias sociaux en ligne. On y ajoute toutefois la notion de trace numérique, en travaillant sur le stockage des données, et sur leur persistance dans les serveurs malgré leur suppression des affichages sur les profils.

Conclusion

Si le professeur documentaliste ne peut être un enseignant de programmation, son champ d'enseignement, l'information-documentation, suppose une connaissance du fonctionnement du *web* et des algorithmes qui le façonnent. Le sujet des simulations présentées dans cet article est

aussi un moyen d'inviter la profession à s'emparer de cette thématique, sans aller dans le codage, pour ensuite transmettre ces connaissances aux élèves, en complément d'autres approches pédagogiques.

En évolution depuis 2016, ces deux simulations sont le sujet de réflexions continues, avec des bilans réguliers. Les retours des professeur.e.s documentalistes utilisateurs sont bien entendu les bienvenus. Les prochaines étapes ? Rafraîchir l'interface graphique, réfléchir à d'autres thématiques de recherche pour *WebFinder*, à une amélioration du réseau *WebLink*, dont on peut repenser le fonctionnement afin qu'il puisse être public. Tout cela passe par une meilleure observation des élèves en activité et en évaluation afin de cerner ce qu'ils retiennent, à chaque fois, et ainsi corriger l'outil et certaines questions, certaines interactions, pour un meilleur apprentissage.

L'adolescence au cœur de l'économie numérique

Prendre la mesure de la
puissance des réseaux
socio-numériques sans pensée
magique ni complotisme



Varia

◆ **Sophie Jehel**

Maîtresse de conférences HDR en sciences de l'information et de la communication,
Université Paris 8, CEMTI

Les adolescents jouent un rôle prépondérant dans l'économie numérique. Ils contribuent pour une part essentielle au succès commercial des grandes plateformes en ligne. Leur participation aux interactions sur les plateformes peut-elle pour autant être interprétée comme l'effet direct de la puissance de ces médias ? Peut-on analyser leur passion pour leurs comptes de réseaux socio-numériques comme le seul résultat de la puissance de l'économie de l'attention et de la captologie ? Sont-ils vraiment ces marionnettes présentées dans *The social dilemma* dont les activités numériques peuvent se réduire à une réaction aux stimuli des algorithmes ? Je souhaiterais montrer que la situation d'usage est bien plus complexe, si l'on fait attention à ce qui se passe du côté de la réception des messages. Je m'appuierai ici notamment sur l'étude de la réception des images violentes, sexuelles et haineuses que j'ai conduite entre 2015 et 2017 avec une équipe de psychologues cliniciens¹.

Les adolescents : une cible privilégiée des industries numériques

Les adolescents sont devenus une cible commerciale privilégiée des industries numériques et en particulier des réseaux socio-numériques. Parmi les six entreprises les plus valorisées au monde par capitalisation, cinq sont des entreprises numériques, *Apple*, *Microsoft*, *Amazon*, *Alphabet* (ex : *Google*), *Meta* (ex-*Facebook*), dénommées jusqu'à récemment les « GAFAM ». Selon les trimestres, certaines de ces six places leur sont disputées par la *Saudi Arabian Oil Company* ou par les entreprises chinoises *Tencent* ou *Alibaba*. Ces cinq industries numériques offrent des plateformes de réseaux socio-numériques comme *Youtube*, *Instagram*, *Facebook*, *Whatsapp* ou de réseaux de *streaming* direct comme *Twitch*, qui sont devenues incontournables dans la panoplie numérique des adolescents. Elles sont concurrencées dans la conquête des adolescents par

¹ Elle est présentée notamment dans l'ouvrage, Jehel Sophie, Gozlan Angélique (dir.), *Les adolescents face aux images trash sur internet*, Paris, In Press, 2019. J'ai pu en développer l'analyse dans l'ouvrage *L'adolescence au cœur du web*, Ina, à paraître au printemps 2022

Snapchat et *Tiktok*, qui devrait prochainement entrer en bourse. La concurrence se fait sur la capacité de chaque plateforme d'accroître son réseau d'utilisateurs, ce qui est particulièrement le cas en 2020 et 2021 de *Tiktok*, *Twitch* et d'*Instagram*, renforcées encore par les périodes de confinement. L'enjeu pour chacune est alors de monétiser au mieux la présence des publics, d'où l'installation d'écrans publicitaires, mais aussi de systèmes facilitant les transactions commerciales directement depuis les comptes des utilisateurs.

Les adolescents ne sont évidemment pas les seuls publics visés par ces entreprises transnationales. Ils constituent cependant des usagers particulièrement précieux, pour leur capacité à adopter de nouvelles routines numériques, de nouvelles modalités de socialisation et à être ainsi fidélisés pour les années suivantes. Les adolescents correspondent en effet à des usagers particulièrement disponibles, à la recherche d'une connexion fréquente avec leurs pairs et d'une autonomisation, voire d'une forme de distinction, vis-à-vis des générations qui les précèdent, parents et aînés.

Le modèle économique des plateformes numériques repose d'abord sur l'intensification des publications réalisées par les usagers du réseau. Le terme de plateforme numérique est choisi ici comme métonymie pour désigner des industries qui utilisent un dispositif computationnel permettant de mettre en relation différents versants d'une place de marché (usagers, annonceurs, producteurs de contenu, contenus), d'observer et de calculer les différents échanges entre eux. Le dispositif de la plateforme est en effet essentiel dans la gouvernance des usagers de ces réseaux. Dans le discours d'accompagnement des plateformes, « l'injonction à publier² » est un argument stratégique.

Le double sens de la participation numérique

L'injonction à publier trouve un terrain favorable avec l'idéologie du « web participatif », largement promue comme une nouvelle forme de liberté pour les usagers ordinaires, qui peuvent, grâce aux nouvelles technologies et en particulier aux plateformes de réseaux sociaux numériques, développer des formes larges de visibilité et d'accès à la reconnaissance sociale. Ces objectifs, que sont la visibilité et la reconnaissance, sont des moteurs puissants d'expression de soi dans la « société des individus »³, dans laquelle les citoyens cherchent à être connus des autres en tant que différents, singuliers et disposant d'une certaine maîtrise de leurs identités sociales.

Si les usagers ont un désir certain de publier et d'échanger, ils ont aussi des craintes qui freinent ces publications : peur du ridicule, peur de la publicité elle-même, lorsqu'elle s'avère dépasser leurs cercles de connaissances, appréhension vis-à-vis de la difficulté qui consiste à articuler les différents groupes de sociabilité dont les valeurs diffèrent. L'identification de ces résistances a donné lieu du côté des plateformes à la recherche de dispositifs émotionnels susceptibles de les vaincre, que l'on désigne comme ceux du « web affectif »⁴.

Du côté des usagers, l'expression de soi sur les plateformes relève d'une liberté et d'un souci de reconnaissance sociale. Pour les plateformes numériques, ces publications peuvent néanmoins s'analyser comme une manne, un « travail du clic » ou « travail social en réseau »⁵ gratuit ou très peu rémunéré. Il est en effet crucial pour les perspectives de profit et de valorisation boursière de ces plateformes de bénéficier d'une armée la plus large possible d'« ouvriers » capables de produire

² Proulx Serge, *La participation numérique, Une injonction paradoxale*, Paris Presse des mines, 2020

³ Elias Norbert, *La société des individus*, Paris, Fayard, 1991

⁴ Alloing, Camille, Pierre, Julien, « Faire attention, ou l'affectivité en contexte numérique » in Jehel Sophie, Saemmer Alexandra (dir.) *Education critique aux médias en contexte numérique*, Presses de l'ENSSIB 2020, p. 87-94

⁵ Casilli Antonio, *En attendant les robots. Enquête sur le travail du clic*. Paris, Seuil, 2019

des contenus, susciter des échanges et délivrer des données personnelles.

Les adolescents se retrouvent donc aux prises avec des entreprises transnationales qui leur proposent des espaces d'expression quasiment infinis, des perspectives de reconnaissance sociale, voire de monétisation, à une échelle encore inconnue jusque-là. Comment apprécier la puissance et l'impact des réseaux sociaux numériques sur leurs vies ?

Le retour de la théorie des effets puissants des médias est-il si convaincant ?

À l'occasion des révélations passionnantes des anciens employés de ces entreprises qui développent leurs activités sur tous les continents et revendiquent des centaines de millions voire des milliards d'utilisateurs, se profile une puissance commerciale démesurée. Nous avons découvert ou pu confirmer nos intuitions qu'elle s'exerce quasiment sans contrôle public, privilégiant ses intérêts commerciaux sur leur modération en favorisant « l'engagement », c'est-à-dire l'évaluation affective des contenus, grâce au bouton « j'aime », aux modalités de partage, aux notifications. Leurs témoignages ont été relayés notamment dans le documentaire *The Social Dilemma* de Jeff Orlowski, diffusé sur Netflix. Le fonctionnement des réseaux sociaux numériques contribuerait à l'aggravation de la santé mentale des adolescents, se traduisant par une augmentation des tentatives de suicide, et la destruction de la démocratie, en favorisant les discours sectaires, la désinformation et au final la « guerre civile ».

Ce discours participe d'une mobilisation aux États-Unis en faveur d'une régulation économique et politique de ces entreprises. Il a été paradoxalement promu par Netflix, plateforme

de moindre envergure (relativement) mais dont la croissance repose elle aussi sur l'engagement, le *marketing* des réseaux sociaux numériques et les algorithmes de profilage. Le documentaire est utilisé aujourd'hui comme outil de sensibilisation du secteur éducatif, aux États-Unis. Sur le fond de son argumentation, il réhabilite un modèle théorique bien connu de la sociologie des médias depuis 60 ans, celui des effets puissants de la propagande, appelé aussi modèle de la seringue hypodermique ou de l'arme magique (« *magic bullet* »⁶).

Or, la sociologie des médias a mis en évidence la complexité du rapport des publics aux contenus médiatiques, faits de confiance et de défiance, la diversité des interprétations par les publics, dans la lignée des *cultural studies*, selon leur type de socialisation et leurs grilles politiques. L'interactivité des publics permise par les plateformes numériques les conduirait-elle à une dose de « passivité » et de soumission aux manipulations médiatiques supplémentaires ? Il s'agirait là d'un incroyable paradoxe.

L'importance de porter attention au travail émotionnel des publics

J'ai toujours considéré que les médias audiovisuels constituaient, de par leur fonctionnement par répétition de thèmes, de schèmes et de stéréotypes, un environnement qui nourrit les repères et les représentations du monde, ainsi que le postule la *cultivation perspective* de Gerbner⁷. Nous l'avons développé dans un premier ouvrage avec Divina Frau-Meigs⁸, en partant de l'analyse d'une semaine de programmes de fiction à la télévision française et des représentations des violences ainsi que du modèle politique sécuritaire et autoritaire qu'elles véhiculaient. Mais il nous manquait la connaissance de leur réception. Depuis, je me

⁶ Lasswell Harold D., *Propaganda Techniques in the World War*, NY, Peter Smith, 1938

⁷ Gerbner George, Gross Larry, Morgan Michael, Signorelli Nancy, "Growing up with television the cultivation perspective", in Bryant James, Zillman Dolf (dir.) *Perspectives on media effects*, Hillsdale NJ, Lawrence Erlbaum, 1994, p. 17-40

⁸ Frau-Meigs Divina, Jehel Sophie, *Les écrans de la violence. Enjeux économiques et responsabilités sociales*, Paris, Economica, 1997

suis attachée dans mes recherches à connaître plus précisément les critères des choix médiatiques des enfants et des adolescents. J'ai pu mettre en évidence, à partir d'une enquête auprès de plus de 1000 préadolescents, le rôle des parents dans la construction d'une médiation vis-à-vis des médias, et de la transmission par certains d'un goût en faveur de représentations violentes⁹. Cela ne minimise pas le rôle des choix de programmation des chaînes ni des radios.

Mes dernières recherches m'ont permis d'approfondir la réception des images violentes, sexuelles et haineuses sur les plateformes numériques. La théorie du « travail émotionnel » d'Arlie R. Hochschild¹⁰ m'a aidée à accéder à la complexité du travail éthique réalisé par les adolescents face à des images qui sont perçues par eux comme violentes, sexuelles ou haineuses, lorsque sur les plateformes ils sont soumis à « l'injonction à la publication » et la manifestation de leurs émotions intimes. Le travail émotionnel concerne l'écart maintenu entre les émotions ressenties et celles qui sont manifestées dans un espace professionnel ou public. La théorie de la manipulation, qui a servi d'inspiration aux repentis des plateformes lorsqu'ils cherchaient chez *Google*, *Youtube*, *Facebook* ou *Twitter* à renforcer « l'engagement » des usagers à travers leur profilage et les fils de recommandation, conduit à croire à une soumission généralisée des publics au travail du clic et aux contenus recommandés.

Les deux axes du travail émotionnel des adolescents sur les plateformes numériques

Enquêtant auprès des adolescents de 15-18 ans comme j'ai pu le faire, accompagnée d'une équipe de psychologues cliniciens¹¹, j'ai pu représenter les coordonnées du travail émotionnel qu'ils

réalisent par rapport à deux axes principaux. Face à ces images, ils peuvent éviter leur visionnage, les outils de « scrolling » le permettent, même si les plateformes déclenchent la lecture directe des vidéos à la moindre digipulation, c'est-à-dire au passage du doigt ou de la souris sur l'écran. Exposés à des images sexuelles sur les sites de téléchargement illégaux, les filles m'ont expliqué passer leur temps à appuyer sur la petite croix, pour fermer ces écrans. **Un premier axe** oppose ainsi ceux qui veulent se confronter à ces images-chocs et ceux qui préfèrent les esquiver. Un second axe permet de définir les modalités de leurs réactions que j'ai situées entre le « déconstrôler » et le « contrôler du déconstrôler ». J'emprunte ces notions à Cas Wouters¹², et à son analyse du déconstrôler des mœurs depuis la seconde moitié du XXe siècle, dans la lignée de l'hypothèse de la civilisation des mœurs de Norbert Elias. Le déconstrôler explique pour une part l'abaissement de la censure des contenus sexuels et violents sur le *Web*. Du fait de ce mouvement de « déconstrôler » des mœurs, les individus ont été socialisés à une certaine forme de tolérance vis-à-vis des contenus médiatiques, et à des modalités d'expression de leurs émotions moins corsetées que dans les décennies précédentes. Les sociétés européennes et états-uniennes tolèrent une expression plus libre des émotions. Mais de nouvelles formes de contraintes et d'interdits sont également apparus par rapport à certaines émotions, celles du mépris social, du racisme ou de la haine notamment. Ce second mouvement est un « contrôler du déconstrôler ».

Les cinq modalités de l'adhésion aux images

Lorsqu'une image vient tracasser la conscience des jeunes utilisateurs du *Web*, certains vont plonger dans l'image, sans pouvoir s'en détacher,

⁹ Jehel Sophie, *Parents ou médias, qui éduque les préadolescents ? Enquête sur leurs pratiques TV, jeux vidéo, Internet, radio*, Toulouse, Erès, 2011

¹⁰ Hochschild Arlie Russell, *Le prix des sentiments, au cœur du travail émotionnel*. Paris, La Découverte, 1e ed.1983, 2017

¹¹ Jehel Sophie, Gozlan Angélique (dir.), *Les adolescents face aux images trash sur internet*, Paris, In Press, 2019

¹² Wouters Cas, *La civilisation des mœurs et des émotions : de la formalisation à l'informalisation*. In Bonny Yves, De Queiroz Jean-Manuel, Neveu Eric (dir.), *Norbert Elias et la théorie de la civilisation. Lectures critiques*, Rennes, PUR, 2003, p. 147-168

même après le visionnage, et sont en grande difficulté pour en donner une interprétation. Ce sont des comportements d'adhésion dont j'ai exploré la diversité. J'en ai distingué cinq formes.

- **L'adhésion-croyance** aux images violentes ou pornographiques se caractérise par une croyance en la « vérité » des images, dans la continuité de ce que Barthes avait désigné par « effet de réel ». Elle court-circuite la nature de représentation des images.

- **L'adhésion-dissociation** est propre aux engagements affectifs forts pendant le visionnage d'un film ou d'un jeu vidéo. Les joueurs intensifs sont souvent incapables de se détacher des significations explicites du jeu, mais sans confondre pour autant les règles du jeu et celles de la réalité sociale dans laquelle ils vivent, lorsque la partie s'arrête.

- **L'adhésion-sidération** relève d'une perte de contrôle face à certaines images, elle manifeste la puissance affectante de l'image-choc, qu'Angélique Gozlan a appelé *punctum-choc*. Les adolescents qui vivent dans des univers de délinquance ont tendance à rechercher ces images qui semblent stopper la pensée. Le choc de l'image est lié autant à la cruauté des actions représentées qu'à la vulnérabilité du sujet, dont il peut signaler une résonance à des vécus traumatiques.

- **L'adhésion-jouissance** est une autre forme de déconstruire le contrôle des émotions et s'accompagne d'un plaisir de la souffrance d'autrui, d'une dyspathie pour les victimes, réelles ou fictionnelles.

- **L'adhésion rationalisation** est au contraire la manifestation d'un surcontrôle. La destruction de la victime, dans les images violentes, est interprétée par le spectateur comme méritée, en raison de la transgression d'un interdit, notamment religieux, bien au-delà des normes de la société (d'où la dimension de sur-contrôle). Le dogme posé empêche l'expression d'une compassion ou même d'un doute sur la nature du « crime »

commis par la victime. Il m'est apparu comme une source de radicalisation cognitive, souvent articulé à des discours d'évitement strict de toute sexualité.

Ces formes d'adhésion se sont trouvées dans cette recherche dans tous les groupes de l'échantillon, mais avec une prépondérance particulière dans le groupe des jeunes vulnérables suivis par la protection judiciaire de la jeunesse et engagés dans des activités illégales (souvent du trafic de drogue).

Les trois autres stratégies de réception des images-chocs

A côté de ces modalités d'adhésion aux images qui pourraient correspondre - avec des nuances importantes - au modèle de la manipulation par les médias, j'ai observé trois autres stratégies de réception qui, sans toutes manifester une forme d'autonomie dans l'interprétation, relevaient d'une résistance aux affordances¹³ des dispositifs et à l'injonction à la publication :

- **L'évitement** des images, en particulier pour les images sexuelles, plus souvent par les filles.

- **L'indifférence**, le fait de passer rapidement, en cherchant à n'y prêter aucune attention, sur des images non désirées qui pourraient sinon, déclencher des émotions.

- La quatrième stratégie était celle de **l'autonomie**, c'est-à-dire la capacité, selon les cas, de manifester son empathie ou son point de vue, ou de le garder pour soi, de publier une image de soi ou de ne pas le faire. Ces adolescents s'appuyaient sur des ressources différentes. Dans les milieux favorisés sur le plan scolaire et social, plus souvent par l'intellectualisation, c'est-à-dire la capacité à resituer une image dans son contexte culturel ou médiatique. Dans les milieux populaires, la mise à distance du message reçu prenait appui plus souvent sur une empathie ou un sentiment de solidarité vis-à-vis des victimes d'actes violents.

¹³ L'affordance est un terme utilisé en psychologie puis en ergonomie pour désigner les potentialités d'actions offertes par un objet et son design

Le rôle de l'éducation aux médias et le développement de la réflexivité

Cette recherche m'a conduit à reprendre la question des moyens de l'éducation aux médias et à l'information, qui est devenue depuis 2015 un sujet de préoccupation des pouvoirs publics et des professionnels de l'éducation. Comment favoriser l'autonomie des jeunes lorsque ceux-ci sont engagés dans des modalités de réception de l'ordre de l'adhésion, de l'évitement ou de l'indifférence ? Comment ouvrir des espaces de réflexivité qui permettent de confronter des interprétations des images, et d'exprimer la diversité du répertoire des émotions qui traversent les publics, et ne sont pas réductibles aux différents boutons

et émojis proposés par les dispositifs formatés des réseaux socionumériques ? L'analyse des images, le développement de démarches d'enquêtes réalisées par les jeunes eux-mêmes, sur certaines images, sur la réception de ces images, sur les usages de certains de ces boutons, me semblent des pistes à développer. Favoriser aussi une expression créative qui ne soit pas la répétition d'un code de bonne conduite, mais l'exploration de zones plus secrètes, plus ambivalentes qui caractérisent le travail éthique des utilisateurs sur les plateformes, sont au cœur des recherches que je développe désormais.

Prochaine publication *L'adolescence au cœur de l'économie numérique. Travail émotionnel et risques sociaux*, aux éditions de l'INA, collection Etudes et controverses (pour janvier 2022).

Références bibliographiques

ALLOING, CAMILLE, PIERRE, JULIEN, « Faire attention, ou l'affectivité en contexte numérique » in Jehel Sophie, Saemmer Alexandra (dir.) *Education critique aux médias en contexte numérique*, Presses de l'ENSSIB 2020, p. 87-94.

BARTHES ROLAND, « L'effet de réel », *Communications*, 1968, 11, p. 84-89.

CASILLI ANTONIO, *En attendant les robots. Enquête sur le travail du clic*. Paris, Seuil, 2019.

ELIAS NORBERT, *La société des individus*, Paris, Fayard, 1991.

FRAU-MEIGS DIVINA, JEHEL SOPHIE, *Les écrans de la violence. Enjeux économiques et responsabilités sociales*, Paris, Economica, 1997.

GERBNER GEORGE, GROSS LARRY, MORGAN MICHAEL, SIGNORELLI NANCY, "Growing up with television the cultivation perspective", in Bryant James, Zillman Dolf (dir.) *Perspectives on media effects*, Hillsdale NJ, Lawrence Erlbaum, 1994, p. 17-40.

GOZLAN ANGÉLIQUE « Du punctum à la fascination : quand les images sur les réseaux sociaux font résonance aux processus adolescents », in Jehel Sophie, Gozlan Angélique, *Les adolescents face aux images trash sur internet*, 2019, In Press.

HOCHSCHILD ARLIE RUSSELL, *Le prix des sentiments, au cœur du travail émotionnel*. Paris, La Découverte, 1e ed.1983, 2017.

JEHEL SOPHIE, *Parents ou médias, qui éduque les préadolescents ? Enquête sur leurs pratiques TV, jeux vidéo, Internet, radio*, Toulouse, Erès, 2011.

LASSWELL HAROLD D., *Propaganda Techniques in the World War*, NY, Peter Smith, 1938.

PROULX SERGE, *La participation numérique, Une injonction paradoxale*, Paris Presse des mines, 2020.

WOUTERS CAS, *La civilisation des mœurs et des émotions : de la formalisation à l'informalisation*. In Bonny Yves, De Queiroz Jean-Manuel, Neveu Eric (dir.), *Norbert Elias et la théorie de la civilisation. Lectures critiques*, Rennes, PUR, 2003, p. 147-168.



Présidente :
Catherine Novel

Suivi éditorial :
Sandrine Vigato

GROUPE IRO - La Rochelle

Abonnement servi aux adhérents
Vente au numéro : 10 €

www.apden.org

ISSN 1260-7649