



HAL
open science

Didactisation de l'indexation automatique à travers les moteurs de recherche : étude des encarts informationnels ou infoboxes

Élodie Renvoisé

► To cite this version:

Élodie Renvoisé. Didactisation de l'indexation automatique à travers les moteurs de recherche : étude des encarts informationnels ou infoboxes. Education. 2020. dumas-02893472v2

HAL Id: dumas-02893472

<https://dumas.ccsd.cnrs.fr/dumas-02893472v2>

Submitted on 30 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Master MEEF

« Métiers de l'Enseignement, de l'Éducation et de la Formation »

Mention second degré

Mémoire

Didactisation de l'indexation automatique à travers le fonctionnement des moteurs de recherche

Étude des encarts informationnels ou *infoboxes*

Mémoire présenté en vue de l'obtention du grade de master

soutenu par

Élodie Renvoisé

le 18 juin 2020

en présence de la commission de soutenance composée de :

Xavier Brilland, directeur de mémoire

Pascal Duplessis, membre de la commission

Remerciements

Ces deux ans de travail ont nécessité du temps, de l'énergie et du remue-méninge. Mais rien de tout cela ne serait possible sans l'entourage professionnel, amical et familial dont j'ai la chance de bénéficier.

Tout d'abord je souhaite remercier chaleureusement Xavier Brilland pour son acuité scientifique, son professionnalisme et sa disponibilité pour nos échanges toujours fructueux. Dans ce type d'épreuve au long cours, il est bon de se sentir soutenue à toute épreuve. Je le remercie également d'avoir créé un groupe de travail au sein des réunions de bassin du Mans qui m'a permis d'avancer sur la notion d'indexation automatique.

Je tiens à remercier les formateur.trice.s de l'INSPE d'Angers de la section Documentation. Cette formation d'excellence m'a permis d'accéder et de réussir le concours, mais aussi de comprendre les enjeux de cette profession. À tous et toutes, un grand merci. Une pensée particulière se tourne vers Pascal Duplessis qui, au cours de ces deux années, a toujours su garder beaucoup de bienveillance dans ses exigences. Une leçon que j'espère garder avec mes futurs élèves.

L'entourage professionnel est encore plus précieux lorsqu'il est pourvoyeur d'émulation intellectuelle. Pour cela, je tenais à remercier le groupe de travail sur l'indexation automatique du bassin du Mans : Chrystèle Cissé, Larissa Allard, Isabelle Magadur, Valérie Caniart et Jean-Noël Amato.

Comme tout travail de recherche, un mémoire de Master 2 ne peut se faire sans le concours de collaborateurs aussi efficaces que talentueux. C'est le cas de Jean-Noël Amato, étudiant en Master 1 de Documentation, qui a repensé la mise en page de l'infographie pour la rendre plus "fun". Pari réussi. C'est le cas aussi de mon amie Adeline Gauthier, qui a su appliquer ses talents d'observation et de précision. Enfin, c'est également le cas de mon amie Valérie Caniart qui, grâce à son expertise de la langue française et de sa typologie, a vérifié toutes les petites erreurs qui se nichaient dans le manuscrit. Ce mémoire n'aurait pas été de cette qualité sans vous.

Je souhaitais tout particulièrement remercier les 46 professeur.e.s documentalistes ayant répondu au questionnaire, d'avoir pris de leur temps pour rendre ce travail de recherche possible.

Une formation crée des liens entre les personnes, et d'autant plus quand elle est intense comme celle-ci. C'est pour cela que j'ai une pensée pour Jessica, Elsa, Lucie, Sandy et Charlotte. Mille merci à Jennyfer et Berangère pour m'avoir suggéré des exemples d'infoboxes de définition. Merci aussi à Marion pour la discussion que nous avons eue à propos de la didactique. J'ai vraiment apprécié cet échange. Il arrive parfois que ces liens se transforment en amitié. Je voulais chaleureusement remercier pour leur soutien, Julie, Hélène, Adeline et Sandrine.

Certes ces deux années sont intenses pour la personne qui les vit, mais aussi pour l'entourage familial. De plus, la période de confinement a nécessité de repenser l'équilibre familial entre travail et vie de famille. Pour cela, je voudrais remercier affectueusement mon mari Olivier de m'avoir épaulée pendant ces deux ans et d'avoir assumé la vie quotidienne. Je voudrais remercier ma fille Helmie également pour avoir accepté, de façon passagère, que sa maman passe plus de temps à travailler qu'à jouer avec elle.

À ma fille, Helmie

*Notre cerveau possède, dès la naissance, un talent que les meilleurs logiciels d'intelligence artificielle ne parviennent pas à imiter :
la faculté d'apprendre.*

Stanislas Dehaene, *Apprendre ! Les talents du cerveau, le défi des machines*, 2018

Sommaire

INTRODUCTION.....	12
1 APPROCHE ÉPISTÉMOLOGIQUE ET DIDACTIQUE DE L'INDEXATION AUTOMATIQUE.....	16
1.1 DE L'INDEXATION AUTOMATIQUE AU WEB SÉMANTIQUE : CHANGEMENT DE CADRE CONCEPTUEL POUR LES MOTEURS DE RECHERCHE	16
1.1.1 DÉFINITIONS ET FONCTIONNEMENT D'UN MOTEUR DE RECHERCHE.....	16
1.1.2 HISTOIRE DU DÉVELOPPEMENT DES MOTEURS DE RECHERCHE.....	19
1.1.3 ÉVOLUTION DE L'INDEXATION AUTOMATIQUE DANS LES MOTEURS DE RECHERCHE.....	21
<i>UNE BRÈVE HISTOIRE D'INDEXATION.....</i>	<i>21</i>
<i>FONCTIONNEMENT DE L'INDEXATION AUTOMATIQUE DANS UN MOTEUR DE RECHERCHE.....</i>	<i>24</i>
1.1.4 ÉVOLUTION DE L'INDEXATION AUTOMATIQUE DANS LE WEB DE DONNÉES ET LE WEB SÉMANTIQUE. .28	
<i>WEB DE DONNÉES OU WEB SÉMANTIQUE ? UNE QUESTION DE DÉFINITION.....</i>	<i>28</i>
<i>FONCTIONNEMENT DE L'INDEXATION AUTOMATIQUE AU SEIN DU WEB SÉMANTIQUE.....</i>	<i>30</i>
<i>LE TRIPLET RDF.....</i>	<i>31</i>
<i>LE LANGAGE SPARQL.....</i>	<i>32</i>
<i>L'ONTOLOGIE.....</i>	<i>32</i>
1.2 UN CAS CONCRET : LES « INFOBOXES ET AUTRES ENCARTS INFORMATIONNELS – ENTRE INDEXATION AUTOMATIQUE, AGRÉGATION D'INFORMATION ET/OU WEB SÉMANTIQUE ?.....	34
1.2.1 RÉCENTE HISTOIRE D'INFOBOXES.....	34
1.2.2 LA RECHERCHE AGRÉGÉE : UNE « CONDENSATION » DE L'INFORMATION DANS LA PAGE DE RÉSULTATS.....	36
1.2.3 LES « ENCARTS DÉFINITIONNELS » : UN NOUVEAU TYPE D'INFOBOX, D'AGRÉGATION DE CONTENU OU AUTRE ?.....	37
1.3 LA DIDACTISATION DE L'INDEXATION AUTOMATIQUE : UN ENJEU PÉDAGOGIQUE POUR LES ÉLÈVES ET LES PROFESSEUR.ES DOCUMENTALISTES.....	41
1.3.1 DIDACTIQUE ET TRANSPOSITION DIDACTIQUE EN INFORMATION-DOCUMENTATION.....	41
1.3.2 LES MÉTHODES PÉDAGOGIQUES ET PROGRESSIONS EN INFORMATION-DOCUMENTATION.....	43
<i>LA PÉDAGOGIE EN INFORMATION-DOCUMENTATION.....</i>	<i>43</i>
<i>PROGRESSIONS LINÉAIRE ET SPIRALAIRE.....</i>	<i>45</i>
1.3.3 L'INDEXATION AUTOMATIQUE DANS LE SECONDAIRE VUE PAR L'INSTITUTION.....	46
<i>CYCLE 3 : (CM1-CM2) 6E.....</i>	<i>46</i>
<i>CYCLE 4 : 5E – 4E – 3E.....</i>	<i>46</i>
<i>LYCÉE : SCIENCE NUMÉRIQUE ET TECHNOLOGIE (SNT) ET SPÉCIALITÉ MANAGEMENT, SCIENCES DE GESTION ET NUMÉRIQUE.....</i>	<i>47</i>
<i>COLLÈGE-LYCÉE : CADRE DE RÉFÉRENCE DES COMPÉTENCES NUMÉRIQUES (CRCN).....</i>	<i>49</i>
1.3.4 L'INDEXATION AUTOMATIQUE VUE PAR LES PROFESSEUR.E.S DOCUMENTALISTES.....	50
<i>CYCLE 3 : (CM1-CM2) 6E.....</i>	<i>50</i>
<i>CYCLE 4 : 5E – 4E – 3E.....</i>	<i>51</i>
<i>LYCÉE : SCIENCE NUMÉRIQUE ET TECHNOLOGIE (SNT).....</i>	<i>51</i>
1.3.5 DIDACTISATION DE L'INDEXATION AUTOMATIQUE.....	52

<i>DIDACTISER L'INDEXATION AUTOMATIQUE OU L'INDEXATION AUTOMATISÉE ?</i>	52
<i>DIDACTISER L'INDEXATION AUTOMATIQUE</i>	53
1.4 CADRE CONCEPTUEL ET HYPOTHÈSES DE RECHERCHE	63
1.4.1 CADRE CONCEPTUEL DE RECHERCHE.....	63
<i>CONCEPTS SCIENTIFIQUES RETENUS</i>	63
<i>PROBLÉMATISATION ET PROBLÉMATIQUE</i>	65
1.4.2 HYPOTHÈSE DE RECHERCHE.....	67
2 PROJET DE RECHERCHE	72
2.1 MÉTHODOLOGIE DE COLLECTE DE DONNÉES	72
2.1.1 ÉCHANTILLONNAGE ET CRITÈRES DE SÉLECTION.....	72
2.1.2 PRÉSENTATION DU QUESTIONNAIRE.....	73
2.2 MÉTHODOLOGIE DE TRAITEMENT DES DONNÉES	77
2.2.1 TYPES DE DONNÉES.....	77
2.2.2 PRÉSENTATION DES DONNÉES.....	78
3 RÉSULTATS ET DISCUSSION	82
3.1 RÉSULTATS	82
3.1.1 CADRE GÉNÉRAL DES RÉSULTATS.....	82
3.1.2 RÉPONSE AUX HYPOTHÈSES.....	84
3.2 DISCUSSION DES RÉSULTATS	100
3.3 PROPOSITION DE PROGRESSION PÉDAGOGIQUE DE L'ENSEIGNEMENT DE L'INDEXATION AUTOMATIQUE À PARTIR DES INFOBOXES	105
3.4 CONCLUSION	108
3.5 DISCUSSION SUR LES BIAIS DE L'ÉTUDE	109
3.5.1 POINT DE VUE CRITIQUE.....	109
3.5.2 OUVERTURES ET ENJEUX PROFESSIONNELS.....	111
3.6 CONCLUSION GÉNÉRALE	112
BIBLIOGRAPHIE	115
ANNEXES	121
ANNEXE 1 : EXEMPLES D'INFOBOXES DE DÉFINITION	122
ANNEXE 2 : RÉCOLTE DES SÉANCES PÉDAGOGIQUES DE L'INDEXATION AUTOMATIQUE	124
ANNEXE 3 : QUESTIONNAIRE – INDEXATION AUTOMATIQUE – PROFESSEUR.E.S DOCUMENTALISTES	125
ANNEXE 4 : DONNÉES BRUTES DU QUESTIONNAIRE ET CATÉGORISATION DES VERBATIMS	127
ANNEXE 5 : TRAITEMENT ANALYTIQUE DES DONNÉES	153

Introduction

Au sein de la culture numérique, 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 aux internautes l'accès à l'information. Le *Web sémantique* a même pour objectif de donner une réponse à une requête et non une liste de résultats. 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. L'indexation automatique a une origine beaucoup plus ancienne que les moteurs de recherche. Elle a été inventée à la fin des années 1950 par un chercheur en informatique, Hans Peter Luhn, pour permettre à un ordinateur de reconnaître une chaîne de caractères dans un texte (El Hachani, 1997). Historiquement et conceptuellement, c'est en partie grâce à l'invention de l'indexation automatique que les sciences informatiques et les sciences de l'information se sont rejointes (Timimi & Kovacs, 2006). Dans les années 1970, les premières utilisations de l'indexation automatique étaient à des fins documentaires, pour retrouver un document dans une base de données. A partir du moment où les moteurs de recherche se sont emparés de cette technologie, elle n'a cessé d'être améliorée au point, elle aussi, d'avoir recours à l'apprentissage ordinateur (*machine learning*).

A 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 sont progressivement entrées 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). Cet enseignement est mis en avant car les pratiques adolescentes, et adultes, montraient, et montrent encore, qu'à plus de 90%, les recherches informationnelles s'effectuent via ce moteur de recherche (Duffez, 2020). C'est en 2013 que la notion de « moteur de recherche » entre dans le Wikinotions (voir historique : Reynaud, 2020). L'indexation automatique est

alors enseignée principalement à travers l'étude de la page de résultats (Sogliuzzo, 2013).

Depuis plusieurs années maintenant, 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, ou bien des cartes géographiques ou bien encore, 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. Nous nous interrogeons sur l'impact, pour le développement de la culture de l'information-documentation des élèves, de l'enseignement de l'indexation automatique à travers les *infoboxes*. Est-ce un objet intéressant pour comprendre le fonctionnement des moteurs de recherche via l'indexation automatique, et par extension, est-ce une entrée dans la culture numérique pour les élèves ? Est-ce que les professeur.e.s documentalistes se sont emparé.e.s de la question de l'enseignement de l'indexation automatique, et si oui, le font-il.elle.s à partir des *infoboxes* ? Si les processus de l'indexation automatique ont suivi les dernières avancées du *Web*, est-il nécessaire de revoir la didactisation de cette notion ?

Afin de comprendre les enjeux que peut engendrer l'étude des encarts informationnels, nous nous demandons si les *infoboxes* constituent un nouvel objet d'étude permettant la didactisation des processus d'indexation automatique par les professeur.e.s documentalistes, et une meilleure compréhension par les élèves, du traitement de l'information par les moteurs de recherche. Nous nous attacherons dans un premier temps à relever les changements de paradigme qu'entraîne l'évolution du *Web* et les conséquences sur la didactisation de la notion d'indexation automatique. Après une présentation épistémologique et historique des moteurs de recherche, nous mettrons en évidence les évolutions de l'indexation automatique en lien avec les évolutions des moteurs de recherche et du *Web*. Nous exposerons ensuite le cas concret des encarts informationnels, ou *infoboxes*, afin de caractériser leur mode de fonctionnement. Nous poursuivrons avec les enjeux didactique et pédagogique de la notion d'indexation automatique. Nous terminerons cette première partie par la présentation des hypothèses de recherche issues du cadre théorique. Nous présenterons ensuite le cadre méthodologique dans lequel nous avons effectué notre étude auprès des professeur.e.s documentalistes. Les résultats de cette étude et leur discussion termineront ce travail de recherche.

Partie 1 :

**Approche épistémologique et didactique de
l'indexation automatique**

1 Approche épistémologique et didactique de l'indexation automatique

1.1 De l'indexation automatique au Web sémantique : changement de cadre conceptuel pour les moteurs de recherche

1.1.1 Définitions et fonctionnement d'un moteur de recherche

Dans son mémoire de master 2 en 2015, le professeur documentaliste Julien Rousseau s'est penché sur la question des définitions d'un moteur de recherche (Rousseau, 2015). Celle qu'il a retenue est issue des travaux de Mesguich et Thomas en 2010 : « *un moteur de recherche est un outil destiné à collecter et indexer un grand nombre de pages dans l'objectif de permettre, via les interfaces web, une interrogation par mots-clés sur le contenu de ces pages* ». J. Rousseau, même s'il avait retenu cette définition, déplorait qu'elle ne prenne pas en compte la notion de classement. La définition de Brigitte Simmonot, professeure au Centre de recherche sur les médiations (CREM), en 2012, aborde cette notion en intégrant toutes les caractéristiques d'un moteur de recherche : « *un moteur de recherche web est une application qui collecte automatiquement les publications en ligne, les indexe de manière automatique et, via une interface souvent simplifiée, met en œuvre des algorithmes d'appariement et de classement des résultats en fonction des requêtes des internautes* » (Simmonot, 2012).

Ces définitions étant antérieures au *Web de données* et au *Web sémantique*, il faudrait ajouter, dans les fonctionnalités d'un moteur de recherche, la capacité à agréger des contenus de pages web, qui se surajoute à l'indexation automatique (Krichen et al., 2012).

La compréhension du fonctionnement des moteurs de recherche a été facilitée par la mise à disposition d'algorithmes rendus publics, ainsi qu'à partir de l'histoire du *Web*. Aujourd'hui, seuls les logiciels libres donnent accès à leurs algorithmes, comme le moteur de recherche *Qwant* par exemple. Parfois, les brevets déposés par les compagnies opératrices permettent d'en savoir plus, mais sans ces algorithmes, la seule façon de déterminer leur fonctionnement est de

remonter le fil à partir de la page de résultats à travers des expérimentations réalisées par des individus (Simmonot, 2012). Cependant, Dominique Cardon, professeur de sociologie à Sciences Po Paris, considère, que même si nous avons la possibilité d'avoir accès au contenu précis des algorithmes, cela ne permettrait pas de connaître l'étendue des conséquences de leur fonctionnement (Cardon, 2019). Même si le fonctionnement des moteurs de recherche a beaucoup évolué depuis leur création, une constante de quatre étapes subsiste à partir de la requête : la collecte des données, l'indexation des documents, l'appariement et le classement des résultats, et enfin la page de résultats (ou interface de consultation) (Simmonot, 2012).

- **La collecte des données** repose sur un protocole principal, le *http (hypertext transfer protocol)*. Il permet d'interroger un ou plusieurs serveur.s sur le.s.quel.s une copie du document est stockée (Simmonot, 2012). Une fois le.les serveur.s interrogé.s, un robot de collecte (nommé *spider* ou *crawler*) va explorer des milliards de pages Web à partir d'un ensemble d'adresses URL qui renvoient à d'autres pages au travers des liens hypertextes (Rousseau, 2015 ; Simmonot, 2012).
- Selon Brigitte Simmonot (2012), « ***l'indexation des documents*** est le processus qui consiste à caractériser les documents pour permettre de les retrouver, en représentant leurs notions caractéristiques par des termes ou des indices. » Bien que manuelle au début du fonctionnement du Web, cette indexation est rapidement devenue automatique. Aujourd'hui, deux méthodes complémentaires permettent d'indexer les documents : l'indexation en texte brut, basée sur la reconnaissance de la chaîne de caractères, et l'indexation structurée à l'aide de métadonnées (Delestre & Malandain, 2017). L'exhaustivité de la méthode en texte brut permet d'accéder à une grande partie du contenu du document, tandis que l'utilisation des métadonnées permet de gagner en précision (Delestre & Malandain, 2017).
- **L'appariement et le classement des résultats** pour une requête constituent le cœur du moteur de recherche (Simmonot, 2012), non seulement au niveau de la précision par rapport à la requête des utilisateurs, mais aussi au niveau économique. En effet, lorsque le moteur

de recherche produit une liste de résultats, « à partir de la consultation de l'index d'une base de données constituée par exploration successive du Web » (Mesguich & Thomas, 2010), deux catégories de résultats apparaissent : « les résultats éditoriaux ou résultats « naturels » et les liens commerciaux » (Simmonot, 2015).

- **L'interface de consultation** contient à la fois la zone de requête et la page de résultats. Cela correspond à la « partie émergée de l'iceberg » d'un moteur de recherche comme le décrit Brigitte Simmonot en 2012. En fonction de son besoin d'information, l'utilisateur va formuler une requête dans la barre d'interrogation du moteur de recherche (Rousseau, 2015). Les opérateurs booléens (ET – OU – SAUF), issus des premiers systèmes de recherche d'information dans les années 1970 sont encore utilisés de façon implicite dans la barre de recherche. Mais aujourd'hui, au moment même où l'utilisateur formule sa requête, des suggestions lui sont proposées en fonction de la fréquence des associations de mots-clés, mais aussi en fonction des traces laissées par l'internaute ou sur la popularité des termes, expression et requêtes des autres internautes. Le moteur de recherche ne se contente plus de chercher une adéquation de chaînes de caractères entre une requête et des milliards de pages web, il « cherche », de surcroît, à devancer le besoin d'information de l'utilisateur. L'ergonomie de la page de résultats a également beaucoup changé depuis les années 1990 et reflète les évolutions des algorithmes liés à l'appariement et au classement des résultats. Elle est différente pour chaque moteur de recherche et reflète « l'état d'esprit » dans lequel les concepteurs du moteur de recherche souhaitent faire apparaître les résultats. Encore une fois, la présentation antérieure avec Titre, URL, « snippet » (assemblage de fragments de texte) existe toujours, mais elle est souvent sous ou à côté d'une « infobox »¹, ou bien d'une barre d'agrégation de données de type images, vidéos, cartes, etc. (Krichen et al., 2012).

Comme nous le verrons dans la partie consacrée à l'histoire des moteurs de recherche, ces derniers ont évolué de façon synchrone avec les nouvelles fonctionnalités du Web, et les évolutions du Web de documents au Web de

¹« Une infobox (ou infoboîte) est une table de données présentant sommairement des informations importantes sur un sujet. Elle prend la forme d'un cartouche ou d'un encadré, placé en général en haut à droite de l'article. Elle n'est pas obligatoire. » (Wikipédia, 2019)

données vers un Web sémantique (Delestre & Malandain, 2017). Chacune des étapes est contrôlée par un ou plusieurs algorithmes qui reprennent les schémas anciens et en ajoutent de nouveaux. Cependant le fonctionnement de ces algorithmes n'est pas parfait et constitue le reflet de la vision du monde de leurs concepteurs. Même si aujourd'hui, l'utilisation du « *machine learning* », (l'apprentissage machine) et du « *deep learning* », (l'apprentissage profond, qui se rapproche du cerveau humain) permet une automatisation de plus en plus avancée et prédictive des procédures des algorithmes des moteurs de recherche, il n'en demeure pas moins que l'un des biais majeurs observé dans la pratique des internautes, et notamment des élèves, est d'attendre une réponse précise à une question, et non des informations à synthétiser soi-même. Le principe de synthèse par soi-même des informations plutôt que de se contenter d'une unique réponse serait pourtant meilleur pour les apprentissages. C'est ce que l'on appelle l'engagement actif en neuropédagogie (Dehaene, 2018).

1.1.2 Histoire du développement des moteurs de recherche

Le développement des moteurs de recherches est intimement lié à l'avènement du *Word Wide Web* en 1990 de Tim Berners-Lee (Simmonot, 2012). En effet, le Web, qui est une partie d'Internet regroupant tous les sites du réseau mondial et reposant sur le principe des liens hypertextes (Rey, 2016b), constitue une masse croissante de documents numériques, pour laquelle se pose rapidement le problème de l'accès à l'information. De cette problématique, naît la nécessité de classer l'information, de l'indexer et d'y donner accès. Comment à partir d'une requête, retrouver une liste des pages web les plus pertinentes possibles ?

Dès les années 1970, la question de la pertinence se pose mais uniquement au niveau du document lui-même. Les logiciels de traitement documentaire de certaines bibliothèques américaines ciblaient les documents les plus pertinents à partir d'une requête (Bellier & Estéoule, 1988). L'indexation des documents était manuelle à cette époque et si le document renvoyé n'était pas suffisamment pertinent, le document n'était pas affiché. La doctrine était alors « *du silence plutôt que du bruit* ». À l'inverse, avec le premier moteur de recherche sur le Web (*Archie* en 1990 ; Simmonot, 2012), même si l'indexation était manuelle, la pertinence était élargie à « *du bruit, plutôt que du silence* » et cela a tout changé, non seulement dans la façon de construire les algorithmes des moteurs de

recherche, mais aussi dans l'utilisation de ces moteurs de recherche par les internautes.

Depuis le premier moteur de recherche, *Archie*, créé en 1990 par Alan Emtage, qui proposait un système manuel d'indexation des documents, tous les moteurs de recherches ont cherché à automatiser toutes les étapes de la requête jusqu'au classement de la page de résultats à l'aide de robots de collecte, et ce dès l'apparition du *Web* en 1991 (Simmonot, 2012). Le système d'annuaire qui sélectionne, valide et classe les sites de façon manuelle comme *Yahoo !* connaîtra assez vite ses limites face à l'explosion documentaire du *Web* (Rousseau, 2015). Cependant, l'évolution des algorithmes des moteurs de recherche se produit par couches successives, à l'image d'un oignon. Ainsi, le premier moteur procédant à une indexation plein texte des pages web, *Webcrawler*, créé en 1994 par Brian Pinkerton, alors qu'il était étudiant à l'université de Washington, a repris une partie de l'algorithme du moteur de recherche *Worm*, développé par Olivier McBrayan à l'université du Colorado, qui permettait d'indexer uniquement les titres des pages et les adresses URL (*Uniform Resource Locator* ; Simmonot, 2012).

Au moment de la mise en application du moteur de recherche *Google* en 1998, créé par Sergey Brin et Larry Page à l'université de Stanford, il existait de nombreux moteurs basés principalement sur la concordance des mots-clés et leur fréquence d'apparition dans les sites (Simmonot, 2012). L'algorithme *PageRank* du moteur de recherche *Google* est basé sur un procédé de classement des pages de résultats prenant en compte le nombre de liens externes pointant vers les pages pour en estimer leur importance (Simmonot, 2012). C'est le principe de la popularité (Cardon, 2015).

Le modèle initial du développement des moteurs de recherche par des étudiants de l'université est aujourd'hui révolu. Selon Brigitte Simmonot (2012), « *Les progrès les plus marquants réalisés [par les moteurs de recherche] sont relatifs à leur portée (la taille de leur index), la diversité des documents auxquels ils donnent accès, ainsi qu'à l'ergonomie des interfaces et aux fonctionnalités à valeur ajoutée qu'ils proposent.* ». Aujourd'hui, nous voyons l'évolution des fonctionnalités de l'algorithme *PageRank* depuis la requête jusqu'à la page de résultats, qui tente de devancer le besoin d'information des internautes. Un nouvel algorithme, *RankBrain*, est basé sur un procédé d'intelligence artificielle, le

machine learning. « L'objectif est de générer des informations structurées et sémantisées liées à différents types de requêtes » (Ertzscheid, 2017). À travers les moteurs de recherche et leurs pages de résultats nous voyons émerger le *Web de données* dont l'objectif est de mettre en ligne, en plus des documents, des données structurées « interprétables » par des programmes informatiques (Delestre & Malandain, 2017). Le *Web sémantique*, imaginé aussi par Tim Berners-Lee et ses collaborateurs en 2001, va plus loin que le *Web de données* et repose sur des procédés d'apprentissage automatique (*machine learning*) afin que les machines soient capables de trouver du sens dans le contenu des pages web (Berners-Lee, 2001; Cardon, 2019). Autrement dit, il s'agit que les machines « comprennent » le contenu des pages web et soient capables de reformuler une réponse à une requête. Comme l'écrivait Brigitte Simmonot en 2012 : « *Le moteur [Google] tend ainsi à se poser comme l'unique source d'information qui permettrait aux internautes de se passer de consulter les documents primaires.* » Pour ce faire, l'évolution des procédés algorithmiques liés à l'indexation automatique a été déterminante.

1.1.3 Evolution de l'indexation automatique dans les moteurs de recherche

- **Une brève histoire d'indexation...**

L'indexation, et notamment l'indexation automatique, a été au cœur même du fonctionnement des premiers moteurs de recherche dès les années 1990 (Simmonot, 2012). Cependant, l'indexation et son ancêtre, l'index, sont des procédés documentaires bien plus anciens et qui, par la suite sont devenus des domaines scientifiques de la linguistique et de la documentation (Amar, 2000; Timimi & Kovacs, 2006). L'indexation automatique fait le lien entre le domaine de l'informatique et des sciences de l'information notamment grâce à l'algorithme d'indexation de l'informaticien Hans-Peter Luhn dans les années 1950.

L'index est un préalable à l'indexation, qui, au XVII^e siècle, est vu comme « *un indice ou répertoire qu'on met à la fin ou au commencement d'un livre, pour le soulagement du lecteur, afin qu'il **trouve facilement** les endroits dont il aura besoin* » (Kovacs & Dibie, 2006). On retrouve cette idée facilitatrice de l'index dans une définition plus récente issue du dictionnaire de l'information de Cacaly et al. (2004), où il « sert à **faciliter** la recherche d'information dans un document à

l'aide d'une liste ordonnée de mots qui sont référencés au sein de ce document » (Cacaly et al., 2004).

L'indexation a la même finalité que l'index, à savoir **faciliter** l'accès à une information d'un ou plusieurs documents, mais avec une dimension technique « à partir d'un sujet, ou d'une combinaison de sujets appartenant à un langage documentaire (ex : un thésaurus) ou au langage naturel » (Cacaly et al., 2004). L'indexation requiert « une démarche d'analyse qui peut être constituée d'un index, d'un résumé d'article ou d'ouvrage, etc. » (Cacaly et al., 2004). Une autre définition parue dans l'ouvrage d'Arlette Boulogne (2004), maître de conférences en sciences de l'information, ancienne directrice de l'Institut national des techniques de la documentation, reprend l'indexation avec une dimension pratique comme un « processus destiné à représenter, au moyen des termes ou indices d'un langage documentaire ou au moyen des éléments d'un langage libre, les notions caractéristiques du contenu d'un document ou d'une question, en vue d'en **faciliter** la recherche, après les avoir identifiées par l'analyse. » (Boulogne, 2004). De nouveau, la fonction et la finalité facilitatrices de l'indexation sont retrouvées, mais avec une insistance particulière sur la nécessité d'un langage contrôlé ou non. Dans la définition de Jacques Maniez, ancien maître de conférences en sciences de l'information, la définition de l'indexation reprend la vision sémiotique de l'index : l'indexation est une « approche différentielle de la signification, (le sens d'un terme est précisé par celui des termes équivalents ou interdits). Les concepts du thésaurus sont spécialement définis à fin d'indexation à partir d'un fonds documentaire donné pour en **faciliter** l'interrogation ultérieure. » La définition de la norme AFNOR Z 47-102 de 1978, insiste également sur la nécessité d'un langage contrôlé, sans faire mention de la fonction facilitatrice de l'accès à l'information de l'indexation : « L'indexation est l'opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans ce document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse ». Dans le cadre d'une journée professionnelle des rencontres de l'URFIST² de Rennes, son directeur d'alors, Alexandre Serres, dégage l'ambiguïté qui réside dans les différentes définitions de l'indexation et notamment sur les questions d'objets d'indexation, de langage et de finalité. Il résume l'indexation à deux fonctions principales que sont :

²URFIST : Unité régionale de formation à l'information scientifique et technique

- L'indication de la teneur du document, c'est-à-dire son contenu
- Permettre une recherche efficace, c'est-à-dire l'accès au contenu.

Trois opérations sont nécessaires pour réaliser ces deux fonctions :

- L'opération de catégorisation, c'est-à-dire définir la cible de l'information (document, collection, etc.)
- L'opération d'anticipation permettant des traitements ultérieurs, c'est-à-dire anticiper les réponses potentielles à une requête
- L'opération de lexicalisation, c'est-à-dire quelle clé, quel langage, utiliser pour indexer un document, une collection, etc. (Par exemple, le langage contrôlé)

Les trois objets de l'indexation définis par Alexandre Serre vont se retrouver dans les définitions et les fonctions de l'indexation automatique (cible, anticipation et clé), cependant ces objets vont être de plus en plus sophistiqués algorithmiquement car liés à une interface informatique via les moteurs de recherche. Selon le travail de master de Mabrouka El Hachani à l'ENSSIB³ de Lyon (1997), l'indexation automatique est « *l'opération qui consiste à faire reconnaître par l'ordinateur des termes figurant dans le titre, le résumé, le texte complet et parfois même l'indexation humaine, et à employer ces termes, soit tels quels, soit après conversion en d'autres termes équivalents ou conceptuellement voisins, pour en faire des critères incorporés dans le fichier de recherche et utilisables pour retrouver le document* » (El Hachani, 1997). Il n'y a plus la notion facilitatrice de l'indexation précédente dans cette définition, de plus, elle ne prend pas encore en compte les métadonnées comme termes « *à faire reconnaître par l'ordinateur* ». En effet, pour que l'indexation automatique « *vise à améliorer la performance des recherches documentaires exprimées en termes de pertinence et d'exhaustivité* » (Cacaly et al., 2004), il est nécessaire qu'elle prenne en compte les métadonnées supplémentaires aux « *composantes classiques des documents analogiques (Titres, auteurs, année de publication, etc.)* » qu'ont les documents numériques (Broudoux, 2012). Pour Évelyne Broudoux, maître de conférences en sciences de l'information et de la communication au laboratoire DICEN-IDF⁴ (2012), « *le processus d'indexation consiste donc pour une part à poser les bases*

³ENSSIB : École nationale supérieure des sciences de l'information et des bibliothèques.

⁴DICEN-IDF : Dispositifs d'information et de communication à l'ère numérique (Paris).

d'une future contextualisation du document par les lecteurs, par une analyse de contenu et par les métadonnées existantes ». Étonnamment, depuis que l'indexation automatique existe, la notion même d'indexation a pu évoluer pour devenir : « *l'indexation peut être manuelle, elle consiste à associer un ensemble d'informations (métadonnées) à un document lors de son enregistrement dans la base de données.* » (Delestre & Malandain, 2017). Dans cette dernière définition purement techniciste, les facettes de l'indexation comme facilitatrice de l'accès au contenu du document et comme analyse du document sont occultées. D'après ces auteurs, même si l'indexation humaine procède d'une meilleure analyse et d'une meilleure expertise du contenu d'un document, elle n'en demeure pas moins variable en fonction de la personne qui indexe et du temps écoulé entre deux indexations d'un même document. L'indexation automatique, quant à elle, semble être la seule solution viable sur le long terme pour pallier la variabilité de l'indexation humaine. Cependant, les auteurs rappellent que le principal inconvénient de l'indexation automatique est qu'elle indexe les contenus sans interprétation et cela peut générer des complications notamment pour « *l'indexation de texte brut* »⁵ (*plain text*) qui nécessite précisément un accès au sens du contenu d'un document (Delestre & Malandain, 2017).

Nous retiendrons que l'indexation automatique est un processus d'analyse du contenu d'un document informatisée pour lequel l'accès au contenu du document est augmenté par un index de données et de métadonnées, et dont l'efficacité de la recherche est améliorée par des algorithmes toujours plus perfectionnés facilitant la catégorisation (cible), l'anticipation (requête) et la lexicalisation (clé : recherche en texte brut).

- **Fonctionnement de l'indexation automatique dans un moteur de recherche**

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, qui se complexifie aujourd'hui à l'heure des algorithmes qui favorisent l'apprentissage des ordinateurs par eux-mêmes (*machine learning*) (Delestre & Malandain, 2017).

⁵L'indexation en texte brut, peut faire référence à l'indexation automatisée définie par Olivier Le Deuff (2014), comme une indexation permettant le traitement de l'ensemble des mots des documents. Avec l'indexation automatisée, les techniques documentaires sont réinvesties dans les techniques informatiques pour le traitement « plain texte », ou texte brut des documents.

La première étape pour qu'un document soit indexé par un moteur de recherche est la **segmentation**, c'est-à-dire le découpage d'un document en entités sémantiques (Figure 1). Cela revient à découper une suite de caractères en mots qui ont un sens. C'est une étape complexe, dans le sens où le découpage des mots n'est pas simplement lié aux espaces entre les mots, mais aussi aux traits d'unions ou aux apostrophes. Par exemple, « rendez-vous » qui a un trait d'union, constitue un seul et même mot (Delestre & Malandain, 2017).

En second lieu vient la **normalisation** (Figure 1). Cette étape se déroule principalement au moment de la requête dans la barre de recherche du moteur. Cette étape permet de « normaliser » la représentation des mots pour un ordinateur (Delestre & Malandain, 2017). Par exemple, faire comprendre que c.o.d., cod et COD sont identiques. Il existe deux formes de normalisation :

- *La normalisation de forme textuelle*, c'est-à-dire la normalisation des abréviations (l'exemple du c.o.d.), le passage des majuscules en minuscules, la suppression des accents, etc.
- *La normalisation linguistique* permettant de transformer chaque mot en une forme « canonique » ou standard.
 - *La lemmatisation*, permet de supprimer les flexions des mots (ex. « mangera, mangions » pour le verbe « manger »). Pour obtenir les lemmes, le moteur d'indexation devra procéder à une analyse syntaxique lui permettant de trouver la fonction du mot dans la phrase et donc la catégorie syntaxique associée.
 - *La racinisation* permet de supprimer les suffixes d'un mot pour ne garder que le radical. Contrairement à un lemme, un radical ne correspond pas forcément à un mot. C'est le système le plus utilisé pour la recherche informationnelle, mais il est moins précis car basé sur un radical (Delestre & Malandain, 2017).

L'**analyse syntaxique**, la troisième étape, permet de trouver la fonction syntaxique de chaque mot dans une phrase (Figure 1). Elle est basée sur trois méthodes avec des complexités croissantes en termes algorithmiques. Premièrement les *méthodes probabilistes*, c'est-à-dire la probabilité qu'un mot soit suivi par un autre, qu'un mot d'une catégorie syntaxique soit suivi par un mot d'une autre catégorie syntaxique. Cette méthode est l'une des plus ancienne. Hans-

Peter Luhn déclara en 1958, à propos de son algorithme d'indexation automatique : « *Au lieu de tirer l'information au hasard comme le fait normalement le lecteur, la nouvelle méthode automatique choisit les phrases d'un article qui représentent le mieux l'information pertinente* » (El Hachani, 1997). Deuxièmement, les *méthodes à base de règles* qui permettent d'explicitier la construction des phrases par des règles grammaticales. Enfin, les *méthodes basées sur le fonctionnement des réseaux neuronaux* qui permettent un apprentissage artificiel de l'ordinateur (*machine learning*). L'idée principale est de disposer d'un très grand corpus de textes déjà annotés et d'injecter ces données (et métadonnées) dans un algorithme basé sur les réseaux de neurones, pour que l'ordinateur « apprenne » à reconnaître de nouveaux textes à partir de la consultation de ce corpus de textes. Par exemple, l'algorithme « SyntaxNet » du moteur de recherche Google est basé sur ce principe (Delestre & Malandain, 2017).

Enfin, pour la quatrième étape, le moteur de recherche procède à une **structuration et une pondération de l'information** pour indexer les documents (Figure 1). Cela consiste à identifier tous les termes de tous les documents et à associer à chaque terme une liste de documents le contenant : c'est le principe de l'index inversé. Chaque terme d'un document a un « poids » différent en fonction de la fréquence de ces termes dans le corpus. La pondération prend en compte : le nombre de termes dans le corpus, le poids du terme dans le document, sa fréquence, et le nombre de documents contenant ce terme. Plus certains mots sont souvent rencontrés en compagnie d'autres mots dans une phrase, plus ces mots ont un poids important (El Hachani, 1997). Le poids d'un terme est dépendant de la taille des documents. Il est nécessaire, par conséquent, de normaliser le poids du terme par la taille du document (Delestre & Malandain, 2017).

Étapes de l'INDEXATION AUTOMATIQUE dans un MOTEUR de RECHERCHE

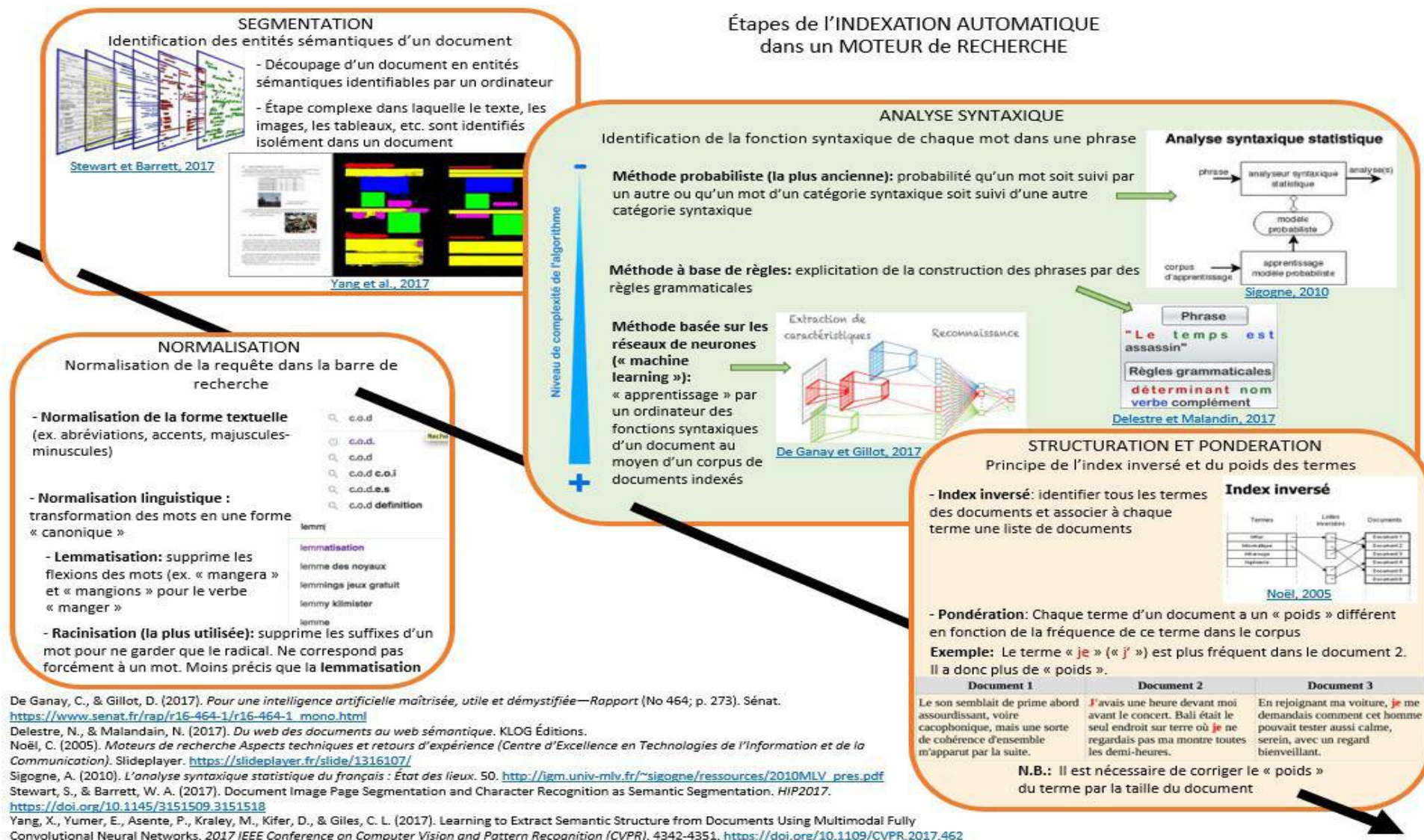


Figure 1 : Étapes de l'indexation automatique dans un moteur de recherche.

1.1.4 Évolution de l'indexation automatique dans le Web de données et le Web sémantique

- **Web de données ou Web sémantique ? Une question de définition**

Comme nous venons de le voir dans la partie précédente (1.1.3), la révolution majeure de l'indexation automatique réside dans la prise en compte des métadonnées : d'une part les métadonnées générées manuellement (ex. par un professionnel de la documentation) et les métadonnées générées automatiquement par les algorithmes de « lecture » des documents (Delestre & Malandain, 2017). La gestion des milliards de données et métadonnées issues de l'indexation automatique des documents a nécessité de faire interagir ces données afin que les moteurs de recherche commencent à « trouver du sens » parmi elles. L'objectif était d'améliorer de plus en plus la page de résultats au point d'arriver à l'idéal de Tim Berners-Lee *et al.* (2001), à savoir : « *The Semantic Web will bring structure to the meaningful content of the Web* » (Traduction personnelle : *Le Web sémantique va apporter une structure et du sens au contenu des pages Web*).

Le passage du *Web de documents* vers le *Web sémantique* se fait progressivement via le *Web de données* (ou *Web of Linked data*). Certains auteurs comme Olivier Le Deuff, maître de conférences au laboratoire Médiation, information et communication à l'université Bordeaux 3, considèrent que le *Web sémantique* est équivalent au *Web de données*. Il définit le *Web sémantique* comme un outil du *Web* permettant de « *séparer et d'organiser le contenu d'un document, afin de pouvoir ré-exploiter les métadonnées plus aisément* » (Le Deuff, 2014). A l'inverse, Muriel Amar, conservatrice des bibliothèques et chargée d'études à la Bibliothèque publique d'information, et Bruno Menon, maître de conférences au département des sciences de l'information-documentation de Paris 8, parlent en 2011 de *Web de données* au lieu de *Web sémantique* en argumentant que Tim-Berners Lee préférerait le terme de « *Web de données* » dans une publication de 2007 (Amar & Menon, 2011). Dans leur ouvrage publié en 2017, Nicolas Delestre et Nicolas Malandin, maîtres de conférences en informatique à l'INSA⁶ de Rouen, intitulé *Du Web des documents au Web sémantique*, expliquent qu'il existe une différence technique entre le *Web de données* et le *Web sémantique*. Alors que le *Web de données* est basé sur le

⁶INSA : Institut national des sciences appliquées.

déréférencement⁷ (accès à toutes les données) et l'interopérabilité⁸ des données (lisibles par toutes les interfaces), le *Web sémantique* est basé, de surcroît, sur les ontologies⁹ (l'accès au sens du contenu des pages *Web* par les algorithmes). Pour ces auteurs, le *Web de données* consiste à « *mettre en ligne des documents, des données structurées « interprétables » par des programmes informatiques* ». Dans sa thèse de doctorat en sciences et technologies de l'information et mathématiques, publiée en langue anglaise en 2015, Luis Daniel Ibáñez, insiste sur le fait que le *Web de données* permet la publication et surtout l'inter-relation entre des milliards de données (Ibáñez, 2015). Selon le dictionnaire de l'ENSSIB (2013), le *Web sémantique* « *vise à faciliter l'exploitation des données structurées pour donner du sens au contenu des pages Web en permettant leur interprétation par des machines. Il ne s'agit pas d'un web à part (comme le disait déjà Tim Berners-Lee et al. en 2001), mais plutôt d'une extension, d'une amélioration du Web courant où chaque donnée acquiert un sens défini, afin de créer un réseau d'informations structurées disponibles en ligne et **facilement** réutilisables et traçables* » (« Web sémantique », 2013). Dans cette définition du *Web sémantique*, nous retrouvons à la fois l'idée de facilité, que l'on trouvait déjà dans l'index du XVIIe siècle, mais aussi, la structuration et l'interprétation par les machines du *Web de données* et l'accès au sens du contenu des pages *Web* définissant le *Web sémantique*. Si nous revenons au sens originel du mot « sémantique » grâce au *Dictionnaire historique de la langue française* d'Alain Rey (2016a) :

- **Sémantique** : vient du grec *sêmantikos* « qui signifie, fait connaître », formé sur le verbe *sêmainen* « marquer d'un signe », « faire savoir, expliquer », lui-même dérivé de *sêma*, *sêmatos* « caractère distinctif ». En didactique : théorie visant à rendre compte des phénomènes signifiants dans le langage. Etude des significations de tout système de

⁷Déréférencement : permet d'obtenir un objet référencé par l'indexation, grâce à une requête HTTP. L'objet est constitué à la fois de la ressource web, et des données qui la décrivent. (Delestre et Malandin, 2017). ATTENTION ! Ne pas confondre avec le déréférencement d'un contenu dans un moteur de recherche qui renvoie au droit à l'oubli (<https://www.cnil.fr/fr/le-dereferencement-dun-contenu-dans-un-moteur-de-recherche>).

⁸Interopérabilité : standardisation des données et des techniques pour qu'elles deviennent utilisables et interprétables dans différents contextes du Web (Amar et Menon, 2011).

⁹Ontologie : en ingénierie des connaissances, une ontologie est une façon de modéliser un domaine en identifiant les concepts y afférant, en les organisant les uns par rapport aux autres, en identifiant les propriétés qui les relient et enfin en désignant les objets qui les représentent (Delestre et Malandin, 2017).

communication, en particulier d'une langue naturelle ou de l'un de ses éléments porteurs de sens.

Pour Delestre et Malandin (2017) : « *Le Web sémantique relève de la reformulation à partir du sens des mots et non plus de leur identification en tant que chaîne de caractères. Le Web 3.0 (ou Web sémantique) va permettre aux machines d'analyser le Web pour le « comprendre » et répondre à des questions plus complexes posées par l'homme. Les données de plusieurs sites seront agrégables par les machines. »*

A partir du moment où les documents sont indexés automatiquement puis que leurs données et métadonnées sont interreliées et structurées de telle sorte qu'elles « fassent du sens » pour un ou des algorithmes de moteur de recherche, nous pouvons dire que nous fonctionnons au sein du *Web sémantique*. Pour nous, le *Web de données* est inclus dans le *Web sémantique*.

- **Fonctionnement de l'indexation automatique au sein du Web sémantique**

Dans ses informations aux utilisateurs, l'entreprise Google explique succinctement le fonctionnement de son moteur de recherche :

« Beyond simple keyword matching, we use aggregated and anonymized interaction data to assess whether search results are relevant to queries. We transform that data into signals that help our machine-learned systems better estimate relevance. » « These relevance signals help Search algorithms assess whether a webpage contains an answer to your search query, rather than just repeating the same question »

(Traduction personnelle : Au-delà d'un simple appariement de mots-clés, nous utilisons des données en interactions agrégées et anonymes si les résultats de la recherche sont pertinents par rapport à la requête. Nous transformons ces données en signaux qui aident nos systèmes en apprentissage-machine (*machine-learned system*) afin de mieux estimer la pertinence. Ces signaux pertinents aident ensuite les algorithmes de « Search » à discriminer les réponses à la requête, des seules répétitions de la même question.)

Au sein de cette description nous pouvons relever les termes d'« agrégation de données en interaction », « pertinence par rapport à la requête » (ce qui est le

but premier d'un moteur de recherche), « transformation de données en signaux », « systèmes en apprentissage-machine », « amélioration de la pertinence », « discriminer les réponses à la requête ». Le fait de transformer des données en signaux, ainsi que d'utiliser un apprentissage-machine de ces signaux pour améliorer la pertinence ressemble à des procédés utilisés par le *Web sémantique* pour les moteurs de recherche. L'analyse syntaxique, troisième étape de l'indexation automatique (voir 1.1.3. : Evolution de l'indexation automatique dans les moteurs de recherche), est justement basée sur l'apprentissage-machine, ou apprentissage automatique (Cardon, 2019), afin de trouver la fonction syntaxique d'un mot dans une phrase (Delestre & Malandain, 2017).

Pour réaliser le potentiel du *Web sémantique*, le *World Wide Web Consortium* (W3C) a défini trois standards :

- 1) le triplet RDF (Resource Data Framework) pour la représentation de l'information dans le Web sémantique ;
- 2) le langage SPARQL pour questionner le Web sémantique ;
- 3) le OWL (Web Ontology Language) pour définir des représentations plus précises afin de prédire des aspects intrinsèques de la réalité (Montoya, 2016).

- **Le triplet RDF**

Le triplet RDF est formé d'un URI (*Uniform Resource Identifier*) d'un prédicat et d'une valeur (Figure 2).



Figure 2 : Schématisation d'un triplet RDF : URI, Prédicat, Valeur. Ce triplet représente l'information « Le prénom de <http://www.wikidata.org/entity/Q3309133> est Michel » (Delestre et Malandain, 2017).

Les URI permettent d'identifier toutes sortes d'entités nommées « ressources ». Ces ressources peuvent être un ensemble d'URL (*Uniform Resource Locator*), ou d'autres types de données au format standardisé du langage XML (*Extensible Markup Language*) représentant des nombres, des dates ou des durées, des données binaires pour les images et le son, des données

booléennes, etc. (Delestre & Malandain, 2017). Un URI est dit déréférencable lorsque son utilisation dans une requête HTTP (Hypertext Transfer Protocol) permet d'obtenir une ressource *Web* contenant des données qui la décrivent. (Delestre & Malandain, 2017). Un URI est différent d'un URL en ce sens que l'URI référence une entité, alors que l'URL référence les documents décrivant cette entité. Dans les types de données qui constituent l'entité, ou la ressource, on peut retrouver les données et métadonnées issues de l'indexation automatique.

Le triplet RDF renvoie donc une valeur (« Michel » dans l'exemple de la Figure 2), obtenue à partir d'un prédicat (par exemple « Prénom ») au sein d'une entité référencée par l'URI. Ce schéma peut être complexifié, si les triplets RDF sont composés de plusieurs URI, de plusieurs prédicats ou de plusieurs valeurs (Delestre & Malandain, 2017).

- **Le langage SPARQL**

Nous n'allons pas nous étendre sur le langage SPARQL car il ne fait pas appel à l'indexation automatique en tant que tel. Cependant, ce langage permet d'interroger la « gigantesque base de données » (Amar & Menon, 2011) où chaque ressource/entité est représentée par un URI. Il signifie Structured Protocol And RDF Query Language (SPARQL). C'est à la fois un langage d'interrogation et de modification des données. Ce langage est aussi un protocole de communication entre le client, le logiciel qui émet la requête, et le serveur, le logiciel qui traite la requête et retourne les résultats au client (Delestre & Malandain, 2017).

- **L'ontologie**

Enfin, l'ontologie est au cœur du *Web sémantique*. Le mot « ontologie » vient du grec *ontos* « être » et *logos* « étude ». Une ontologie est donc une étude de l'être (Rey, 2019). En philosophie, une ontologie est une étude des propriétés générales de l'être. Selon Delestre et Malandain (2017), en ingénierie des connaissances, l'ontologie est « *une façon de modéliser un domaine en identifiant les concepts y afférant, en les organisant les uns par rapport aux autres, en identifiant les propriétés qui les relient et enfin en désignant les objets qui les représentent* ». Pour Jacques Chaumier (2006) « *une ontologie fournit le vocabulaire spécifique à un domaine de connaissances et selon un degré de formalisation variable en fixant le sens des concepts et des relations qui les*

unissent ». Autrement dit, l'ontologie permet de coder du lien sémantique entre des objets, des mots, afin d'en identifier les concepts et d'en faire du sens, comme en biologie entre les protéines et les gènes qui les codent.

Il est important de distinguer le thésaurus, qui peut être un élément essentiel de l'indexation manuelle et automatique lors de l'utilisation de descripteurs (langage normalisé), de l'ontologie. « *Alors que les thésaurus organisent les termes d'un domaine, les ontologies organisent les concepts de ce domaine* » (Delestre & Malandain, 2017). Comme l'explique très justement Manuel Zacklad, professeur en sciences de l'information et de la communication au Conservatoire national des arts et métiers, en 2007, « *les ontologies sont organisées par thèmes, les thésaurus par concepts et les classifications par sujets. La finalité n'est pas pour une compréhension et une utilisation humaine, mais pour un langage machine* » (Zacklad, 2007). Ces deux définitions permettent de comprendre que l'organisation d'une ontologie a un degré conceptuel supérieur au thésaurus et n'est pas compréhensible par un humain.

Dans la partie 1.1.3 « *Évolution de l'indexation automatique dans les moteurs de recherche* », nous avons vu que le principal inconvénient de l'indexation automatique est qu'elle ne permet pas d'interpréter les contenus pouvant générer des problèmes pour l'indexation de texte brut (Delestre & Malandain, 2017). Avec les ontologies, l'indexation des objets est faite de telle sorte que le lien sémantique entre ces objets est « compris » et « appris » par une machine. Les ontologies permettent aux moteurs de recherche d'être plus performants dans la pertinence des réponses à une requête, même à partir du langage naturel. Pour bien comprendre à quoi servent les ontologies, nous pouvons citer la page Wikipédia (2019) sur l'ontologie informatique : « *L'ontologie est aux données ce que la grammaire est au langage* » (« Ontologie (informatique) », 2019).

Le langage informatique utilisé par le W3C pour exprimer les ontologies est le *Ontology Web Language* (OWL), qui veut dire langage ontologique du *Web*. Et « owl » en Anglais veut aussi dire « chouette » ou « hibou », ce qui permet à certains auteurs de faire des jeux de mots à partir de cet acronyme. « *OWL repose sur le Web de données liées, il utilise donc les URI pour désigner tous ces éléments.* » (Delestre & Malandain, 2017). Le Dublin Core est également une

ontologie reconnue par le W3C spécialisée pour « *décrire des documents de manière simple et standardisée* » (« Dublin Core », 2020).

Par conséquent, l'arrivée dans le *Web sémantique* permet aux moteurs de recherche d'aller plus loin que l'indexation automatique, même si elle est toujours présente, en interrogeant les données et métadonnées des documents du *Web*, en les inter-reliant et en générant du lien sémantique entre les chaînes de caractères grâce aux ontologies et à l'apprentissage-machine. L'apprentissage-machine est un procédé d'intelligence artificielle, mais c'est encore l'humain qui apprend à la machine, alors qu'avec l'apprentissage profond (ou *deep-learning*), la machine est capable d'apprendre par elle-même (Cardon, 2019). Cette dernière technologie est celle qui est stimulée actuellement par les compagnies développant les moteurs de recherche.

1.2 Un cas concret : les « infoboxes » et autres encarts informationnels – entre indexation automatique, agrégation d'information et/ou web sémantique ?

1.2.1 Récente histoire d'infoboxes

Le terme « *infobox* » a d'abord été associé à la structuration des formulaires Wikipédia. Toutes les recherches du terme « *infobox* » dans des moteurs de recherche spécialisés (Cairn, Persée, etc.) renvoient à des articles anglophones et francophones qui ont trait au travail sur l'encyclopédie collaborative Wikipédia (ex. Bissan, 2014; Bizer et al., 2009; Krichen et al., 2012).

Définition Wikipédia : « *Une infobox (ou infoboîte) est une table de données présentant sommairement des informations importantes sur un sujet. Elle prend la forme d'un cartouche ou d'un encadré, placé en général en haut à droite de l'article. Elle n'est pas obligatoire. Une infobox doit rester synthétique quant à son contenu. Elle ne doit pas remplacer l'article mais le compléter, fournir une vue globale, par l'apport d'informations générales. Pour chaque type d'article, un certain nombre de modèles existent ; ils sont normalisés par la communauté via les projets éditoriaux. Si une infobox standard ne convient pas à un article ou que plusieurs choix existent, le choix de l'infobox (modèles, contenus) passe par une discussion entre éditeurs pour trouver un consensus.* » (Aide:infobox in Wikipédia, Figure 3a). Comme l'explique la définition de Wikipédia, cette *infobox* est générée

à partir d'un modèle pré-rempli par un ou plusieurs auteurs. Cette *infobox* est à différencier de celles que l'on trouve sur la page de résultats des moteurs de recherche comme Google ou Bing par exemple. En effet, comme le montre l'exemple de la Figure 3b, les informations de l'*infobox* de Wikipédia s'y retrouvent, mais d'autres informations y sont ajoutées. Le procédé qui est utilisé pour générer cette seconde génération d'*infobox* est le « *linked-data* » via la base de données DBpedia par exemple (Bizer et al., 2009). Cette application du *Web de données* permet non seulement de récolter les données issues des *infoboxes* Wikipédia via le triplet RDF (Resource Description Framework), mais aussi de lier et d'enrichir son contenu avec plusieurs millions d'information contenues dans des centaines d'autres pages web pointant vers DBpedia (Bizer et al., 2009 et partie 1.1.4). Ainsi, des données textes structurées, mais aussi des images et/ou des vidéos, vont être associées aux données issues de l'*infobox* Wikipédia pour générer l'*infobox* de la page de résultats. Ces autres types d'*infobox* sont appelées « *Ontology infobox* », ou *infobox* d'ontologie (Barati et al., 2019). Dans ces *infoboxes* d'ontologie, une partie de l'information est sourcée, ici Wikipédia (Figure 3b) et une autre partie ne l'est pas, celle provenant des sites web pointant vers la source centralisante qui est DBpedia. Les *infoboxes* d'ontologie sont souvent situées en haut à droite de la page de résultats (ex. l'*infobox* d'ontologie « Chocolat » dans le moteur de recherche Google¹⁰).

Par conséquent, une *infobox* Wikipédia est un cartouche, ou un encadré, placé en général en haut à droite de la page 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. L'origine de l'information est manuelle, c'est-à-dire que l'auteur.e renseigne les informations importantes dans un gabarit pré-formaté. Lorsque l'*infobox*, contenant des informations issues de la page Wikipédia, se situe sur la page de résultats (souvent en haut à droite), des algorithmes vont aussi pointer vers d'autres pages du *Web* à partir d'une source centralisante et ajouter des informations supplémentaires. Nous sommes alors en présence d'« *infoboxes* d'ontologie » qui sont issues des données liées du *Web de données* (voir partie 1.1.4 ; Tableau 1).

¹⁰<https://www.google.com/search?client=firefox-b-e&q=chocolat>. Consultée le 03/01/2020



Figure 3a : *Infobox* originelle de Wikipédia où les informations sont générées par le.s auteur.e.s de la page. Cette *infobox* est modifiable par tout à chacun en cliquant sur « modifier »¹¹.

Figure 3b : Exemple d'*infobox* d'ontologie pour la requête Père Noël dans le moteur de recherche Google. *Infobox* générée par DBpedia par intercroisement de données dans le *Web de données*¹². Les données DBpedia ici, sont utilisées pour la récolte d'images.

1.2.2 La recherche agrégée : une « condensation » de l'information dans la page de résultats

Les interfaces d'agrégation d'information sont à différencier des *infoboxes* dans le procédé de récolte des informations (Tableau 1). Alors que les *infoboxes* de type Wikipédia sont renseignées par les auteurs et que les *infoboxes* d'ontologie sont récoltées de façon automatique par un processus de données liées (*linked-data*), l'agrégation de contenu, ou d'information, consiste à « rassembler des informations provenant de sources différentes, et à les présenter dans une seule interface » (Murdock & Lalmas, 2008). L'agrégation de

¹¹https://fr.wikipedia.org/wiki/Père_Noël – Consultée le 02/01/2020

¹²<https://www.google.com/search?client=firefox-b-e&q=p%C3%A8re+no%C3%ABl>.

Consultée le 02/01/2020

contenu s'effectue à partir de recherches agrégées qui permettent « *d'extraire et de regrouper automatiquement des granules d'information* » (Krichen et al., 2012) à partir de différents moteurs de recherche verticaux (voir Figure 4). Les deux seuls moteurs de recherche, à notre connaissance, à utiliser ce type d'agrégation de contenu actuellement sont Bing et Google. D'autres moteurs de recherche comme Qwant, Ecosia, Duckduck go ou Gigablast par exemples ont fait le choix de présenter ce type de contenu, non pas sous forme agrégée mais sous forme de facettes dans des onglets représentés au-dessus ou à la marge gauche de la page.



Figure 4 : Exemple d'interface à partir d'agrégation de contenu pour la requête Père Noël dans le moteur de recherche Bing.

1.2.3 Les « encarts définitionnels » : un nouveau type *d'infobox*, d'agrégation de contenu ou autre ?

Aujourd'hui nous voyons apparaître un autre type *d'infobox* ou plutôt d'encart d'agrégation de contenu, qui s'apparente à un dictionnaire pour le moment car seules des définitions y sont présentes (Tableau 1). Si nous prenons l'exemple de la requête « information », seule *l'infobox* d'ontologie issue de la ressource Wikipédia apparaît, alors qu'avec la requête « définition information », l'équivalent d'une page de dictionnaire apparaît (Figure 5a, b). Les deux exemples montrés ici sont issus des seuls moteurs de recherche, à notre connaissance, qui proposent un « encart définitionnel » Google (Figure 5a) et Bing (Figure 5b).

Dans un premier temps, il est frappant de voir que certaines définitions se ressemblent et d'autres diffèrent. Toutefois, ces différences sont déjà présentes

lorsque l'on consulte des dictionnaires papier. En revanche, il est plus intéressant de s'attacher à l'origine de ces définitions : d'où proviennent-elles ? Ainsi, la définition du terme « information » associée à la rubrique du « droit » du moteur de recherche Google (Figure 5a), s'inspire de sites qui ne font pas autorité dans le domaine du droit : un site pour un dictionnaire de langue français-arabe, un site de soutien scolaire en ligne pour étudiant et un lien vers la page d'un réseau social professionnel¹³. En ce qui concerne le moteur de recherche Bing, les différentes définitions ne sont pas référencées en fonction d'une discipline, comme le droit ou les sciences, comme sur le moteur de recherche Google, mais il semblerait que les définitions puissent être reformulées (Figure 5b). Cependant, il semblerait que ce ne soit pas toujours le cas (Voir autres exemples Annexe 1). En effet, la recherche plein texte de la première définition de l'encart de dictionnaire du moteur de recherche Bing (Figure 5b) renvoie aux mots isolément de la définition et non à la définition entière. Contrairement à la page de résultats du moteur de recherche Google, sur Bing, les premiers sites issus de la recherche plein texte font référence au CNRTL¹⁴, l'un des dictionnaires les plus complets faisant autorité sur le *Web*. A première vue, le moteur de recherche Google construit son dictionnaire à partir d'un système d'agrégation de contenu du type triplets RDF du *Web de données* (URI, prédicat, valeur, voir chapitre 1.1.4), alors que le moteur de recherche Bing procéderait parfois à une reformulation depuis d'autres dictionnaires avec un procédé qui ressemble aux ontologies du *Web sémantique* (voir chapitre 1.1.4 ; pour d'autres exemples du même type, voir l'Annexe 1).

¹³Pour rechercher les sources à l'origine de cette définition, nous avons effectué un copier-coller de cette définition en ajoutant des guillemets dans la barre de recherche du moteur de recherche Google.

¹⁴CNRTL : Centre national de ressources textuelles et lexicales

information

nom féminin

- DROIT**
Enquête pour établir la preuve d'une infraction, pour en découvrir les auteurs.
Ouvrir une information contre X.
- Renseignement** (sur qqch, sur qqch.)
Des informations confidentielles.
- Action de s'informer.
Une réunion d'information.
- Renseignement ou événement** qu'on porte à la connaissance d'une personne, d'un public.
Une information exclusive (anglisme scoop).
• **L'information action d'informer** le public, l'opinion (par les médias).
- SCIENCES**
Ce qui peut être transmis par un signal ou une combinaison de signaux (message) selon un code commun et par un canal ; ce qui est transmis (objet de connaissance, de mémoire).
Traitement informatique de l'information.

Traduire "information" en

Moins

Commentaires

information

[information]

NOM FEMININ

- Ensemble de renseignements sur quelqu'un, sur quelque chose.
"Recueillir de l'information sur un sujet." - "Des informations confidentielles ont été diffusées par erreur."
- Évènement porté à la connaissance d'un public.**
"Une information de dernière heure (et non un "scoop")."
- au pluriel*
Actualités radiodiffusées ou télévisées.
"Écouter le bulletin d'informations à la radio."
- Ensemble des activités qui ont pour objet la collecte, le traitement et la diffusion des nouvelles auprès du public.**
"Le traitement de l'information." - "«Il est de la responsabilité des pouvoirs publics de garantir les conditions économiques et d'indépendance des médias faute de quoi la liberté d'information et le droit à une information plurielle ne seraient que des concepts vides de toute réalité»"

Translate information to

Powered By Oxford Dictionaries - Corsica - Bing Translator

Feedback

Google

"Enquête pour établir la preuve d'une infraction, pour en découvrir les auteurs."

Tous Actualités Vidéos Images Shopping Plus Paramètres Outils

Environ 2 résultats (0,37 secondes)

Le Dictionnaire Majeur Des eleves - Francais-Francais-Arabe: ...
<https://books.google.fr/books> - Traduire cette page
 Hana Kabbani et Mirela Zade - 2008 - Langage Arts & Disciplines
 Enquête pour établir la preuve d'une infraction, pour en découvrir les auteurs. Ouvrir une information contre X. 2.Action de s'informer. Une réunion d'information.

EXPOSE - Commentaires Composés - 851 Mots - Etudier.com
<https://www.etudier.com> • Dissertations •
 L'information est une enquête pour établir la preuve d'une infraction, pour en découvrir les auteurs. 4)-LA TECHNOLOGIE DE L'INFORMATION DE LA ...

JEAN - BERTIN AYOMBE WA - AYOMBE - Responsable du ...
<https://www.linkedin.com> • ...
 View JEAN - BERTIN AYOMBE WA - AYOMBE'S profile on LinkedIn, the world's largest professional community. JEAN - BERTIN has 2 jobs listed on their profile.

Bing

"Ensemble de renseignements sur quelqu'un, sur quelque chose."

ALL IMAGES VIDEOS MAPS NEWS SHOPPING MY SAVES ABOUT

2 880 000 Results Date Language Region

RENSEIGNEMENT : Définition de RENSEIGNEMENT [Translate this page](#)
<https://www.cnrtl.fr/lexicographie/renseignement> •
 2. ART MILIT., POLICE. Ensemble de connaissances concernant l'ennemi, indispensables à l'état-major de la Défense nationale pour une action militaire efficace, à la police pour mener ses enquêtes et arrêter les délinquants.

RÉFÉRENCER : Définition de RÉFÉRENCER [Translate this page](#)
<https://www.cnrtl.fr/definition/referencer> •
 1. Si l'acte humain n'a pas une double référence, d'une part à l'histoire dans laquelle il s'insère et d'autre part, à une norme supérieure qui le juge, mais une seule, à savoir la première, il s'ensuit nécessairement qu'il ne peut tirer sa valeur propre que de son rôle historique.

CONCEPTS LE SOIN / L'ACCOMPAGNEMENT / LE ROLE DU ...
www.synora-commerce.fr/ch_bischoffwiller_extranet/Doc/Ressources/LE_SOIN... - PDF file
 CONCEPTS LE SOIN / L'ACCOMPAGNEMENT / LE ROLE DU SOIGNANT INTRODUCTION RAPPEL: Définition du métier L'aide-soignant exerce son activité sous la responsabilité de l'infirmier, dans le cadre du rôle propre

Figure 5a : Résultat de « l'ecart définitionnel » de la requête « définition information » pour le moteur de recherche Google ainsi que les 3 premiers résultats de la recherche *plein texte* de la première définition.

Figure 5b : Résultat de « l'ecart définitionnel » de la requête « définition information » pour le moteur de recherche Bing ainsi que les 3 premiers résultats de la recherche *plein texte* de la première définition.

Pour résumer, les quatre types *d'infoboxes* peuvent être décrites de la façon suivante :

	<i>Infobox Wikipédia</i>	<i>Infobox d'ontologie</i>	<i>Interface d'agrégation de contenu</i>	<i>Infobox type Dictionnaire</i>	
Description	→ Un cartouche/encart → En haut à droite de la page Wikipédia	→ Un cartouche/encart → En haut à droite de la page de résultats	→ Un ou plusieurs encarts → En haut, à droite ou entre les liens vers les sites de la page de résultats	→ Un cartouche/encart → Systématiquement en haut de la page de résultats	
Contenu	Synthèse de l'information /Informations générales	→ Synthèse ou premières lignes de la page Wikipédia → Autres informations provenant d'autre sites Web (texte, images, etc.).	Pas de synthèse de l'information	→ Extraction ou reformulation de l'information → Information structurée ou non en fonction des disciplines	
Fonction	→ Complète un article → Est toujours sourcé	→ Facilite l'accès à l'information → Seule l' information issue de la page Wikipédia est sourcée.	→ Met en valeur une sélection d'articles → Est toujours sourcé	→ Apporte des définitions à la manière d'un dictionnaire → Pas de source de l'information	
Origine de l'information	Information d'origine manuelle principalement	Information d'origine automatique exclusivement.	→ Information obtenue par extraction et regroupement automatique de portions (granules) d'informations → L'information provient de différents moteurs de recherche verticaux	<i>Google</i> → Information d'origine automatique → Les définitions sont structurées en fonction des disciplines → Les définitions proviennent de sites ne faisant pas toujours autorité (cf. <i>addendum</i>)	<i>Bing</i> → Information d'origine automatique → Souvent une seule définition / Pas ou peu de structuration des définitions → Les définitions proviennent de sites faisant en général autorité (cf. <i>addendum</i>)
Méthodologie	Les auteurs de la page Wikipédia ont la possibilité de compléter manuellement un gabarit d'infobox parmi un choix déjà existant.	Pour l'origine informatique : récolte d'information par centralisation des liens vers différents sites (DBpedia). => <i>Web de données</i>	Triplets RDF => <i>Web de données</i>	Ressemble à de l'agrégation de contenu structuré => <i>Web de données</i>	Ressemble à de la reformulation à partir de plusieurs dictionnaires => <i>Ontologie du Web sémantique ?</i>

Tableau 1 : Catégorisation des trois types *d'infoboxes* qui sont présentes sur une page de résultats, ensemble ou non, après une requête sur un moteur de recherche.

1.3 La didactisation de l'indexation automatique : un enjeu pédagogique pour les élèves et les professeur.e.s documentalistes

1.3.1 Didactique et transposition didactique en Information-documentation

Alors que la didactique est une science qui s'intéresse à la « *recherche attentive de la spécificité des contenus* » (Martinand, 2014), et « *leurs relations à l'enseignement et aux apprentissages* » (Reuter et al., 2013), la transposition didactique cherche à faire le lien entre les concepts scientifiques et les notions à enseigner aux élèves. Comme l'a défini Yves Chevallard et Marie-Alberte Joshua en 1991, la transposition didactique consiste en un « *travail qui d'un objet de savoir savant passe à un savoir à enseigner pour devenir un objet d'enseignement* » (Chevallard & Joshua, 1991). A cette transposition didactique des savoirs savants aux savoirs à enseigner, s'ajoute les « *pratiques sociales de référence ou PSR* » définies par Jean-Louis Martinand, comme l'intégration de la sphère professionnelle dans la conception des programmes (Martinand, 2014) C'est-à-dire comment les « *activités de production, des activités d'ingénierie, voire des activités domestiques, etc., peuvent servir de références à des activités scientifiques scolaires ?* » (Astolfi et al., 2008b).

Comme toutes les disciplines scolaires, même si elle n'est pas reconnue en tant que telle au niveau de l'Institution, l'information-documentation a une didactique et une transposition didactique qui lui sont propres. Cependant, elle répond à des règles communes à d'autres disciplines, notamment la didactique des sciences (ex. Astolfi et al., 2008a; Chevallard, 1985).

En didactique des sciences, la notion de didactique présente trois aspects (Astolfi et al., 2008a) :

- 1) Se centrer sur des champs conceptuels délimités
- 2) Le « savoir-savant » ne fournit pas de façon satisfaisante un contenu d'enseignement à appliquer tel quel en y apportant une simplification descendante. Il est nécessaire de diversifier le plus possible les pratiques sociales de référence.
- 3) Il y a besoin de créer des concepts nouveaux régulièrement.

Dans un entretien pour les éditions Éducagri en 2010, Jean-Pierre Astolfi a défini qu'une discipline « naît d'un changement de paradigme, qu'elle contient des concepts et des savoirs à enseigner » (Fleury & Billon, 2010). Alors que les concepts se rapportent à la discipline scientifique, les savoirs à enseigner se rapportent à la discipline scolaire.

La transposition didactique a des limites qu'il est important de prendre en compte (Schneuwly, 2008) : les savoirs scientifiques sont fragmentés en éléments (que l'on nommera ici caractéristiques), notamment dus à la séquentialisation des contenus d'enseignement, et les savoirs à enseigner sont transposés par l'enseignant en savoirs enseignés (c'est ce que l'on appelle la transposition didactique interne, cf. Duplessis, 2017), entraînant un changement de signification des concepts initiaux. De plus, la définition des contenus d'enseignement eux-mêmes comportent de nombreuses contraintes : l'état de l'art des connaissances scientifiques à un temps « t », les pratiques des élèves, les objectifs institutionnels, les compétences propres du corps enseignant et le développement cognitif, le désir d'apprendre des élèves et leurs représentations (Astolfi et al., 2008a). Ces limites peuvent être, en partie, réduites par un procédé de « contre-transposition didactique », défini par la chercheuse en didactique de l'information-documentation Muriel Frisch, qui consiste à utiliser les pratiques professionnelles des professeur.e.s (ici les professeur.e.s documentalistes), afin de remonter les expériences sur le « terrain » de l'apprentissage des notions par les élèves (Frisch, 2008).

En information-documentation, les concepts se rapportent à la discipline scientifique des sciences de l'information et de la communication et des sciences de l'information-documentation, alors que les savoirs à enseigner se réfèrent aux notions info-documentaires (Duplessis, 2017). Les pratiques sociales de références concernent les domaines de la bibliothèque, de la documentation, de l'édition, de l'informatique documentaire, du journalisme, de la communication d'entreprise, etc. (Duplessis, 2017). En l'absence de programme officiel, les professeur.e.s documentalistes disposent d'outils qui leur permettent de faire le travail de transposition didactique. Par exemple le programme « Vers un curriculum en information-documentation » publié par l'association des professeurs documentalistes de l'Éducation nationale (APDEN) en 2014, et sa plateforme collaborative de notions info-documentaires, Wikinotions Info-Doc

(Association des professeurs documentalistes de l'éducation nationale (APDEN), 2017). Cette plateforme contient actuellement 131 notions, classées par ordre alphabétique. Chaque notion est définie à partir des 10 domaines de l'information-documentation, associée à une ou plusieurs autres notions y faisant référence (Duplessis, 2016). Chaque notion est décomposée en caractéristiques au niveau débutant (souvent niveau Collège) et avancé (souvent niveau Lycée). Les caractéristiques, écrites sous forme de phrases verbales, servent directement en tant que telles pour définir la notion avec des niveaux de formulations attendus par les élèves à la fois aux niveaux débutant et avancé. Les définitions sont souvent suivies d'exemples et de contre-exemples, d'exemples de séquences pédagogiques, mais également de ressources scientifiques à partir desquelles la notion a été définie.

Il est important de retenir, que pour réaliser la transposition didactique d'une notion en information-documentation, il est nécessaire de s'appuyer sur des ressources scientifiques (transposition didactique), institutionnelles (quand elles existent) et professionnelles (pratiques sociales de références et séances pédagogiques) afin de définir les caractéristiques de cette notion ainsi que les niveaux de formulation attendus.

1.3.2 Les méthodes pédagogiques et progressions en information-documentation

- **La pédagogie en information-documentation**

« Les questions pédagogiques renvoient à trois éléments, les problèmes didactiques, les problèmes de normalisation, la question des apprentissages et de la progression. » (Gardiès, 2008). Alors que la pédagogie consiste à « définir des formes d'interventions [envers les élèves] situées dans une culture » (Gardiès, 2008), la didactique et sa transposition se réfèrent plutôt aux contenus de ces enseignements et au lien entre les savoirs et pratiques sociales de référence et les apprentissages (Frisch, 2012). Alors que les méthodes pédagogiques sont trans-disciplinaires, la didactique est toujours propre à une discipline. Ici, nous placerons les méthodes pédagogiques dans la culture de l'information au sein de la discipline de l'information-documentation.

Plusieurs méthodes pédagogiques existent, dont beaucoup sont utilisées en information-documentation. Elles dépendent de l'implication des élèves dans

leurs apprentissages et de la relation enseignant-élèves, par rapport au savoir. Plus l'élève est « acteur » de ses apprentissages, plus la méthode va être qualifiée d' « active ». A l'inverse, plus le cours est transmis de façon magistrale et descendante, plus la méthode sera qualifiée d' « expositive » (ou affirmative) (Duplessis, 2014). Sur son site professionnel, « Les Trois couronnes », Pascal Duplessis publie en 2014 un récapitulatif des méthodes pédagogiques utilisées en information-documentation. Trois méthodes pédagogiques sont répertoriées : expositive, interrogative et active. Alors que la méthode expositive a pour but de « *faire savoir* » ou de « *faire faire* », la méthode interrogative s'oriente plutôt dans le « *faire dire* » à l'élève et la méthode active a pour objectif de « *faire faire pour faire dire* » (Duplessis, 2014). Le rôle de l'enseignant dans la méthode expositive sera de dispenser un discours, un exposé, magistral ou une méthodologie, une procédure à reproduire. Dans ce cas le mode de pensée de l'élève est déductif, c'est-à-dire qu'il déduit, qu'il est « emmené » (Rey, 2019) dans le savoir mémorisé ou dans la procédure à reproduire (Duplessis, 2014). La méthode interrogative va conduire l'enseignant à interroger les élèves sur leurs représentations et faire avancer le cours pour présenter de manière guidée les savoirs aux élèves. Les élèves peuvent alors avoir plusieurs modes de pensée : déductive, inductive (remonter des cas concrets aux généralités) ou dialectique (raisonnement par le dialogue)(Duplessis, 2014). La méthode active, ou de découverte, amène l'enseignant.e à faire en sorte que les élèves posent le problème par eux-mêmes à partir de l'observation d'un ou plusieurs objets d'étude ou par une démarche d'investigation. Le rôle de l'élève sera de résoudre ce problème afin de conceptualiser la notion. Le mode de pensée de l'élève sera donc inductif et/ou hypothético-déductif (Duplessis, 2014).

Par conséquent, il existe un large champ de méthodes pédagogiques utilisables en information documentation. Les méthodes actives sont celles qui mettent le plus les élèves dans un processus de réflexion, lequel permet de partir d'un cas, d'un problème, et de progresser vers la généralité de la notion, et donc de son apprentissage. Cependant, il existe parfois certaines confusions parmi les professeur.e.s documentalistes dans la dénomination de ces méthodes et notamment la méthode active. Ce n'est pas parce que les élèves sont en « activité », qu'ils sont intellectuellement dans un mode inductif ou hypothético-déductif. Parmi les séances pédagogiques présentées dans l'Annexe 2 par exemple, aucune ne précise de façon explicite la méthode pédagogique

employée. Il est souvent nécessaire de faire l'analyse de la séance pour pouvoir la caractériser, comme cela est fait dans la base de données d'analyses de séances pédagogiques en information-documentation ID'Base¹⁵.

- **Progressions linéaire et spiralaire**

La progression pédagogique en elle-même « *ne relève pas que de critères scientifiques et didactiques relatifs à l'acquisition de contenus scolaires, mais d'un ensemble complexe de déterminations et d'enjeux (institutionnels, sociologiques, idéologiques, axiologiques)* » (Nonnon, 2010). Alain Mercier et Christian Buty (2004) citent Yves Chevillard pour parler de l'importance de la progression pour maintenir une cohérence didactique entre la discipline et son enseignement : « *La compréhension des systèmes d'enseignement modernes suppose d'identifier les lois de l'organisation de toute matière d'enseignement en progressions* » (Mercier et Buty, 2004). Deux types de progressions existent : la progression linéaire et la progression spiralaire.

En 1960, Jérôme Bruner introduit la notion de progression spiralaire permettant de qualifier ce que les didacticiens et pédagogues de l'époque préconisaient pour l'apprentissage, à savoir qu' « *apprendre est un processus continu qui suppose une reprise constante de ce qui est déjà acquis et une complexification progressive* » (Queval, 2010). Contrairement à la progression linéaire qui n'a pas pour objectif de revenir sur les notions apprises antérieurement, la progression spiralaire propose aux élèves de revoir une notion et d'y ajouter un niveau de complexité supplémentaire à chaque niveau d'enseignement (Queval, 2010). Cependant, en 1992, Jean-Pierre Astolfi, dans son ouvrage « *L'école pour apprendre* », met en garde sur le risque d'une progression circulaire et non d'une progression spiralaire au niveau des enseignements (Queval, 2010). En effet, il est nécessaire de mettre l'élève dans un niveau cognitif supérieur de la notion au fur et à mesure de la progression de l'élève. Dans le cas contraire, l'élève ne progresse pas et le sentiment de « tourner en rond » persiste et peut jouer sur sa motivation.

La progression spiralaire semble surtout utilisée en mathématiques et science physiques (Martin, 2016, Académie de Lyon), mais elle a toute sa place en information-documentation afin de reprendre les notions avec des niveaux de

¹⁵ID Base : Fiches pédagogiques à l'usage des professeurs documentalistes.
<http://idbase.esmeree.fr/table/methode>

complexité croissante. D'autant plus que ,dans le *Wikinotions* par exemple, différents niveaux de formulation sont proposés. Il est donc possible d'aborder une notion de manière spiralaire avec les élèves, pour peu que ces élèves puissent suivre de façon régulière, au cours de leur cursus, les enseignements délivrés par les professeur.e.s documentalistes.

1.3.3 L'indexation automatique dans le secondaire vue par l'Institution

- **Cycle 3 : (CM1-CM2) 6^e**

La notion d'indexation, n'apparaît jamais en tant que telle dans les programmes du cycle 3. Pour le secondaire, le cycle 3 représente uniquement la classe de sixième. Selon l'analyse de Pascal Duplessis de « *l'emprise de l'information-documentation dans le cycle 3* » sur son site professionnel « *Les trois couronnes* », cette notion n'est pas répertoriée pour elle-même (Duplessis, 2019a). Cependant, selon les 10 domaines de l'information-documentation enseignés par les professeurs-documentalistes, définis par le professeur de Documentation à l'INSPE d'Angers, Pascal Duplessis, la notion d'indexation est à rapprocher des domaines liés aux « Centre de ressources », à la « Recherche d'information », à la « Veille », au « Moteur de recherche » (Duplessis, 2016). Dans l'analyse des notions info-documentaires du cycle 3, le domaine de la « Recherche d'information » est identifié neuf fois, celui du « Centre de ressource » une seule fois dans les programmes et ceux de « Veille » et de « Moteurs de recherche » n'apparaissent pas. Par conséquent, la notion d'indexation et d'indexation automatique, peut y être abordée un certain nombre de fois dans les programmes, notamment à travers la recherche d'information en français principalement pour le cycle 3.

- **Cycle 4 : 5^e – 4^e – 3^e**

De même que pour le cycle 3, la notion d'indexation automatique, n'apparaît jamais en tant que telle dans les programmes du cycle 4. Selon l'analyse de Pascal Duplessis de « *l'emprise de l'information-documentation dans le cycle 4* » paru sur son blog professionnel en 2019, ni la notion d'indexation, ni celle d'indexation automatique ne sont évoquées. Les domaines info-documentaires présents dans ces programmes, qui peuvent être liés à l'indexation et à l'indexation automatique sont : les domaines de la « Recherche d'information » (six fois), de la « Veille » (une fois) et du « Moteur de recherche » (une fois). Le

domaine lié au « Centre de ressources » n'est pas évoqué. Nous voyons ici apparaître le domaine du « Moteur de recherche » qui, nous le verrons par la suite, est souvent associé à la notion d'indexation automatique. Cependant, comme nous le pouvons le voir ici, la notion même de moteur de recherche est très rarement abordée en tant que telle dans le cycle 4 dans les programmes institutionnels.

- **Lycée : Science numérique et technologie (SNT) et spécialité management, sciences de gestion et numérique**

Dans le programme de science numérique et technologie (SNT) du nouveau programme commun de seconde générale et technologique, publié au Bulletin Officiel spécial n°1 du 22 janvier 2019, le terme « indexation » apparaît 6 fois (Ministère de l'Éducation nationale et de la jeunesse, 17 janvier 2019). Cette notion est absente des autres programmes du Lycée général, mais pas du Lycée technologique où elle est présente dans le programme du cycle terminal de « management, sciences de gestion et numérique » en STMG¹⁶ (Ministère de l'Éducation nationale et de la jeunesse, 19 juillet 2019). A notre connaissance, la notion d'indexation n'apparaît dans aucun des programmes généraux du lycée professionnel. Même dans le module d'accompagnement « Algorithmique et programmation » du programme de mathématiques du niveau seconde professionnelle, la notion d'indexation n'est pas évoquée.

Dans le programme de SNT, il est fait état du fonctionnement des moteurs de recherche au sein du chapitre sur « *le Web* ». Plus précisément, la méthode de « *référencement naturel* » est décomposée en trois parties dont l'une consiste en « *l'analyse du contenu des pages Web et leur **indexation** sur les mots qu'elles contiennent (constitution d'un annuaire inversé qui associe chaque terme les URL des pages où il apparaît)* ». Parmi les exemples d'activités proposées dans ce programme de SNT, l'une consiste à « *Réaliser à la main **l'indexation** de quelques textes sur quelques mots puis choisir les textes correspondant à une requête* ». Au sein du chapitre sur « *les données structurées et leur traitement* », « ***l'indexation** des bases de données* » est évoquée dans la partie sur « *les repères historiques* » et spécifiquement pour les années 1970. Au sein de ce même chapitre, dans la partie « *les algorithmes et les programmes* », un paragraphe évoque l'indexation humaine et l'indexation automatique :

¹⁶STMG : sciences et technologie du management et de la gestion.

« La recherche dans des données structurées a d'abord été effectuée selon une **indexation** préalable faite par l'homme. Des algorithmes ont ensuite permis **d'automatiser l'indexation** à partir de textes, d'images ou de sons ».

Dans la partie « les machines », un paragraphe évoque une dissociation entre métadonnées et indexation lors de la recherche au sein des fichiers par les ordinateurs :

« Des recherches dans les fichiers se font à l'intérieur même des ordinateurs, soit sur la base de leurs métadonnées, soit sur la base d'une **indexation** (à la manière des moteurs de recherche sur le Web) ».

Or, d'après la partie 1.1.3 sur l'évolution de l'indexation automatique dans les moteurs de recherche, nous avons vu que l'indexation automatique se base également sur les métadonnées des documents¹⁷.

Dans le programme de spécialité « management, sciences de gestion et numérique » de la classe de terminale STMG, la notion d'indexation est liée au document avec les notions de numérisation et de structuration des documents dans un objectif d'échange d'informations afin de favoriser la communication. Dans le chapitre sur la structuration de l'information de ce programme, les métadonnées sont présentées comme un facteur favorisant l'indexation.

« Pour valoriser leur patrimoine informationnel, les organisations structurent la gestion de leur contenu, en y associant notamment des métadonnées qui favorisent son indexation ».

Ici, les métadonnées sont bien incluses dans un processus d'indexation. Par ailleurs, la distinction entre indexation humaine/manuelle et indexation automatique n'est pas précisée. L'indexation n'est pas liée au fonctionnement d'un moteur de recherche ou d'une base de données, mais à la structuration de l'information dans un document, c'est-à-dire, l'**indexation documentaire**.

Par conséquent, les notions d'**indexation** et d'**indexation automatique** sont pleinement inscrites dans le programme de SNT pour les élèves de seconde générale et dans un programme de spécialité de terminale technologique. Cette observation confirme la place que peuvent prendre les professeur.e.s

¹⁷Nous utilisons ici le terme de « document » et non de « fichier » car le fichier est associé à l'informatique alors que le document est associé à la documentation.

documentalistes au sein de ces enseignements avec l'apport de l'information-documentation.

Une enquête des Travaux Académiques Mutualisés (TraAM) en Documentation, intitulé « *Les professeurs documentalistes et l'enseignement de SNT* », publiée le 19 Mars 2020, montre que plus de la moitié (54,7%) des professeur.e.s documentalistes ayant répondu (439 réponses reçues) n'interviennent jamais en SNT (Chambaud et al., 2020). Près du tiers interviennent pour quelques classes et moins de 20 % interviennent pour toutes les classes. Les enseignements de SNT sont principalement réalisés par les professeurs de mathématiques, de physique-chimie et des sciences de l'ingénieur. Cependant, ceux-ci reconnaissent parfois leur manque de compétences dans certains domaines du programme, notamment celui lié aux réseaux sociaux. Or, c'est le thème le plus abordé par les professeur.e.s documentalistes ayant répondu (33,2%). Cette étude montre également, que ce sont principalement les professeur.e.s documentalistes disposant de séances déjà préparées qui ont eu la possibilité d'intervenir dans le cadre de ce nouveau programme.

- **Collège-Lycée : Cadre de référence des compétences numériques (CRCN)**

À la rentrée de Septembre 2020, va se mettre en place une nouvelle certification des compétences numériques qui sera effective pour les élèves de 3^e et de Terminale. Cette nouvelle certification remplace l'ancien B2i (Brevet informatique et internet), pour le collège et le lycée, et le C2i (Certificat informatique et internet) pour l'enseignement supérieur. Cette certification numérique, pourra être obtenue en utilisant la plateforme en ligne d'évaluation et de certification des compétences numériques des citoyens francophones, appelée PIX (pour pixel). Elle s'appuie sur le Cadre de référence des compétences numériques (CRCN) au niveau français, qui lui-même émane du DigComp (*Digital competencies*) au niveau européen.

Selon le Décret n°2019-919 du 30 Août 2019 relatif au « *développement des compétences numériques dans l'enseignement scolaire, dans l'enseignement supérieur et par la formation continue, et au cadre de référence des compétences numériques* », seize compétences numériques sont attendues, réparties dans cinq domaines de compétences. Chaque compétence contient huit niveaux de progression, dont les cinq premiers concernent l'enseignement secondaire (Ministère de l'Éducation nationale et de la jeunesse, 30 août 2019).

La notion spécifique d'indexation n'est pas présente au sein du CRCN. Cependant, parmi les cinq domaines de compétences, le domaine 1, relatif à « *Information et données* », et plus spécifiquement le sous-domaine 1.1 concernant « *Mener une recherche et une veille d'information* » requièrent des compétences qui peuvent être liées à l'indexation automatique. Par exemple, dans le domaine 1.1, il est spécifié de « *Mener une recherche et une veille d'information pour répondre à un besoin d'information et se tenir au courant de l'actualité d'un sujet tout en étant en mesure de vérifier les sources et la fiabilité de l'information (avec un moteur de recherche, au sein d'un réseau social, par abonnement de flux ou de lettres d'information, ou tout autre moyen).* » Plusieurs notions info-documentaires peuvent être décelées ici. En ce qui concerne l'indexation et l'indexation automatique, le fait de mener une recherche à l'aide d'un moteur de recherche peut nécessiter d'introduire la notion d'indexation automatique pour en comprendre le fonctionnement. De même, mener une recherche au sein d'un réseau social, peut faire émerger la notion d'indexation en langage naturel (notamment les «*# hashtags* » sur Tweeter) par opposition au langage contrôlé (par exemple, l'indexation par descripteurs sur un logiciel de gestion des bibliothèques, SIGB).

1.3.4 L'indexation automatique vue par les professeur.e.s documentalistes

- **Cycle 3 : (CM1-CM2) 6^e**

Parmi les quelques séances pédagogiques récoltées¹⁸ (Annexe 2) parlant d'indexation et/ou d'indexation automatique de façon explicite ou pour comprendre le fonctionnement d'un moteur de recherche, seule une séance était adressée aux élèves de 6^e. Dans cette séance, réalisée par Marie Guillet, professeure documentaliste dans l'Académie de Toulouse, l'objectif est de faire comprendre aux élèves la différence entre l'**indexation humaine** et l'**indexation automatique** (Guillet, 2013).

¹⁸Origine des séances récoltées : Moteur de recherche (Qwant – blogs professionnels), ID'Base (Base de données de séances pédagogiques en information-documentation), Edu'Base (Base de données de séances pédagogique dans toutes les disciplines, dont la documentation), Eduscol (site institutionnel à destination des professionnels de l'Éducation nationale).

- **Cycle 4 : 5^e – 4^e – 3^e**

Or, parmi les quelques séances pédagogiques sélectionnées abordant la notion d'indexation et/ou d'indexation automatique, trois sont réalisées au collège, dont deux au cycle 4 (Annexe 2). Sur le dictionnaire en ligne des notions info-documentaires édité par l'APDEN, une séquence pédagogique est proposée à destination d'élèves de 3^e leur permettant de différencier l'indexation de l'indexation documentaire. Sur l'espace pédagogique des professeurs documentalistes de l'Académie de Poitiers « Doc'Poitiers », la notion d'indexation automatique n'est pas traitée pour elle-même mais à travers le fonctionnement d'un moteur de recherche (Académie de Poitier « Comprendre le fonctionnement des moteurs de recherche », 2019). Cette séance, nommée « Comprendre le fonctionnement des moteurs de recherche... pour mieux choisir ! », et à destination des élèves de 5^e, permet d'analyser le fonctionnement de plusieurs moteurs de recherche et donc de comparer, quand cela est possible, leur système d'indexation.

- **Lycée : Science numérique et technologie (SNT)**

Avant la mise en place du programme de SNT, la notion d'**indexation** n'était pas du tout évoquée dans les programmes du lycée général. Cela n'a pas empêché les professeur.e.s documentalistes d'enseigner cette notion. Parmi les quelques séances récupérées sur différents sites ou blog publiant des séances pédagogiques (Annexe 2), une séance fondatrice publiée par Frédéric Rabat, professeur documentaliste de l'académie de Normandie, s'intéresse au fonctionnement du moteur de recherche Google pour la classe de seconde (Rabat, 2008). Dans cette séance, F. Rabat propose de découper le traitement de la requête, dans le moteur de recherche Google, en trois étapes, dont la deuxième est l'**indexation** (effectuée par un robot d'indexation). Il s'inspire des travaux scientifiques pour effectuer ce découpage. En prolongement, il semble intéressant pour F. Rabat de comparer le fonctionnement d'autres moteurs de recherche mais aussi l'**indexation** à partir d'un langage contrôlé (descripteurs) et celle à partir d'un langage naturel (ex. tag ou *folksonomie*¹⁹). Cette séance a fortement inspiré la professeure documentaliste Gaëlle Sogliuzzo de l'académie de Nantes qui publia sur son blog « *Culture de l'info et des médias en lycée* » une séance intitulée

¹⁹Définition de *folksonomie* : « Le terme de *folksonomie* est apparu récemment sur le web pour désigner le phénomène d'indexation des documents numériques par l'utilisateur. On rencontre également fréquemment la mot *tag* qui désigne en quelque sorte un mot-clé » (Le Deuff, 2006)

« Comment fonctionne un moteur de recherche : L'exemple de Google » (Sogliuzzo, 2013). Au sein de cette séance, la professeure intègre l'**indexation** au sein des éléments constitutifs d'un moteur de recherche en définissant un « *moteur d'indexation* ». Par conséquent, la notion d'**indexation automatique** était déjà surtout abordée en seconde pour le lycée avant 2019, dans le cadre de l'accompagnement personnalisé (AP) principalement (Annexe 2).

1.3.5 Didactisation de l'indexation automatique

- **Didactiser l'indexation automatique ou l'indexation automatisée ?**

Au sein des différentes séances pédagogiques, traitant de l'indexation, décrites dans la partie 1.3.2 « *l'indexation automatique dans le secondaire* » et dans l'Annexe 2, certaines employaient le terme **d'indexation** pour parler de l'**indexation automatique**, d'autres employaient le terme **d'indexation automatisée** pour se différencier de l'indexation automatique. La notion d'« *indexation automatisée* » est présente sur la plateforme professionnelle de notions info-documentaires, *Wikinotions* et dans le dossier « Vers un curriculum en information-documentation » publié par l'APDEN (Reynaud, 2017 ; APDEN, 2016a, b ; Tableau 2), ainsi que dans le mémoire de master recherche de Julien Rousseau sur la didactisation des moteurs de recherche (qui reprend la définition du Wikinotions, cf. Rousseau, 2015) et, enfin, dans l'ouvrage scientifique d'Olivier Le Deuff « *La documentation dans le numérique* » (Le Deuff, 2014). Or, il s'avère que le sens donné à l'indexation automatisée n'est pas le même entre les ressources professionnelles et l'ouvrage scientifique. Pour le chercheur en sciences de l'information et de la communication, Olivier Le Deuff, « *l'indexation automatisée paraît parfois plus efficace [que l'indexation automatique], car le traitement porte sur l'ensemble des mots des documents* » (Le Deuff, 2014). Autrement dit, l'**indexation automatisée** se rapprocherait d'une indexation automatique augmentée pour laquelle la difficulté majeure de l'accès au sens *via* l'indexation plein texte serait possible. Pour la plateforme professionnelle *Wikinotions*, le professeur documentaliste Florian Reynaud définit, pour son niveau de formulation avancé, l'**indexation automatisée** comme :

« Une opération informatique et automatique qui concerne des documents numériques en ligne. Elle suppose une méthode

logicielle qui conduit à la création d'un index. Cette indexation automatisée permet de retrouver facilement des données informationnelles parmi plusieurs documents, à l'aide d'un moteur de recherche.

La méthode logicielle de l'indexation automatisée s'appuie sur une programmation d'algorithmes qui prend en considération la fréquence des mots. Elle suppose des opérations diverses :

- l'extraction des mots vides (comme des mots fonctionnels, tels des articles ;*
- la prise en considération de la place des mots dans le document (titrairie, chapeau, casse...) ;*
- l'opération de stemming ou lemmatisation (effacement des terminaisons pour retrouver les racines des mots).*

L'indexation automatisée peut concerner des textes, mais aussi d'autres natures d'information, à savoir des vidéos, des documents sonores, et des images. Pour les images, il peut s'agir d'une indexation réalisée à partir de métadonnées liées à l'image, mais aussi à partir d'une reconnaissance du contenu graphique de l'image » (Reynaud, 2017 ; Tableau 2).

Dans cette définition, à part l'application exclusive à des documents numériques en ligne, les autres éléments renvoient *sensu stricto* au fonctionnement de l'indexation automatique tel que décrit dans la partie 1.1.3.

Par conséquent, afin de clarifier la notion d'indexation automatisée, il semblerait nécessaire de la remplacer par la notion originelle d'indexation automatique. Dans un niveau de formulation avancé, comme proposé dans le Wikinotions, il pourrait être envisagé d'apporter une remarque supplémentaire : « *l'indexation automatisée consiste en une indexation automatique « avancée » qui permet une indexation plein texte et l'accès au sens des mots d'un document* ».

- **Didactiser l'indexation automatique**

De ce fait, il convient de se demander comment didactiser la notion d'indexation automatique. Didactiser la notion d'**indexation automatique** n'est pas une question nouvelle, mais au regard de son évolution technique due à l'évolution du *Web* et surtout des moteurs de recherche, il est nécessaire de

compléter les définitions originelles et de revoir les niveaux de formulation pour les élèves de façon, peut-être, plus progressive. Parmi les séances pédagogiques évoquées précédemment dans ce chapitre (Annexe 2), il apparaît de façon quasi systématique que l'indexation automatique est rattachée au fonctionnement des moteurs de recherche. Dans son mémoire de Master 2, Julien Rousseau ajoute, en parlant de la notion de **moteur de recherche**, qu'on peut « *lui associer les notions de **base de données** et d'**indexation automatique** qui renvoient toutes les deux à deux autres notions organisatrices que sont l'**indexation** et l'**espace informationnel*** » (Rousseau, 2015). Il est évident que l'entrée par le fonctionnement des moteurs de recherche pour aborder la notion d'indexation automatique est pertinente.

L'APDEN, dans sa proposition de programme « Vers un curriculum en information-documentation » propose d'aborder la notion d'*indexation automatique* dès la classe de 5^e (APDEN, 2016a). En tant que notion associée, l'**indexation automatique** n'est pas abordée en tant que telle, mais en association avec les notions de **mots-clés** et de **moteur de recherche**. A partir de la classe de 4^e, et aussi en 3^e, l'APDEN préconise d'aborder la notion d'indexation en général, à travers les notions spécifiques d'**indexation documentaire** et d'**indexation automatisée** (APDEN, 2016a). L'**indexation** est associée au mot-clé comme « *opération intellectuelle* » pour répondre à un besoin d'information dans une recherche informationnelle. Au lycée, l'indexation est toujours associée aux mots-clés et aux moteurs de recherche avec une déclinaison plus fine des types d'**indexation** : **documentaire, automatisée et libre**. En 1^{re} et Terminale, l'indexation des moteurs de recherche est mise en parallèle de l'indexation humaine dans le catalogue du CDI (APDEN, 2016b). Cependant, d'autres enseignants, et notamment Marie Guillet, professeure documentaliste dans l'académie de Toulouse, propose de faire comprendre aux élèves d'une classe de 6^e la notion d'**indexation automatique**, par opposition à la notion d'**indexation humaine** (Guillet, 2013; Annexe 2). Dans cette même séance l'**indexation humaine** est abordée à partir d'un **espace informationnel** virtuel (comme le catalogue du portail documentaire en ligne) et l'**indexation automatique** à travers un **moteur de recherche**. Il est intéressant de noter que l'on parle ici d'**indexation humaine** et rarement d'**indexation manuelle** comme nous l'avons vu dans la partie sur le cadre définitionnel de l'indexation automatique. Cette notion

d'**indexation humaine**, par opposition à l'**indexation automatique**, est très présente dans les séances pédagogiques et dans les outils de transposition didactique des notions info-documentaires comme le Wikinotions. Il est probable que le terme d'**indexation humaine** soit plus parlant que celui d'indexation manuelle pour les élèves. Par ailleurs, l'**indexation humaine** peut être abordée par les nouvelles formes d'indexation de type *folksonomie* (tags, #tag, etc.) sur les réseaux sociaux. C'est ce qu'Hélène Mulot, sur son Blog « *l'Odysée d'LN : je tisse ma toile* » propose aux élèves de collège à travers le défi lecture sur le réseau social de partage de commentaires de livres « *Babelio* »²⁰(Mulot, 2013 ; Annexe 2). Sur *Wikinotions*, une séance consacrée aux « *Espaces et documents du CDI* » pour la classe de 3^e s'intéresse à la notion organisatrice d'**indexation** en parallèle de la notion d'**indexation documentaire** (Reynaud, 2014). Le niveau de formulation attendu par les élèves pour la notion d'**indexation documentaire** est la suivante : « *opération humaine de traitement intellectuel d'un document* ». Il est précisé de « *montrer qu'une indexation humaine est subjective en fonction d'une intentionnalité* ». Pour le chercheur en sciences de l'Information et de la communication, Olivier Le Deuff, « *l'intérêt pour les travaux réalisés en **indexation documentaire**, notamment en ce qui concerne les thésaurus, s'est accru, si bien que les systèmes d'**indexation informatique** travaillent de plus en plus de pair avec de l'indexation de type documentaire* » (Le Deuff, 2014). Ainsi, l'indexation documentaire peut être humaine, mais elle est liée à un langage contrôlé.

Dès lors, l'**indexation humaine** peut nécessiter un langage contrôlé (thésaurus), ou non (mots-clés ; *tags*). L'**indexation documentaire**, quant à elle, est systématiquement liée à un langage contrôlé. Elle est l'oeuvre de professionnels de l'information. L'**indexation automatique**, enfin, est obligatoirement informatique mais peut nécessiter des techniques d'indexation documentaire.

- **Didactiser l'indexation automatique par les pratiques sociales de référence**

Comme toute notion scolarisable, la notion d'**indexation automatique** s'appuie sur des concepts scientifiques, des programmes institutionnels et des ressources professionnelles (Tableau 3). Comme nous l'avons résumé dans la

²⁰Babelio : <https://www.babelio.com/>. Consulté le 17/04/2020.

partie 1.1.3 « *Evolution de l'indexation automatique dans les moteurs de recherche* », l'indexation automatique est « *un processus d'analyse du contenu d'un document informatisé pour lequel l'accès au contenu du document est augmenté par un index de données et de métadonnées, et dont l'efficacité de la recherche est améliorée par des algorithmes toujours plus perfectionnés facilitant la catégorisation (cible), l'anticipation (requête) et la lexicalisation (clé : recherche en texte brut)* ». Les quatre étapes qui régissent le fonctionnement de l'indexation automatique au sein des moteurs de recherche sont : « *la segmentation, la normalisation, l'analyse syntaxique et la structuration et pondération* » (Delestre et Malandin, 2017).

Pour le milieu professionnel des bibliothèques, **l'indexation** renvoie à un outil pour les bibliothécaires permettant l'accès et la description précise et concise du sujet traité dans les documents (Association des bibliothécaires de France, 2013). En documentation, **l'indexation automatique** se différencie de **l'indexation humaine** en présentant des avantages en termes de rapidité pour certains types de documentation, comme celle de la presse (Accart, 2015). Différents logiciels d'indexation automatique sont d'ailleurs utilisés par les professionnels de la documentation. Ils travaillent principalement à partir du texte intégral (c'est-à-dire, la lecture plein texte) et en constituant un index à partir de tous les termes (c'est-à-dire le principe de l'index inversé²¹). Différents types d'analyse sont alors combinées par ces logiciels : morphologique, syntaxique et sémantique (Accart, 2015). Aujourd'hui, les nouveaux systèmes d'indexation automatique des professionnels de la documentation intègrent l'apprentissage automatique (*machine learning*) (Suominen, 2019). Un nouveau logiciel, *Annif*, développé à la bibliothèque nationale de Finlande, fonctionne comme la plupart des systèmes d'indexation automatique, à savoir une « préparation » du document, en fragmentant les phrases et les mots (c'est la segmentation), en normalisant les mots des documents (conversion en minuscule, retraitage des mots vides, lemmatisation des déclinaisons, gestion des pluriels, etc.). Ensuite, un index inversé est généré sur tous les documents, de façon à comparer les « mots-sujets » utilisés, ou bien à les comparer avec les termes d'un langage contrôlé. Les mots les plus « pertinents » sont ensuite sélectionnés parmi la liste de mots générés, préalablement classés. Deux types d'approches complémentaires sont utilisées

²¹cf. Partie 1.1.3 Fonctionnement de l'indexation automatique

par les algorithmes d'indexation automatique pour la sélection des « mots-sujets candidats » pour l'indexation des documents : l'approche lexicale et l'approche associative (Suominen, 2019). Alors que l'approche lexicale est basée sur la fréquence des termes (c'est la pondération) et leur comparaison avec un langage contrôlé, l'approche associative se base sur l'apprentissage machine (ou *machine learning*) qui trouve des corrélations entre les mots du document et l'ensemble des mots-sujets, en exploitant un grand nombre de données (Suominen, 2019). Par conséquent, l'indexation automatique, pour les professionnels de la documentation, à pour objectif de **faciliter l'accès aux documents** par les usagers à partir de requête par mots-sujets, ce que développent les logiciels actuels de gestion des bibliothèques (SIGB). On retrouve les étapes de **segmentation du document**, de la lecture **plein texte** et de la **normalisation des mots du document** lui-même et non nécessairement de la requête (ce qui diffère des moteurs de recherche). Ces logiciels, restent toutefois extrêmement liés à **un langage contrôlé**, même dans la version récente du logiciel *Annif*, pour l'approche lexicale de l'indexation des mots-sujets d'un document. Cependant, l'association de cette approche avec la technique associative, permet aussi de se rapprocher de l'indexation automatique à partir d'un **langage non contrôlé**. Elle demande une technologie plus avancée qu'est l'apprentissage machine.

Dans le monde de l'entreprise, « l'extraction automatique de connaissances » constitue un enjeu majeur pour la performance des organisations (Soulier, 2005) Pour cela, l'**indexation automatique** des documents est un outil essentiel (Balmisse, 2015). En effet, « *pour être performant, un service de recherche d'information doit être en mesure de comprendre et surtout de maîtriser les contenus qu'il rend accessible* » (Balmisse, 2015). L'**indexation automatique** « *a pour but de représenter la connaissance tout en conservant toutes les nuances et toutes les subtilités que peut exprimer le langage naturel sans perte d'information et de connaissance* » (Planté, 2009). L'extraction de connaissances « *aura pour rôle de procéder à l'enrichissement sémantique des contenus avec des métadonnées de manière totalement automatique* » (Balmisse, 2015). Les techniques utilisées par les entreprises pour extraire l'information des documents sont très similaires à celles des moteurs de recherche actuels et les entreprises s'en inspirent grandement (Balmisse, 2015). L'analyse de **segmentation** reste basée sur le **mot** ou la **phrase**, sur le « **poids** » **des mots** dans les documents,

c'est-à-dire sur leur fréquence (Plantié, 2009). L'analyse syntaxique est peu utilisée. Cependant des modèles de **comparaison des documents** entre eux sont très développés par une analyse vectorielle (Plantié, 2009), ce qui est moins le cas des moteurs de recherche par exemple. La réduction de la redondance de l'information touche à la fois les entreprises et les professionnels de l'information (Accart, 2015), c'est pour cela que, malgré l'automatisation de l'indexation, certaines parties de l'indexation restent manuelles car plus précises (Balmissse, 2015).

- **Didactique de l'indexation automatique dans les programmes scolaires**

Pour l'Institution, dans le programme de SNT de la classe de seconde, **l'indexation** [automatique] est liée au fonctionnement des moteurs de recherche et participe au référencement naturel des pages Web : « *La méthode de recherche appelée **référencement naturel** se décompose en trois grandes activités, réalisées par les moteurs de recherche : (1) le parcours automatique du Web pour collecter les pages visitées (aspiration des pages Web effectuée par des robots) ; (2) l'analyse du contenu des pages et leur **indexation** sur les mots qu'elles contiennent (constitution d'un **annuaire inversé** qui associe à chaque terme les URL des pages où il apparaît) ; (3) la troisième activité, réalisée à chaque fois qu'un internaute fait une requête, construit une liste ordonnée des pages (classement) comportant les mots clés de la requête. Leur ordre dépend notamment de leur popularité (principe des liens), de leur pertinence (aux mots de la requête), et de l'ordre des termes de la requête* ». Nous pouvons voir ici que les robots évoqués, sont des robots d'indexation, qui permettent de générer un index issu de documents stockés sur des serveurs. L'analyse du contenu est différenciée de l'indexation automatique, or, c'en est une finalité. De plus, parmi les quatre étapes de fonctionnement de l'indexation automatique, seule celle concernant l'annuaire inversé est évoquée. Cependant, ce n'est qu'une sous-étape de la dernière étape concernant la structuration et la pondération des mots dans un document (Figure 1). La requête, ainsi que l'importance de l'ordre des mots dans cette requête sont abordées, mais leur importance dans l'étape de normalisation n'est pas précisée. En outre, toujours dans le programme de SNT, dans la partie « *Les données structurées et leur traitement* », il est rappelé que « *Des recherches dans les fichiers se font à l'intérieur même des ordinateurs, soit sur la base de leurs **métadonnées**, soit sur la base d'une **indexation** (à la*

manière des moteurs de recherche sur le Web) ». Là encore, la notion d'indexation est opposée à celle de métadonnées. Notons que dans la définition scientifique, l'analyse des métadonnées fait partie intégrante de l'indexation automatique. Ce choix institutionnel de réduire l'indexation automatique à quelques éléments de sa définition, voire de l'opposer à certaines de ses caractéristiques, pose la question de sa didactisation pour les élèves.

Nous avons vu précédemment les niveaux de formulation attendus pour la notion d'indexation automatisée pour la plateforme collaborative de notions info-documentaires, *Wikinotions* (Reynaud, 2017). Nous avons vu également que cette notion est identique à la notion d'indexation automatique, à ceci près qu'elle était spécifique aux documents numériques. Si nous reprenons dans leur intégralité les caractéristiques de cette notion pour les deux niveaux nous obtenons (Tableau 2) :

Niveau débutant	Niveau avancé
<ul style="list-style-type: none"> . Consiste en une opération informatique et automatique . Concerne des documents numériques en ligne . Suppose une méthode logicielle . Conduit à la création d'un index . Permet de retrouver facilement des données informationnelles : <ul style="list-style-type: none"> → parmi plusieurs documents → à l'aide d'un moteur de recherche 	<ul style="list-style-type: none"> . S'appuie sur une programmation d'algorithmes . Prend en considération la fréquence des mots . Suppose des opérations diverses : <ul style="list-style-type: none"> → L'extraction des mots vides (comme des mots fonctionnels, tels des articles) → La prise en considération de la place des mots dans le document (titrairie, chapeau, casse...) → L'opération de <i>stemming</i> ou lemmatisation (effacement des terminaisons pour retrouver la racine des mots) → Peut concerner : des textes, des images, des documents sonores, de la vidéo
Tableau 2 : Caractéristiques de la notion d'indexation <i>automatisée</i> selon le Wikinotions (Reynaud, 2017).	

Pour le niveau débutant, la caractéristique « *Consiste en une opération informatique et automatique* » semble redondante. En effet, l'automatisme nécessite, par essence un programme informatique. Pour « *Concerne des documents numériques en ligne* », il est possible d'ajouter « **ou numérisés** », car les documents manuscrits numérisés sont indexés de façon automatique aujourd'hui²². Pour le niveau avancé, la caractéristique « *S'appuie sur une programmation d'algorithmes* » est là encore redondante. Les algorithmes, pour être effectifs, nécessitent un langage de programmation. Alors que l'algorithme est une suite de procédures mathématiques suivant une logique, le langage de

²²cf. la segmentation de l'infographie sur l'indexation automatique ; Figure 1 ; Tableau 4

programmation en est l'application pour un logiciel particulier. Par conséquent cette caractéristique pourrait devenir « **S'appuie sur plusieurs algorithmes** » (Tableau 4). Il semble important d'apporter la caractéristique suivante en haut de la liste : « **Analyse de façon informatisée le contenu intellectuel d'un document** » (Tableau 4), l'analyse de contenu étant le propre de l'indexation initialement (voir le premier point de la partie 1.1.3, « *Une brève histoire d'indexation* »). De plus, par rapport au concept scientifique, et pour rappeler la caractéristique du niveau débutant « *Conduit à la création d'un index* », une formulation telle que « **Génère et exploite un index de données et de métadonnées** » (Tableau 4) semble essentielle à ajouter dans la liste des caractéristiques pour cette notion. La caractéristique « *Prend en considération la fréquence des mots* » ressemble aux opérations diverses décrites dans les caractéristiques citées ensuite, car elle correspond à la quatrième étape de l'indexation automatique : structuration et pondération (Figure 1). En effet, dans le niveau avancé, l'évocation des caractéristiques concernant les étapes de l'indexation automatique est pertinente. Il convient de les actualiser et de les regrouper. Cela pourrait devenir : « **Suppose quatre étapes : La segmentation (identification des éléments ayant du sens dans un document), la normalisation (normalisation des mots-clés dans la barre de recherche d'un moteur de recherche), l'analyse syntaxique (identification de la fonction syntaxique de chaque mot dans une phrase) et la structuration et pondération (principe de l'index inversé et du poids des mots)** » (Tableau 4). Il y a une confusion entre la lemmatisation et la racinisation dans la caractéristique suivante : « *L'opération de stemming ou lemmatisation (effacement des terminaisons pour retrouver la racine des mots)* ». En effet, l'effacement des terminaisons est la racinisation, alors que la lemmatisation consiste en l'effacement de la flexion des mots, mais c'est un détail vers lequel il est inutile d'aller pour comprendre l'indexation automatique avec les élèves (Delestre & Malandain, 2017). Enfin la caractéristique qui précise « *Peut concerner : des textes, des images, des documents sonores, de la vidéo* » peut être conservée (Tableaux 2 et 4).

L'INDEXATION AUTOMATIQUE du concept à la notion.

Définition scientifique	Définition institutionnelle	Définition professionnelle
Concept scientifique	Dans le programme de SNT-2nde	Notion à didactiser (Wikinotions, bibliothèques et entreprises)
<p><i>Processus d'analyse informatisée du contenu d'un document pour lequel l'accès au contenu du document est augmentée par un index de données et de métadonnées, et dont l'efficacité de la recherche est améliorée par des algorithmes toujours plus perfectionnés facilitant la catégorisation (cible), l'anticipation (requête) et la lexicalisation (clé : recherche en texte brut).</i></p> <p><i>4 étapes : segmentation, normalisation, analyse syntaxique, structuration et pondération.</i></p>	<p><i>Processus d'analyse du contenu des pages Web à travers l'indexation de leurs mots par la méthode de l'annuaire inversé.</i></p> <p><i>L'indexation automatique fait partie des trois étapes de référencement naturel des pages Web. L'indexation automatique est aussi présente lors d'une recherche dans une base de données.</i></p>	<p><i>Elle suppose une méthode logicielle qui conduit à la création d'un index. Cette indexation automatisée permet de retrouver facilement des mots normalisés parmi plusieurs documents en « lecture » plein texte, à l'aide d'un moteur de recherche, ou d'un logiciel professionnel de la documentation</i></p> <p><i>La méthode logicielle de l'indexation automatisée s'appuie sur une programmation d'algorithmes qui prend en considération la fréquence des mots, sa normalisation et la création d'un index inversé. Elle suppose des opérations diverses :</i></p> <ul style="list-style-type: none"> <i>- l'extraction des mots vides (comme des mots non fonctionnels, tels des articles).</i> <i>- la prise en considération de la place des mots dans le document (titraillle, chapeau, casse...), leur syntaxe et leur sémantique :</i> <i>- l'opération de stemming o u lemmatisation (effacement des terminaisons pour retrouver les racines des mots).</i> <p><i>L'indexation automatique est basée sur langage contrôlé et/ou libre</i></p> <p><i>L'indexation automatisée peut concerner tous les types de documents (texte, image, vidéo, etc.) et utiliser les métadonnées de ces documents.</i></p>

Tableau 3 : Comparaison des définitions de l'indexation automatique au niveau scientifique, institutionnel et professionnel.

Notion didactisée : INDEXATION AUTOMATIQUE

	Débutant	Intermédiaire	Avancé
Caractéristiques de la notion	.Parcours et repère le contenu des documents numériques automatiquement	.Analyse le contenu d'un document de façon informatisée	.Analyse tout ou partie du contenu d'un document via plusieurs algorithmes (Lorsque l'analyse porte sur le sens du texte , on parle d' indexation automatisée).
	Consiste en une opération informatique	.Suppose une programmation informatique	.Suppose quatre étapes dans un moteur de recherche : - la segmentation (identification des différentes zones dans un document : ex. zone de texte, zone d'image, zone de tableau, etc.) - la normalisation (« lecture » et reconnaissance des mots-clés dans la barre de recherche d'un moteur de recherche), - l' analyse syntaxique (identification de la fonction grammaticale de chaque mot dans une phrase) - la structuration et la pondération (comptabilisation des mots et de leur fréquence dans un document)
	Concerne tout type de document numérique	.Concerne tout type de document numérique (textes, images, sons, vidéos).	.Concerne tout type de document numérique, ou numérisé (textes, images, sons, vidéos).
	.Permet de retrouver facilement et rapidement des informations contenues dans un ou plusieurs documents	.Permet de retrouver facilement des informations : - contenues dans plusieurs documents - à partir d'un moteur de recherche	.Permet de retrouver facilement des informations : - à partir d'une base de données - à partir d'un moteur de recherche
	.Génère et utilise un index de données et de métadonnées .	- Génère et exploite un index de données et de métadonnées (donnée servant à décrire une autre donnée de façon normalisée).	
Niveaux de formulation	L' indexation automatique est une opération informatique qui permet de parcourir et de repérer facilement et rapidement le contenu des documents numériques automatiquement .	L' indexation automatique analyse le contenu intellectuel d'un document de façon informatisée . Elle suppose une programmation informatique . Elle concerne tout type de document numérique (textes, images, sons, vidéos). Elle permet de retrouver plus facilement des informations contenues au sein de plusieurs documents et/ou à partir d'un moteur de recherche . Elle génère un index de données et de métadonnées .	L' indexation automatique analyse tout ou partie du contenu d'un document à partir de plusieurs algorithmes. Elle concerne tout type de document numérique (textes, images, sons, vidéos). Elle permet de retrouver facilement des informations dans une base de données ou dans un moteur de recherche . Elle suppose quatre étapes dans un moteur de recherche : - la segmentation , qui consiste en l'identification des différentes zones dans un document (ex. zone de texte, zone d'image, zone de tableau, etc.) - la normalisation , qui consiste à « lire » les mots-clés de la requête dans la barre de recherche. - l' analyse syntaxique , qui permet l'identification de la fonction grammaticale de chaque mot dans une phrase. - la structuration et la pondération , qui permettent de comptabiliser les mots dans un document et d'en estimer leur fréquence. Au cours de ce processus, l'indexation automatique génère et exploite un index de données et de métadonnées (données servant à décrire d'autres données).

Tableau 4 : Didactisation de la notion d'indexation automatique par les caractéristiques et les niveaux de formulations débutant, intermédiaire et avancé. En gras, les caractéristiques importantes. En rouge, les éléments ajoutés d'un niveau de formulation à l'autre. Afin d'affiner les caractéristiques et les niveaux de formulation de la notion d'indexation automatique, nous avons travaillé avec l'ensemble du groupe de travail sur l'indexation automatique du bassin du Mans²³

²³Larissa Allard et Isabelle Magadur, professeures documentalistes (LPO Raphaël Elizé, Sablé-sur-Sarthe)

1.4 Cadre conceptuel et hypothèses de recherche

1.4.1 Cadre conceptuel de recherche

- Concepts scientifiques retenus

De l'ensemble des travaux scientifiques en sciences de l'information et de la communication évoqués précédemment, nous retiendrons que l'indexation automatique est un processus qui a grandement évolué avec le *Web* en lien avec le fonctionnement des moteurs de recherche. La définition scientifique sur laquelle nous nous appuyons pour effectuer une transposition didactique de la notion est celle inspirée d'Alexandre Serres (données non publiées) : **l'indexation automatique est un processus d'analyse du contenu d'un document informatisé pour lequel l'accès au contenu du document est augmenté par un index de données et de métadonnées, et dont l'efficacité de la recherche est améliorée par des algorithmes toujours plus perfectionnés facilitant la catégorisation (cible), l'anticipation (requête) et la lexicalisation (clé : recherche en texte brut)**. De plus, nous nous basons sur les quatre étapes qui régissent le fonctionnement de l'indexation automatique au sein d'un moteur de recherche : **la segmentation, la normalisation, l'analyse syntaxique et la structuration de l'information**.

L'arrivée dans le *Web sémantique* permet aux moteurs de recherche d'aller plus loin que l'indexation automatique, en indexant les données et les métadonnées des documents du *Web* afin que les algorithmes les « comprennent », créent du « sens », ou du lien sémantique. Les algorithmes sous-jacents sont basés sur le fonctionnement des ontologies et de l'apprentissage-machine.

Ces nouvelles technologies du *Web* sont, en partie, utilisées dans des cartouches d'information appelés ***infoboxes* ou *encarts informationnels***, nos objets d'étude. Nous avons mis en évidence quatre types *d'infoboxes* (des plus anciens au plus récents) : les ***infoboxes Wikipédia***, les ***infoboxes d'ontologie***, les ***agrégations de contenu*** et les ***infoboxes de type dictionnaire***. Alors que les *infoboxes Wikipédia* sont indexées manuellement, les autres *infoboxes* relèvent de

Christelle Cissé, professeure documentaliste (LPO Le Mans Sud, Le Mans) ;
Valérie Caniart, professeure documentaliste (collège Jacques Prévert, Pontvallain) ;
Jean-Noël Amato, étudiant en Master 1 MEEF Documentation (INSPE Angers).

procédés algorithmiques, issus du *Web de données*, plus complexes, relevant de la récolte d'informations pointant vers d'autres sites *Web* (*infoboxes* d'ontologie), d'extraction et de regroupement automatique de portions d'informations (agrégation de contenu et *infoboxes* type dictionnaire), voire d'une « reformulation » (*infoboxes* type dictionnaire).

Même si le travail de didactisation de la notion d'indexation avait déjà été effectuée par le passé, il était nécessaire de la revoir à l'heure de l'évolution des nouvelles connaissances scientifiques des technologies du *Web* et de l'évolution des professions liées à la documentation (pratiques sociales de référence). Il nous a semblé nécessaire de clarifier la différence entre indexation automatisée et indexation automatique car elles étaient utilisées de façon peu explicitée par les professeur.e.s documentalistes. La définition que nous avons retenue pour l'indexation automatisée est : « ***l'indexation automatisée consiste en une indexation automatique « avancée » qui permet une indexation plein texte et l'accès au sens des mots d'un document*** ». L'indexation automatique présente dans les logiciels des **professionnels de la documentation** montre des différences avec celle utilisée dans les moteurs de recherche, notamment l'utilisation du **procédé de vectorisation** pour retrouver un document à partir des mots-clés. Comme pour la définition originelle de **l'indexation**, pour les professionnels de la documentation, **l'indexation automatique** a aussi pour objectif de **faciliter l'accès aux documents** par les usagers. Nous retiendrons que les professionnels de la documentation favorisent l'utilisation de **l'indexation manuelle**, ou **humaine**, car elle est plus précise que **l'indexation automatique**.

Afin de préciser les différences entre **indexation humaine**, **indexation documentaire** et **indexation automatique**, nous retenons que **l'indexation humaine** peut nécessiter un langage contrôlé (de type thésaurus), ou non (mots-clés, tags). **L'indexation documentaire** est systématiquement liée à un langage contrôlé. **L'indexation automatique** est, quant à elle, obligatoirement informatique mais peut nécessiter des techniques d'indexation documentaire.

Dès lors, pour **enseigner l'indexation automatique** avec les élèves, nous plaçons les méthodes pédagogiques à tester et la transposition didactique de cette notion au sein de la culture de l'information et dans la discipline de l'information documentation. Dans le large potentiel des méthodes pédagogiques

existantes, nous privilégierons les **méthodes actives** et nous favoriserons la **méthode spiralaire** à la manière du *Wikinotions*.

Nous avons également observé que, pour l'Institution, la notion d'**indexation automatique** est plus abordée dans les programmes du cycle 3 que du cycle 4, et qu'elle est surtout privilégiée dans les programmes du lycée général avec la discipline des **sciences numériques et technologie** (SNT) en seconde. Avant cette discipline de seconde générale, les professeur.e.s documentalistes enseignaient l'indexation automatique dans le cadre de l'accompagnement personnalisé (AP) en ce qui concerne les lycée. Il est assez fréquent que les professeur.e.s documentalistes enseignent l'indexation automatique dans le cadre du fonctionnement des moteurs de recherche et ce, dès le collège. Cependant, certain.e.s d'entre-eux.elles l'enseignent pour la différencier de l'indexation humaine ou manuelle, notamment en fin de cycle 3, comme préconisé par l'association des professeur.e.s documentalistes de l'éducation nationale (APDEN) dans le dossier « Vers un curriculum en information-documentation ».

- **Problématisation et problématique**

Parmi les quatre types *d'infobox* résumées dans le tableau 1, celle qui est la plus ancienne et la plus commune aux enseignants et aux élèves est *l'infobox Wikipédia*. De plus, en termes de procédés d'indexation, *l'infobox Wikipédia* reste la plus simple. **L'indexation** peut être **manuelle** ou **automatique** : 1) *l'infobox Wikipédia* qui est indexée manuellement et 2) *l'infobox* de la page de résultats des moteurs de recherche, aussi appelée *infobox d'ontologie*, qui indexe automatiquement les informations de la page Wikipédia et centralise les informations issues de différents sites *Web*. Sur son encyclopédie collaborative en ligne, l'équipe de Wikipédia a publié un article en 2018, « *Wikipédia dans la classe* », donnant des pistes pédagogiques qui peuvent être utilisées par les enseignants (Wikipédia dans la classe, 2018). Parmi ces pistes, la plus répandue est le complément, ou l'amélioration d'une page de l'encyclopédie déjà existante. La plupart des professeur.e.s documentalistes qui se sont emparé.e.s de cet outil en classe ont fait en sorte que les élèves comprennent le fonctionnement de cette encyclopédie collaborative pour développer plusieurs notions info-documentaires : fiabilité de l'information, source, droit d'auteur, bibliographie, écritures collaboratives, auteur, autorité, etc (Séances pédagogiques Wikipédia,

Pearltrees²⁴). A notre connaissance, aucune séance pédagogique en information-documentation intégrant les *infoboxes* Wikipédia comme objet d'étude n'a été publiée dans des ressources professionnelles en ligne. Par conséquent, nous sommes en droit de nous demander si cela aurait un intérêt d'étudier les *infoboxes* en information-documentation et notamment pour comprendre le fonctionnement de l'indexation automatique dans un moteur de recherche.

En fait, l'intérêt est ici de comparer l'indexation manuelle et l'indexation automatique à partir de deux types *d'infoboxes* : les *infoboxes* Wikipédia, et les *infoboxes* d'ontologie qui reprennent les informations de la page Wikipédia et qui apparaissent sur la page de résultats des moteurs de recherche. Ainsi, il est possible de mettre les élèves devant une situation-problème à partir *d'infoboxes* qui se ressemblent, mais qui ne sont pas indexées de la même manière. La question devient donc, d'où provient l'information présentée dans une *infobox* Wikipédia et une *infobox* d'ontologie ? Comment ces *infoboxes* sont-elles générées ?

Dès lors, ne serait-il pas pertinent d'enseigner l'indexation automatique à partir des *infoboxes* Wikipédia et des *infoboxes* d'ontologie, afin d'affiner la compréhension du fonctionnement actuel des moteurs de recherche ? Et surtout, comment l'enseigner avec des élèves ?

Les *infoboxes* sont intéressantes à travailler avec les élèves à trois niveaux :

1. Elles permettent de mieux comprendre le fonctionnement des moteurs de recherche,
2. Elles incitent les élèves à prendre conscience de l'origine de l'information présente dans les *infoboxes* et de les évaluer,

Elles autorisent l'apprentissage de la notion d'indexation automatique pour elle-même, sans aborder nécessairement le fonctionnement des moteurs de recherche.

L'apparition des différents encarts informationnels, ou *infoboxes*, sur certains moteurs de recherche comme Google et Bing pose la question de l'origine

²⁴Séance pédagogique {Wikipédia}. Pearltrees. Consulté 5 juin 2020, à l'adresse <https://www.pearltrees.com/u/24010211-a-propos-de-wikipedia#1975>

de l'information. La plupart de ces *infoboxes* ne renvoient pas nécessairement à une source identifiée. Comment un.e élève, face à plusieurs *infoboxes* sur la même page de résultats d'un moteur de recherche, va-t-il.elle être capable de déterminer la crédibilité et l'origine de l'information présente dans ces encarts informationnels ? L'élève sera potentiellement tenté de se référer à ces ressources pour deux raisons : d'abord parce qu'elles sont présentes sur Google (Moteur de recherche utilisé par 90% des usagers, et qui apparaît comme donnant des résultats et informations fiables selon les élèves ; Rousseau, 2015) et ensuite parce qu'elles sont représentées majoritairement en haut de la première page de résultats, et donc bien visibles. A travers ces *infoboxes*, la question de l'évolution des méthodes d'indexation automatique peut potentiellement être abordée avec les élèves ainsi que l'évolution du *Web* jusqu'au *Web sémantique* actuel. Alors que la notion d'indexation automatique a été abordée assez tôt au début des années 2000 pour comprendre le fonctionnement des premières versions des moteurs de recherche, nous ne savons pas si les professeur.e.s documentalistes se sont emparé.e.s des *infoboxes* dans leurs enseignements et quelles notions il.elle.s abordent potentiellement avec les élèves. Les enjeux pour les professeur.e.s documentalistes de l'utilisation des *infoboxes* pour didactiser l'indexation automatique sont de trois niveaux. Premièrement, l'étude des *infoboxes* permet de comprendre que l'indexation automatique n'est pas uniquement utilisée dans les moteurs de recherche pour donner accès à des pages Web (page de résultats). Deuxièmement, à travers les matériaux *infoboxes*, il est possible de comprendre le fonctionnement de l'indexation qui permet d'obtenir ces contenus informationnels : indexation manuelle vs. indexation automatique. Enfin, comprendre le fonctionnement de ces *infoboxes* offre aux élèves la possibilité de porter une évaluation sur les informations données dans les *infoboxes* : à la fois leur fiabilité, mais aussi leur pertinence. De ce fait, les *infoboxes* constituent-elles un nouvel objet d'étude permettant la didactisation des processus d'indexation automatique par les professeur.e.s documentalistes, et une meilleure compréhension, par les élèves du traitement de l'information par les moteurs de recherche ?

1.4.2 Hypothèse de recherche

Afin d'interroger cette problématique, nous voulons savoir comment les professeur.e.s documentalistes enseignent l'indexation automatique à travers les questions de recherche suivantes :

1) Dans quelle mesure l'indexation automatique est-elle enseignée aujourd'hui ?

Hypothèse A: L'indexation automatique est peu enseignée par les professeur.e.s documentalistes.

Nous avons vu que l'indexation automatique est enseigné par certain.e.s professeur.e.s documentalistes, d'après les séances pédagogiques que nous avons relevées, parlant de l'indexation automatique (Annexe 2). De plus, la notion d'indexation (dans le sens automatique) est présente explicitement dans certains programmes du lycée général et technologique. Cependant, il apparaît tout de même que cette notion est peu enseignée au niveau professionnel et peu recommandée au niveau institutionnel. Nous nous attendons donc à ce que la notion d'indexation automatique soit peu enseignée par les professeur.e.s documentalistes.

Hypothèse B: Il y a un effet générationnel sur l'enseignement de l'indexation automatique.

Comme la notion d'indexation et d'indexation automatique est antérieure à la mise en place des moteurs de recherche. Nous nous attendons à ce que les professeur.e.s documentalistes ayant plus d'expérience se soient plus généralement emparés de cette notion que les professeur.e.s documentalistes avec moins d'expérience.

2) L'enseignant-il.elle.s uniquement dans le cadre du fonctionnement des moteurs de recherche ?

Hypothèse C: L'indexation automatique est majoritairement enseignée dans le cadre du fonctionnement des moteurs de recherche

En effet, nous avons vu dans les séances pédagogiques présentées au début des années 2000 (Annexe 2), que bon nombre d'entre elles utilisaient

comme matériaux les moteurs de recherche pour aborder la question de l'indexation automatique (Partie 1.3.2. : « *L'indexation automatique dans le secondaire* »). Cependant, dans le programme optionnel de seconde de science numérique et technologie, la notion d'indexation automatique est aussi associée à la gestion des bases de données. De même dans le programme de management, sciences de gestion et numérique de terminale STMG (sciences et technologie du management et de la gestion), l'indexation est associée à l'échange de documents numériques. Par conséquent, nous émettons l'hypothèse que l'indexation automatique est majoritairement enseignée dans le cadre du fonctionnement des moteurs de recherche, mais que ce n'est probablement pas la seule porte d'entrée.

3) Utilisent-il.elle.s des « modèles » de séances pédagogiques de référence pour enseigner l'indexation automatique ?

Hypothèse D : Les professeur.e.s documentalistes créent leurs propres séances pédagogiques pour aborder la notion d'indexation automatique.

Dans la mesure où certaines séances sont reprises par d'autres collègues, comme celle de Frédéric Rabat « Une année avec Google (suite) » (Rabat, 2008) en seconde, existe-t-il des séances pédagogiques « types » qui sont reprises par les professeur.e.s documentalistes pour enseigner l'indexation automatique ? En considérant des séances évoquées dans notre cadre théorique, il semblerait que chaque professeur.e documentaliste crée ses propres séances, à quelques exceptions près, notamment celles liées à la notion plus générale des moteurs de recherche. Nous émettons l'hypothèse que la majorité des professeur.e.s documentalistes créent leurs propres séances pédagogiques lorsqu'ils veulent enseigner l'indexation automatique.

4) Considèrent-il.elle.s nécessaire d'intégrer la notion d'indexation en amont de la notion d'indexation automatique ?

Hypothèse E : Les professeur.e.s documentalistes n'abordent pas systématiquement la notion d'indexation en amont de celle d'indexation automatique.

Au niveau historique, nous avons vu que l'indexation, vient de la mise en place de l'index au Moyen-Âge, puis de l'indexation manuelle et enfin de

l'indexation automatique. Est-ce que les professeur.e.s documentalistes considèrent comme essentiel de reprendre le cheminement historique de la pensée des inventeurs pour aborder la question de l'indexation automatique ? En d'autres termes, est-il fondamental d'enseigner l'indexation avant l'indexation automatique ? Dans les séances que nous avons étudiées, tel était rarement le cas. Dans le programme de SNT, il est préconisé d'enseigner l'historique de l'indexation ; humaine dans un premier temps et automatique ensuite. A l'inverse, l'APDEN propose d'aborder la notion spécifique d'indexation automatique dès la classe de 5^e avant d'aborder la notion organisatrice d'indexation, déclinée ensuite en indexation documentaire et automatisée, en classe de 4^e. Nous émettons l'hypothèse, par conséquent, que la notion d'indexation automatique est abordée pour elle-même par les professeur.e.s documentalistes sans nécessairement évoquer la notion d'indexation en amont.

5) Utilisent-il.elle.s les *infoboxes* dans leurs enseignements ? Et si oui, sont-elles un support de l'enseignement de l'indexation automatique ?

Hypothèse F: Les professeur.e.s documentalistes n'utilisent pas les infoboxes dans leur enseignements

Hypothèse G: Quand il.elle.s l'utilisent, les infoboxes ne sont pas un support de l'enseignement de l'indexation automatique

Nous avons vu les potentialités des *infoboxes* pour traiter de la question de l'indexation manuelle et automatique. Nous n'avons pas trouvé de séances en ligne utilisant les *infoboxes* comme matériaux pour aborder la question de l'indexation automatique. Est-ce un biais lié à notre veille ou bien les professeur.e.s documentalistes ne se sont pas emparés des *infoboxes* ? S'il.elle.s les utilisent, abordent-ils la question de l'indexation automatique ?

Partie 2 :

Projet de recherche

2 **Projet de recherche**

2.1 ***Méthodologie de collecte de données***

2.1.1 **Échantillonnage et critères de sélection**

Pour aborder la question de recherche évoquant la manière dont les professeur.e.s documentalistes enseignent l'indexation automatique, nous avons fait le choix de nous appuyer sur une collecte de données basée sur un questionnaire. En effet, le questionnaire nous semblait le plus approprié afin de maximiser notre échantillonnage auprès des professeur.e.s documentalistes. L'objectif n'est pas ici de récolter des opinions sur les caractéristiques mêmes de l'indexation automatique, mais plutôt d'avoir un retour le plus large possible de la façon dont les professeur.e.s documentalistes se sont emparé.e.s de la notion d'indexation automatique et comment ils l'enseignent aux élèves. Étant donné que notre objectif est de questionner des professeur.e.s documentalistes de terrain, il n'a pas été envisagé d'élargir le questionnaire à d'autres professionnels de l'information-documentation, comme des chercheurs ou des enseignants de l'INSPE²⁵.

Les hypothèses que nous posons ne se limitent pas à une seule académie. De plus, comme il y a une hypothèse portant sur un facteur générationnel, il est important que l'échantillonnage soit le plus large possible pour pouvoir ensuite analyser les données et en tirer des conclusions. Par conséquent, le questionnaire que nous avons établi a été diffusé le plus largement possible à l'ensemble des académies. Deux listes de diffusion ont été utilisées : la liste de diffusion des professeur.e.s documentaliste de l'académie de Nantes (Lettre Edu_Num Documentation ; Liste institutionnelle administrée par l'IAN Jean-François Lemarié), et la liste de diffusion des professeur.e.s documentalistes de France (CDI-Doc ; Liste chapeautée par le Réseau Canopé, sous tutelle du Ministère de l'Éducation nationale et administrée par des professeur.e.s documentalistes : Amaury Catel et Laure Luchez).

²⁵INSPE : Instituts nationaux supérieurs du professorat et de l'éducation.

2.1.2 Présentation du questionnaire

Le questionnaire est composé de sept questions majeures déclinées en sous-questions (Annexe 3). Les sept questions sont réparties en fonction des hypothèses et questions de recherche posées en amont (Tableau 5).

- Hypothèse A : Pour tester la première hypothèse selon laquelle les professeur.e.s documentalistes enseignent peu l'indexation automatique (Tableau 5), nous proposons de demander dans un premier temps aux professeur.e.s documentalistes de répondre de façon binaire, s'ils abordent la question de l'indexation automatique ou non (question Q1, Tableau 5 ; Annexe 3). Dans le cas d'une réponse par l'affirmative, nous demandons, à travers une question à réponse ouverte et courte, de préciser le.s niveau.x au.x.quel.s l'indexation automatique est enseignée.
- Hypothèse B : Pour savoir s'il y a un effet générationnel sur l'enseignement ou non de l'indexation automatique (Tableau 5), nous demandons, avec la question 6 du questionnaire (Annexe 3), aux professeur.e.s documentalistes de renseigner dans un encart de réponse ouverte et courte le nombre d'années d'ancienneté en tant que professeur.e documentaliste.
- Hypothèse C : Afin de mettre à l'épreuve l'hypothèse selon laquelle l'indexation automatique est enseignée uniquement dans le cadre du fonctionnement des moteurs de recherche (Tableau 5), nous proposons de cocher, à travers un questionnaire à choix multiples parmi une liste de notions génériques potentielles, celles qui correspondent aux enseignements que les professeur.e.s documentalistes pratiquent (question Q1c, Annexe 3). Les trois notions proposées sont : le moteur de recherche, la veille et le portail documentaire. Nous laissons la possibilité d'en ajouter d'autres avec un encart de réponse ouverte. Le choix entre ces trois notions se justifie par le fait que le moteur de recherche est la principale notion rattachée à l'indexation automatique. La veille et le portail documentaire, sont deux autres notions auxquelles nous supposons que l'indexation automatique peut être rattachée si l'on se réfère aux 10 domaines des notions info-documentaires de P. Duplessis (2016).

- Hypothèse D: Dans le but de constater si les professeur.e.s documentalistes créent majoritairement leurs propres séances pédagogiques pour aborder la notion d'indexation automatique, nous proposons aux professeur.e.s documentalistes de répondre à deux questions (questions Q4 et Q5, Tableau 5, Annexe 3). Dans la première (question Q4), nous leur demandons, à travers un choix binaire, s'il.elle.s utilisent ou adaptent des séances pédagogiques existantes pour réaliser leurs séances qui abordent l'indexation automatique. Dans le cas d'une réponse par l'affirmative, nous leur demandons de préciser quelles sont ces séances. Dans la deuxième question qui se rapporte à cette hypothèse, nous demandons aux professeur.e.s documentalistes, à travers une réponse ouverte, la méthode pédagogique qui leur semble la plus pertinente pour enseigner l'indexation automatique (question Q5, Tableau 5, Annexe 3). Dans cette question Q5, nous avons élargi la question aux moteurs de recherche et aux encarts informationnels (*infoboxes*), car il nous semblait important de ne pas se limiter à une seule notion pour évoquer une méthode pédagogique idoine, et cela permettait de compléter les *hypothèses A et G*.
- Hypothèse E: Nous voulons tester, ici, l'hypothèse selon laquelle les professeur.e.s documentalistes n'abordent pas systématiquement la notion d'indexation en amont de la notion d'indexation automatique. À travers la question Q2 (Tableau 5, Annexe 3), nous demandons de façon assez simple si oui ou non il.elle.s enseignent l'indexation en amont de la notion d'indexation automatique et de justifier dans les deux cas.
- Hypothèses F et G: Ces hypothèses stipulent que les professeur.e.s documentalistes n'utilisent pas les encarts informationnels, ou *infoboxes*, dans leurs enseignements, et que, quand bien même il.elle.s le feraient, ce ne serait pas pour aborder la question de l'indexation automatique (Tableau 5). À travers la question Q3a (Tableau 5, Annexe 3), nous demandons aux professeur.e.s documentalistes s'il.elle.s ont déjà remarqué ces encarts informationnels dans une page de résultats de moteur de recherche, si oui, lesquels. Pour clarifier la description des types d'encarts à identifier, nous avons défini en amont ce qu'était un encart informationnel, ou *infobox*, et nous avons nommé ces trois types d'encarts comme des encarts

« Wikipédia », « Images ou vidéo » et « Définition ». Il aurait été peu opportun, dans un souci de clarté, et donc de précision dans la réponse, de nommer ces trois encarts « Ontologie », « Aggrégation de contenu » et « Définitionnel », même si ce dernier porte clairement son nom. Avec la question 3b (Tableau 5, Annexe 3, 4), nous interrogeons les professeur.e.s documentalistes sur le fait qu'il.elle.s se soient emparé.e.s ou non de la question des encarts informationnels avec les élèves. Afin de déterminer la.les notion.s qu'il.elle.s ont abordée.s avec les élèves dans la mesure où il.elle.s se sont emparé.e.s de la question des *infoboxes* nous leur demandons dans une question ouverte ce qu'il.elle.s souhaiteraient que les élèves retiennent de l'étude des encarts informationnels (question Q3c, Tableau 4 ; Annexe 3).

- Question 7 : Avec cette dernière question, nous demandons l'origine académique des répondant.e.s (question Q7, Annexe 3). Cette question ne correspond pas à une hypothèse en particulier mais elle permet de vérifier s'il existe un biais dans l'origine académique des répondants pouvant expliquer certains écarts de résultats, notamment dans le nombre de répondant.e.s. Nous nous attendons à ce que les professeur.e.s documentalistes de l'académie de Nantes répondent plus facilement, même par la négative, que les autres académies.

Nous avons fait le choix de présenter les questions du questionnaire dans un ordre différent de celui des hypothèses, car nous souhaitons privilégier la cohérence dans l'ordre des questions par rapport à leur imbrication, plutôt que de garder celui des hypothèses. En effet, l'ordre des hypothèses de recherche aurait pu manquer de sens pour les répondant.e.s par rapport aux questions.

Les questionnaires sont anonymes car nous n'avions aucune hypothèse qui justifiait la récolte des données personnelles. De plus, pour respecter au mieux les données personnelles des répondant.e.s, nous avons édité le questionnaire avec un logiciel non propriétaire (Framaform²⁶ de Framasoft).

²⁶<https://framaforms.org/content/creez-et-diffusez-vos-formulaires-facilement>

Question de recherche	Hypothèses	Question(s) permettant d'y répondre
1. Dans quelle mesure l'indexation automatique est-elle enseignée?	A. L'indexation automatique est peu enseignée	Q1a : Enseignez-vous l'indexation automatique ? Q1b : Auprès de quel(s) niveau(x)
	B. Il y a un effet générationnel sur l'enseignement de l'indexation automatique	Q6 : Depuis combien de temps exercez-vous en tant que professeur.e. documentaliste ?
2. Les professeur.e.s documentalistes enseignent-ils l'indexation automatique uniquement dans le cadre des moteurs de recherche	C. L'indexation automatique est enseignée uniquement dans le cadre du fonctionnement des moteurs de recherche	Q1c : Avec quelles notions génériques abordez-vous la question de l'indexation automatique ? Q1d : Dans quel cadre pédagogique ?
3. Utilisent-il.elle.s des « modèles » de référence de séances pédagogiques pour enseigner l'indexation automatique ?	D. Les professeur.e.s documentalistes créent leurs propres séances pédagogiques pour aborder la notion d'indexation automatique.	Q4 : Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ? Q5 : Quelle méthode pédagogique vous paraît la plus pertinente pour enseigner l'indexation automatique, les moteurs de recherche ou les encarts informationnels ?
4. Les professeur.e.s documentalistes, considèrent-ils nécessaire d'intégrer la notion d'indexation en amont de la notion d'indexation automatique ?	E. Les professeur.e.s documentalistes n'abordent pas systématiquement la notion d'indexation en amont de la notion d'indexation automatique.	Q2 : Abordez-vous la notion d'indexation en amont de la notion d'indexation automatique ?
5. Les professeur.e.s documentalistes utilisent-ils.elle.s les infobox dans leurs enseignements ? Et si oui, sont-elles un support de l'enseignement de l'indexation automatique ?	F. Les professeur.e.s documentalistes n'utilisent pas les <i>infoboxes</i> dans leurs enseignements.	Q3a : Avez-vous remarqué les encarts informationnels sur une page de résultats ? Q3b : Abordez-vous la question des encarts informationnels avec les élèves ?
	G. Lorsqu'il.elle.s l'utilisent, ils n'abordent pas la question de l'indexation automatique.	Q3b : Si oui, que souhaiteriez-vous que les élèves retiennent ?

Tableau 5 : Relations entre questions de recherche, hypothèses de recherche et les questions du questionnaire permettant d'y répondre

2.2 **Méthodologie de traitement des données**

2.2.1 **Types de données**

Les analyses statistiques seront réalisées à partir de données qualitatives et quantitatives. Les verbatims seront organisés en fonction de catégories qui feront sens selon les hypothèses posées. Par exemple, à la question 3b du questionnaire, « Que souhaitez-vous que les élèves retiennent de l'étude des encarts informationnels ? », si plusieurs réponses indiquent qu'il est important que les élèves retiennent d'où vient l'information, il sera possible de catégoriser cette réponse en « origine de l'information ». Ensuite, toutes les réponses correspondant à une même catégorie pourront être comptabilisées.

La plupart des variables que nous aurons dans cette étude sont des variables discrètes qualitatives. Les variables discrètes qualitatives correspondent à des valeurs qui expriment une qualité, un état, un statut unique comme le nombre d'années d'exercice (question Q6, Tableau 5). En tant que variable seule, elles peuvent être analysées sous forme de pourcentage et/ou de fréquence. Pour analyser deux variables qualitatives ensemble, il est nécessaire d'effectuer un tableau de contingence (Lemelin, 2004). Un tableau de contingence « est une façon de présenter des données de comptage d'individus classés en catégories. [...] C'est déjà le résultat d'un traitement des données, puisque les individus (observations) ont dû être préalablement classés, puis comptés » (Lemelin, 2004). Plusieurs analyses multivariées peuvent être effectuées à partir des tableaux de contingence (Lemelin, 2004) :

- Mesure du degré d'association entre deux ou plusieurs variables ;
- Estimation des paramètres d'une relation entre deux ou plusieurs variables ;
- Évaluation de la significativité entre deux ou plusieurs variables
- Prédiction de l'appartenance d'un individu à un groupe à partir de ses autres caractéristiques
- Discernement d'une structure dans un ensemble de données.

En fonction de nos hypothèses et des questions que nous avons posées, nous nous bornerons aux deux premières approches d'analyses multivariées.

L'évaluation de la significativité n'a pas été effectuée car le jeu de données ne s'y prêtait pas.

Les contingences (croisements) possibles entre les données permettant d'affiner les hypothèses sont présentées dans la Figure 6. Toutes les données issues d'un tableau de contingence seront analysées selon une méthode de fréquence relative qui s'interprète comme une probabilité (Lemelin, 2014 ; Figure 6). Ici, nous avons choisi la méthode de probabilité conditionnelle car elle permet de tester si la probabilité d'appartenance d'une valeur de la variable α est conditionnée par son appartenance à une variable β . Autrement dit, c'est la probabilité d'appartenir à une variable, étant donné que l'on appartient à une autre variable (Lemelin, 2014). Pour faire ce calcul, il faut effectuer le rapport entre la valeur de la variable α , dans la variable β , et la somme des valeurs de la variable α (Figure 6). C'est un rapport compris entre 0 et 1. Plus le rapport est proche de 1 et plus la probabilité est forte que la variable α appartienne à la variable β . Ce calcul de probabilité conditionnelle nous permettra de tester les relations entre les variables contingentes.

Expérience	Indexation automatique		Total
	OUI	NON	
0-9 ans	5	8	13
10-19 ans	10	8	18
20-30 ans	10	4	14

Figure 6 : Exemple de calcul de probabilité conditionnelle : $5/13 = 0,38$. Ici la probabilité que la catégorie 0-9 ans d'expérience aient répondu « OUI » à la question de l'enseignement de l'indexation automatique est faible.

2.2.2 Présentation des données

Les données brutes seront présentées sous forme de tableaux. Les valeurs de pourcentages seront présentées sous forme de diagrammes circulaires et les valeurs de fréquence seront illustrées sous forme d'histogrammes ou de diagrammes circulaires. Les valeurs de contingence et les valeurs de probabilité conditionnelle seront présentées sous forme de tableaux et d'histogrammes.

Hypothèses	Questions	Types de variable	Contingences	Types d'analyse
A. L'indexation automatique est peu enseignée	Q1a : Enseignez-vous l'indexation automatique ?	Qualitative	<i>Q1b-Q1d : le cadre pédagogique dans lequel est enseigné l'indexation automatique est lié au niveau(x) d'enseignement(s) ?</i>	Tableau de contingence
	Q1b : Auprès de quel(s) niveau(x)	Qualitative		
B. Il y a un effet générationnel sur l'enseignement de l'indexation automatique	Q6 : Depuis combien de temps exercez-vous en tant que professeur.e. documentaliste ?	Quantitative	<i>Q1a x Q6 : l'enseignement de l'indexation automatique est dépendant du nombre d'années d'expérience</i>	Probabilité conditionnelle
C. L'indexation automatique est enseignée uniquement dans le cadre du fonctionnement des moteurs de recherche	Q1c : Avec quelles notions génériques abordez-vous la question de l'indexation automatique ?	Qualitative	<i>Q1c x Q1d : il existe une corrélation entre le cadre pédagogique et la notion générique enseignée</i>	Probabilité conditionnelle
	Q1d : Dans quel cadre pédagogique ?	Qualitative		
D. Les professeur.e.s documentalistes créent leurs propres séances pédagogiques pour aborder la notion d'indexation automatique.	Q4 : Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ?	Qualitative		Pourcentage et/ou fréquence
	Q5 : Quelle méthode pédagogique vous paraît la plus pertinente pour enseigner l'indexation automatique, les moteurs de recherche ou les encarts informationnels ?	Qualitative	<i>Q5 x Q1d : la méthode pédagogique utilisée est corrélée au cadre pédagogique</i>	Probabilité conditionnelle
E. Les professeur.e.s documentalistes n'abordent pas systématiquement la notion d'indexation en amont de la notion d'indexation automatique.	Q2 : Abordez-vous la notion d'indexation en amont de la notion d'indexation automatique ?	Qualitative	<i>Q2 x Q6 : il existe une corrélation entre l'expérience et le pré-requis de la notion d'indexation</i>	Probabilité conditionnelle
F. Les professeur.e.s documentalistes n'utilisent pas les infoboxes dans leurs enseignements.	Q3a : Avez-vous remarqué les encarts informationnels sur une page de résultats ?	Qualitative		Pourcentage et/ou fréquence
	Q3b : Abordez-vous la question des encarts informationnels avec les élèves ?	Qualitative		
G. Lorsqu'il.elle.s l'utilisent, ils n'abordent pas la question de l'indexation automatique.	Q3c : Si oui, que souhaiteriez-vous que les élèves retiennent ?	Qualitative	<i>Q3b x Q6 : il existe une corrélation entre l'expérience et l'enseignement des infoboxes</i>	Probabilité conditionnelle

Tableau 6 : Tableau des types de variables, leurs contingences potentielles et les types d'analyses envisagées pour traiter les tableaux de contingence, en considérant les hypothèses initiales.

Partie 3 :

Résultats et discussion

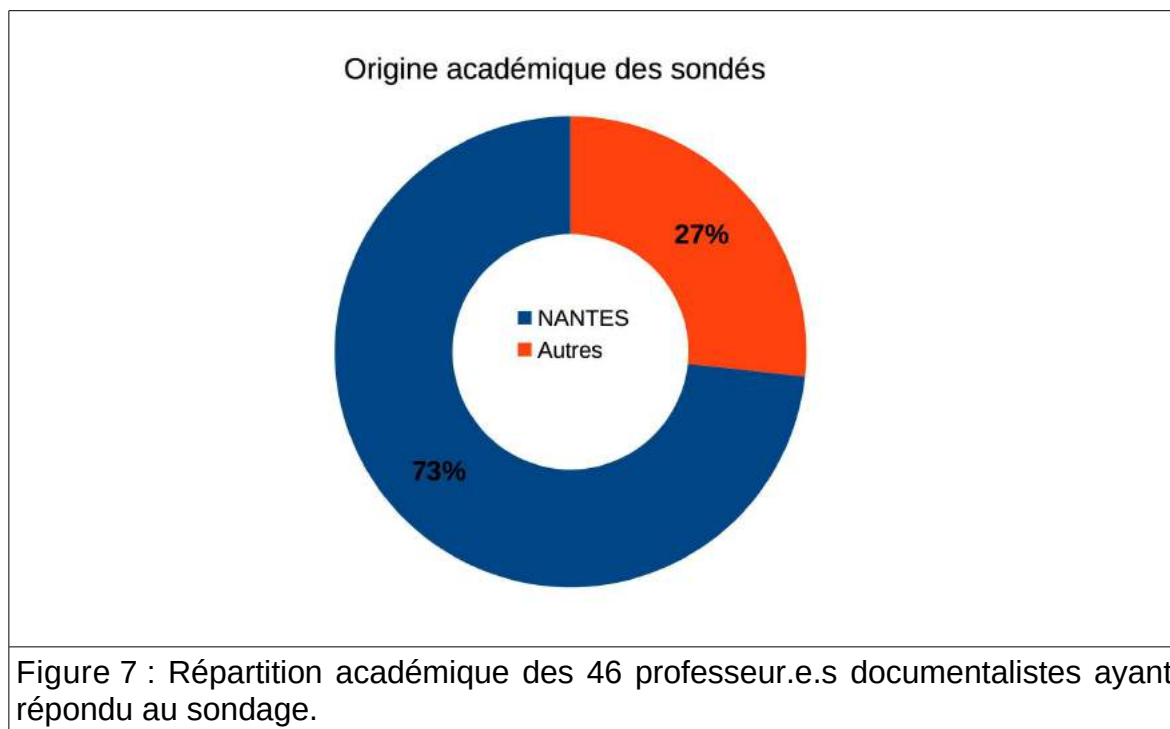
3 Résultats et discussion

3.1 Résultats

3.1.1 Cadre général des résultats

Nous avons interrogé les professeur.e.s documentalistes à travers la France afin de connaître leur pratique d'enseignement autour de la notion d'indexation automatique. Ceci, afin d'interroger la problématique qui demande si les *infoboxes* constituent un nouvel objet d'étude permettant la didactisation des processus d'indexation automatique par les professeur.e.s documentalistes, et une meilleure compréhension, par les élèves, du traitement de l'information par les moteurs de recherche.

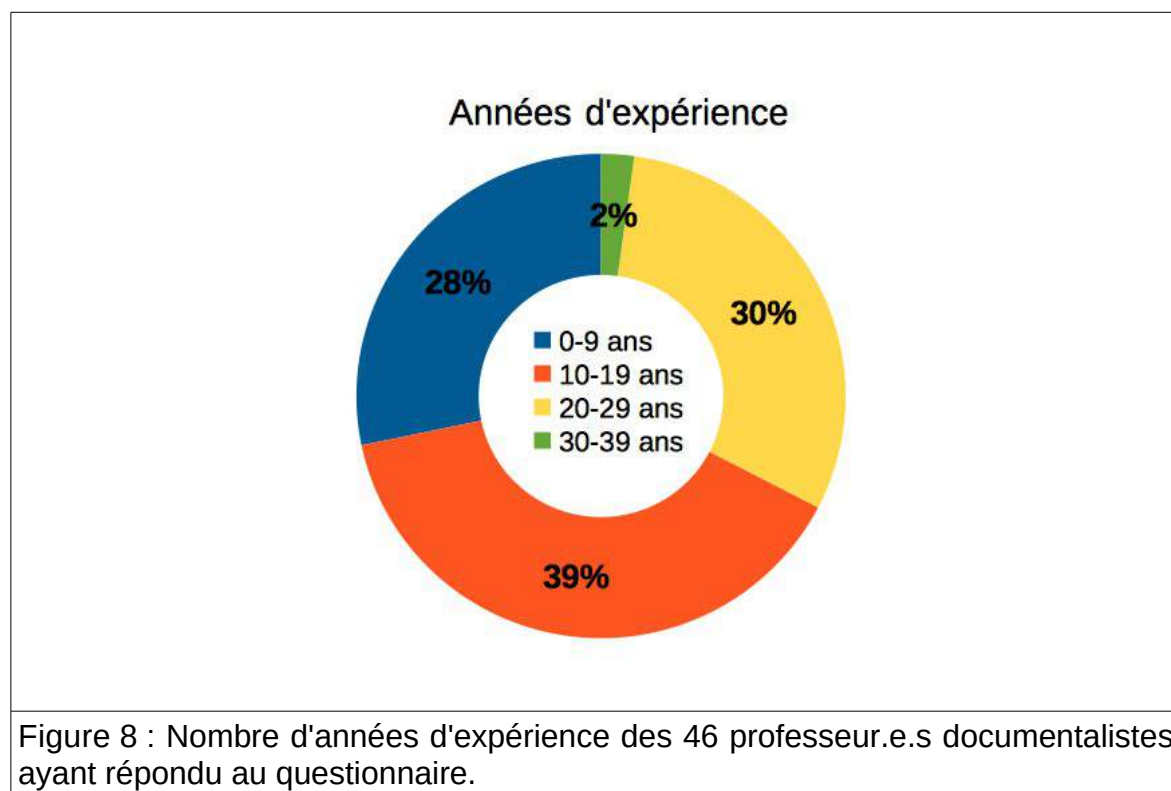
Le questionnaire de recherche que nous avons réalisé pour répondre à cette problématique a été complété par 46 professeur.e.s documentalistes, provenant majoritairement de l'Académie de Nantes (73%, Figure 7). Par conséquent, les résultats de notre questionnaire sont valables pour l'académie de Nantes. Il n'est pas représentatif des autres académies (Figure 7).



La moyenne d'années d'expérience pour les 46 professeur.e.s documentalistes ayant répondu au questionnaire est de 14 ans (Annexe 5 - Q6).

Selon les quatre intervalles de 10 ans d'expérience que nous avons définis, 39% des professeur.e.s documentalistes ayant répondu, ont entre 10 et 19 ans d'expérience (Figure 8). Les jeunes professeur.e.s documentalistes dans le métier (entre 0 et 9 ans d'expérience) représentent 28% et les professeur.e.s documentalistes expérimentés (entre 20 et 29 ans d'expérience) représentent 30% des personnes ayant répondu (Figure 8). Seule une personne ayant 30 ans de métier a répondu (2%, Figure 8). Le panel de notre étude peut être considéré comme représentatif de la profession, puisque ces proportions sont assez proches de la pyramide des âges des professeur.e.s documentalistes de France publiée par l'APDEN (moyenne d'âge de 46 ans et majorité entre 45 et 54 ans ; [APDEN, 2019](#)).

Cependant, nous pouvons supposer que la réponse au questionnaire sous format électronique a pu entraîner un léger biais positif dans la catégorie d'année d'expérience 10-19 ans des répondant.e.s.



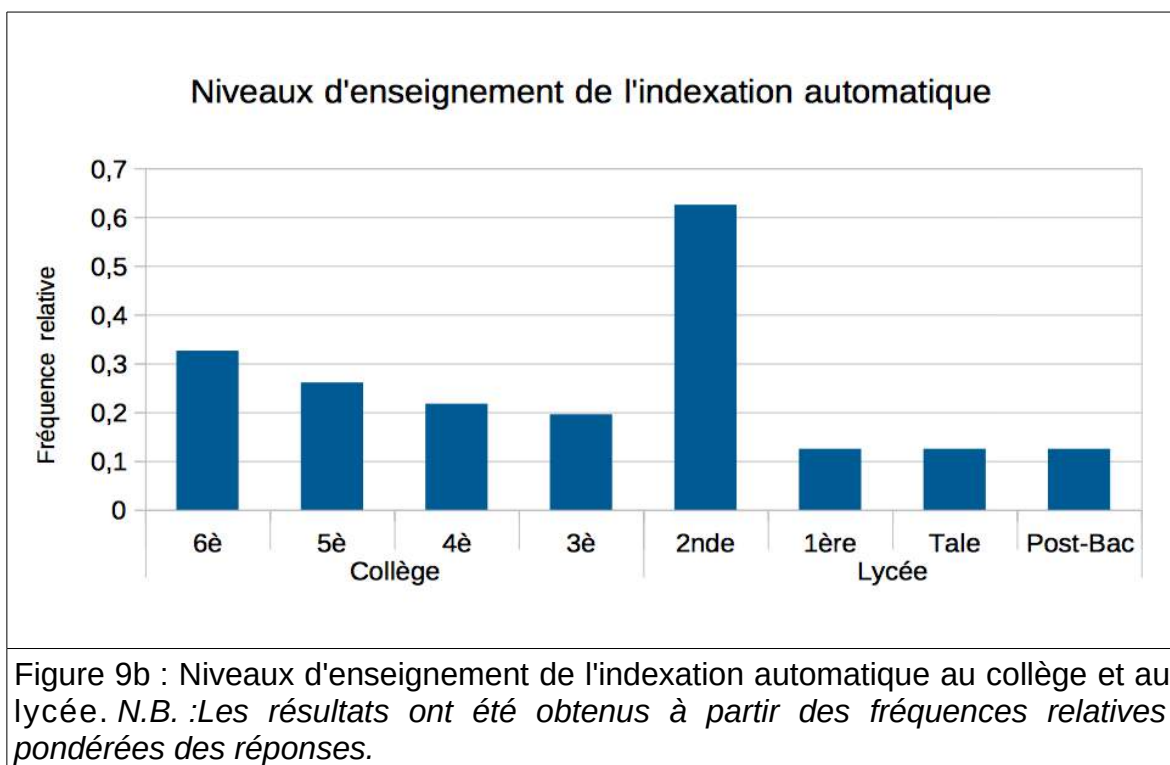
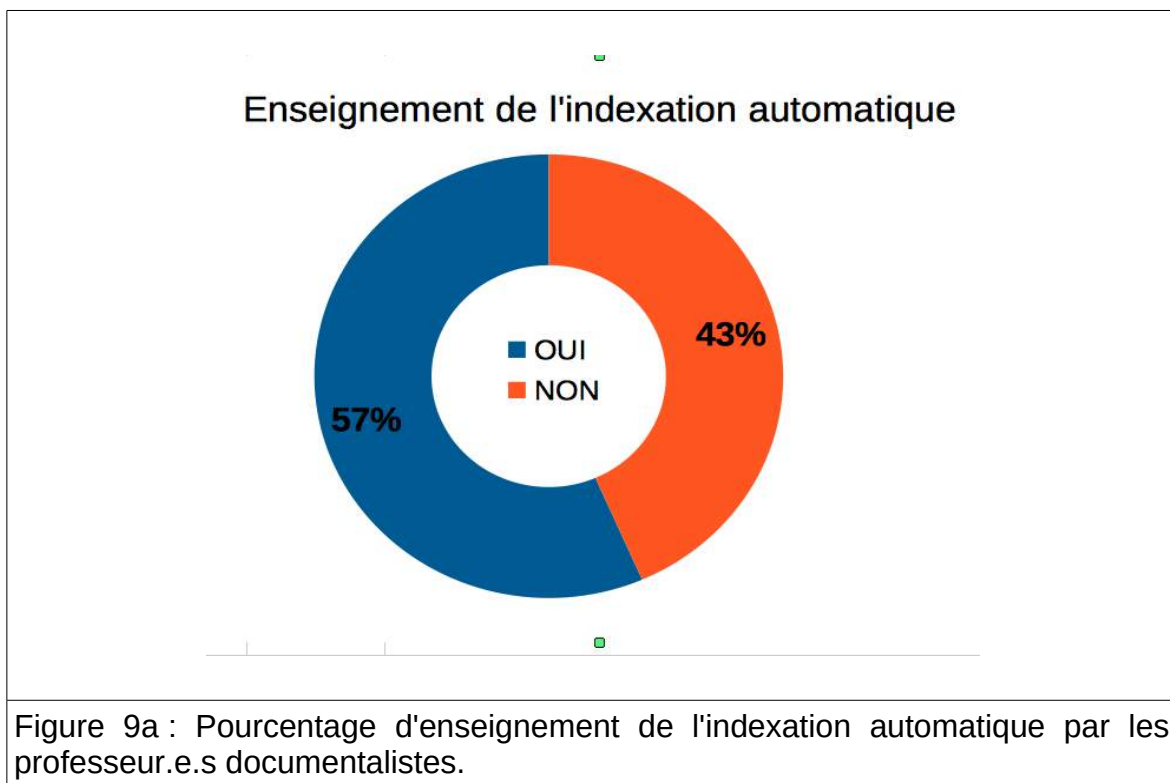
3.1.2 Réponse aux hypothèses

- *Hypothèse A : Les professeur.e.s documentalistes enseignent peu l'indexation automatique*

Réponse A : En fait, la majorité des professeur.e.s documentalistes de notre panel enseignent l'indexation automatique (57%) et surtout en classe de seconde.

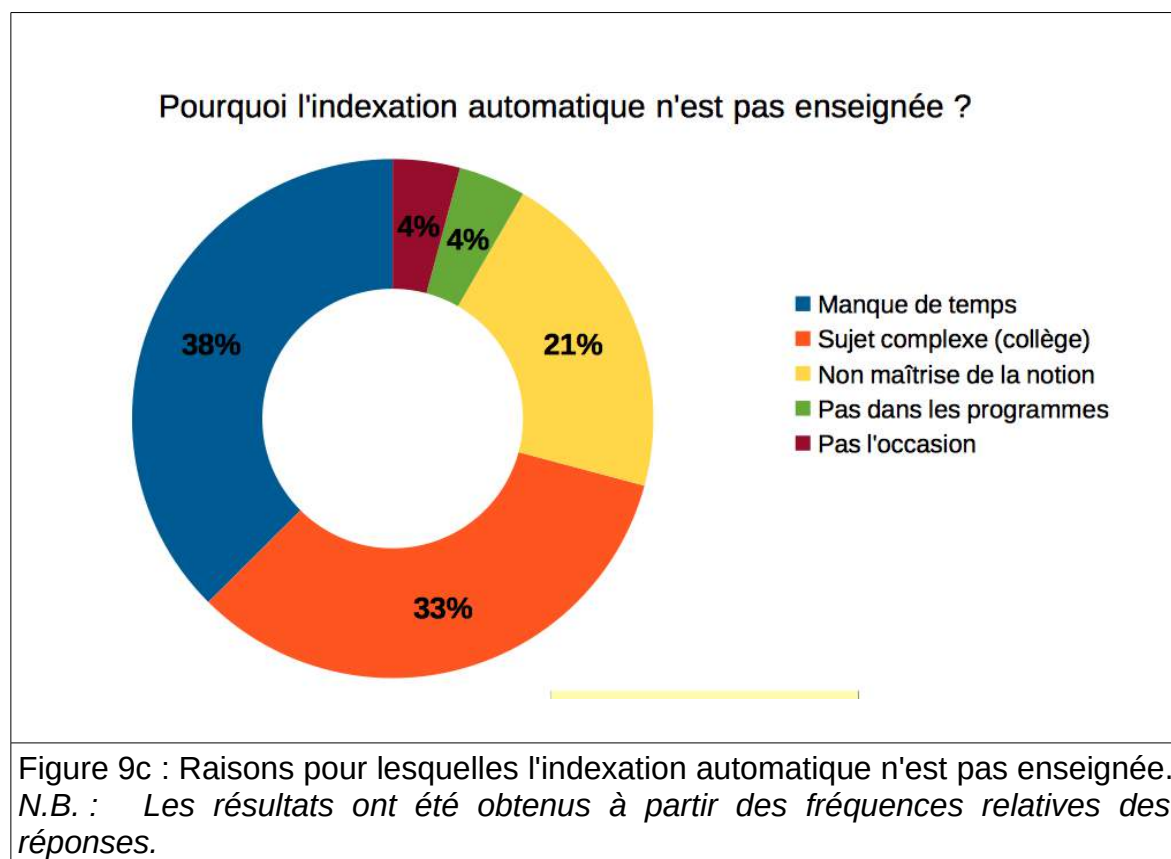
Dans un premier temps nous avons formulé l'hypothèse selon laquelle l'indexation automatique était peu enseignée par les professeur.e.s documentalistes car peu de séances étaient disponibles en ligne. Les résultats du questionnaire montrent que 58% de notre panel intègrent la notion d'indexation automatique dans leurs enseignements (Figure 9a ; Annexe 5 - Q1a) . Sept personnes ont précisé qu'il.elle.s enseignaient l'indexation automatique pour comprendre le fonctionnement des moteurs de recherche (évoqué 5 fois, notamment à travers la pertinence d'une page de résultats), en comparaison avec le portail documentaire (évoqué 3 fois) et pour travailler sur l'identité numérique et la traçabilité (évoqué 1 fois). Une seule personne ayant répondu a indiqué ne pas savoir ce qu'était l'indexation automatique (Annexe 4).

Pour comparer les niveaux d'enseignement de l'indexation automatique, nous avons calculé des fréquences relatives pondérées entre le collège et le lycée. En effet, nous avons obtenu 46 résultats pour le collège contre 16 pour le lycée. Par conséquent, il était nécessaire de pondérer les résultats en fonction du nombre de réponses pour le collège ou pour le lycée. Les résultats obtenus par le calcul des fréquences relatives pondérées montrent qu'au lycée, l'indexation automatique est majoritairement enseignée en seconde (Figure 9b ; Annexe 5 - Q1b). Au collège, où elle est moins enseignée, l'indexation automatique est abordée au niveau 6^e principalement (Figure 9b ; Annexe 5 - Q1b).



Pour la vingtaine de professeur.e.s documentalistes qui disent ne pas enseigner l'indexation automatique, 17 ont donné une raison à cela (Annexe 4). La

majorité de leurs réponses concernent le manque de temps (38% ; Figure 9c ; Annexes 4 et 5 - Q1b). La deuxième raison qui apparaît est la complexité de la notion à enseigner aux élèves, et surtout au niveau collège (33% ; Figure 9c ; Annexes 4 et 5 - Q1b). La troisième raison limitante est le fait de ne pas maîtriser suffisamment la notion personnellement pour être en mesure de l'enseigner aux élèves (21% ; Figure 9c ; Annexes 4 et 5 -Q1b).

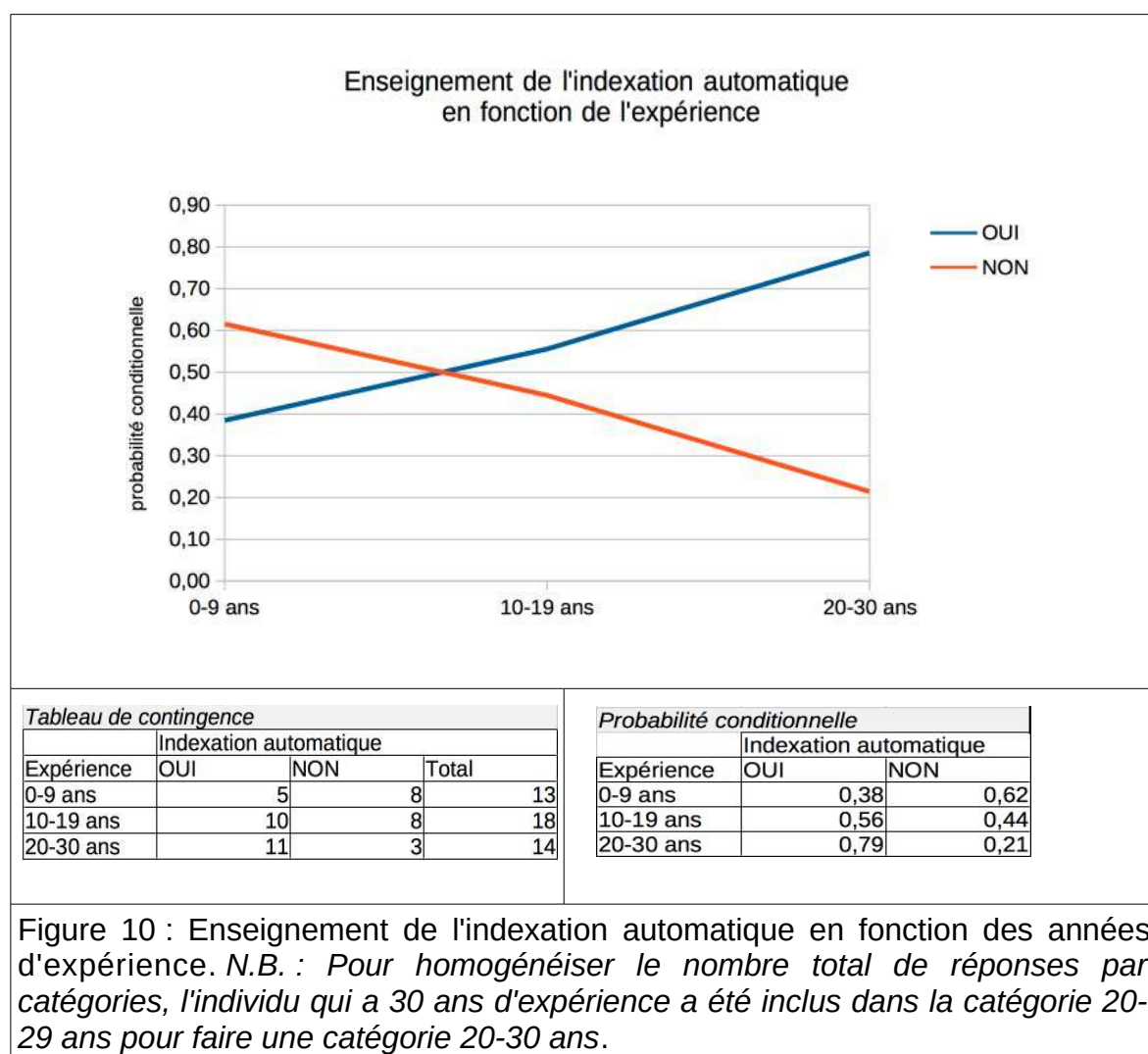


- Hypothèse B : Il y a un effet générationnel sur l'enseignement de l'indexation automatique

Réponse B : Oui il y a un effet. Plus les professeur.e.s documentalistes sont expérimentés, plus ils enseignent l'indexation automatique.

Pour répondre à l'hypothèse selon laquelle il y aurait un effet générationnel sur l'enseignement de l'indexation automatique, il est nécessaire de mettre en regard les données quantitatives liées au nombre d'années d'expérience et les données qualitatives liées à l'enseignement de l'indexation automatique. C'est-à-dire, tester la probabilité que le fait d'enseigner l'indexation automatique est dépendant de l'expérience. Les résultats du calcul de probabilité conditionnelle

montrent que plus les professeur.e.s documentalistes sont expérimenté.e.s et plus il est probable qu'il.elle.s enseignent l'indexation automatique (Figure 10 ; Annexe 5 – Contingence Q1a-Q6). Les professeur.e.s documentalistes ayant peu d'expérience (0-9 ans d'expérience) ont majoritairement répondu « NON » à la question « Enseignez-vous l'indexation automatique ? », alors que les professeur.e.s très expérimenté.e.s (20-30 ans d'expérience), ont très largement répondu « OUI » (Figure 10 ; Annexe 5 – Contingence Q1a-Q6). Pour les professeur.e.s documentalistes en milieu de carrière (10-19 ans d'expérience), il y a peu de différence entre ceux.celles qui l'enseignent et ceux.celles qui ne l'enseignent pas (Figure 10 ; Annexe 5 - Q1a-Q6).



- Hypothèse C : L'indexation automatique est enseignée uniquement dans le cadre du fonctionnement des moteurs de recherche.

Réponse C : L'indexation automatique est majoritairement enseignée dans le cadre du fonctionnement des moteurs de recherche associé aux cadres pédagogiques des SNT et de l'information-documentation.

Pour répondre à l'hypothèse selon laquelle l'indexation automatique est uniquement enseignée dans le cadre du fonctionnement des moteurs de recherche nous avons demandé aux professeur.e.s documentalistes de répondre à deux questions à choix multiples ; l'une sur le cadre des notions génériques (moteur de recherche, portail documentaire, veille et autres) et l'autre sur les cadres pédagogiques (information-documentation, EPI²⁷, AP²⁸, SNT²⁹, interdisciplinarité et autre) dans lesquels il.elle.s enseignent l'indexation automatique. Dans la très grande majorité des cas, l'indexation automatique est enseignée à travers le fonctionnement des moteurs de recherche (68%, *fréquence=0,68* ; Figure 11a ; Annexe 5 - Q1c). L'information-documentation est le cadre pédagogique privilégié par les professeur.e.s documentalistes pour enseigner l'indexation automatique (45%, *fréquence=0,45* ; Figure 11b ; Annexe 5 - Q1d). Cependant, malgré sa récente existence, le cadre pédagogique des sciences numérique et technologie (SNT) est utilisé en second pour enseigner l'indexation automatique, juste après l'information-documentation (14%, *fréquence=0,14* ; Figure 11b ; Annexe 5 - Q1d). Il y a clairement un effet du niveau scolaire d'enseignement dans ce résultat (Figure 11c ; Annexe 5 – Contingence Q1b-Q1d). L'indexation automatique, au travers de l'information-documentation, est essentiellement enseignée au niveau collège, alors que les SNT sont enseignées exclusivement au niveau Lycée (Figure 11c ; Annexe 5 – Contingence Q1b-Q1d). L'indexation automatique est enseignée en interdisciplinarité au même niveau que les SNT (15%, *fréquence=0,15* ; Figure 11b ; Annexe 5 - Q1d) et de façon quasi-égale en collège et en lycée. Il est intéressant de noter que l'accompagnement personnalisé (AP), est un cadre pédagogique peu utilisé pour enseigner l'indexation automatique (12%, *fréquence=0,12* ; Figure 11b ; Annexe 5 -

²⁷EPI : Enseignements pratiques interdisciplinaires.

²⁸AP : Accompagnement personnalisé.

²⁹SNT : Sciences numérique et technologie.

Q1d). Dans la catégorie « Autre » (7%, fréquence=0,07 ; Figure 11b ; Annexe 5 - Q1d), l'indexation automatique est enseignée dans le cadre de : l'« identité numérique » et en lien avec le « portail documentaire » (Annexe 4).

Lorsque l'indexation automatique est enseignée dans le cadre notionnel des moteurs de recherche, elle a plus de probabilité d'être enseignée dans le cadre pédagogique des SNT ($p=0,75$), de l'EPI ($p=0,67$) et de l'information-documentation ($p=0,68$; Figure 11d ; Annexe 5 – Contingence Q1c-Q1d). Le cadre pédagogique « Autre » est aussi probable que le cadre pédagogique de l'information-documentation. Il renvoie aux cadres pédagogiques de l'« éducation à » comme l'« EMI » ou l'« hygiène numérique », ou bien encore d'« autres supports horaires » et « Maths sup, maths spé » (Annexe 4). Cela montre la diversité des cadres pédagogiques possibles permettant d'enseigner l'indexation automatique. La veille est une notion très peu utilisée pour enseigner l'indexation automatique (3%, fréquence=0,03 ; Figure 11a) et lorsqu'elle l'est, c'est dans le cadre de l'interdisciplinarité ($p=0,13$; Figure 11d ; Annexe 5 – Contingence Q1c-Q1d).

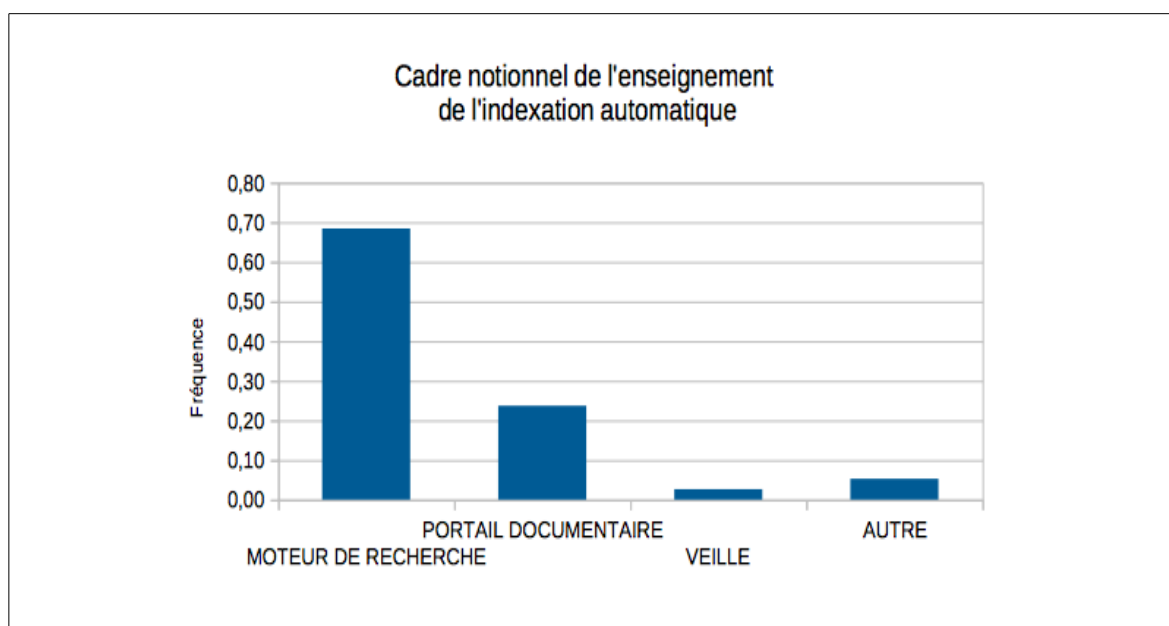


Figure 11a : Notions génériques dans lesquelles l'indexation automatique est enseignée par les professeur.e.s documentalistes. La notion générique de moteur de recherche est la plus utilisée pour enseigner l'indexation automatique.

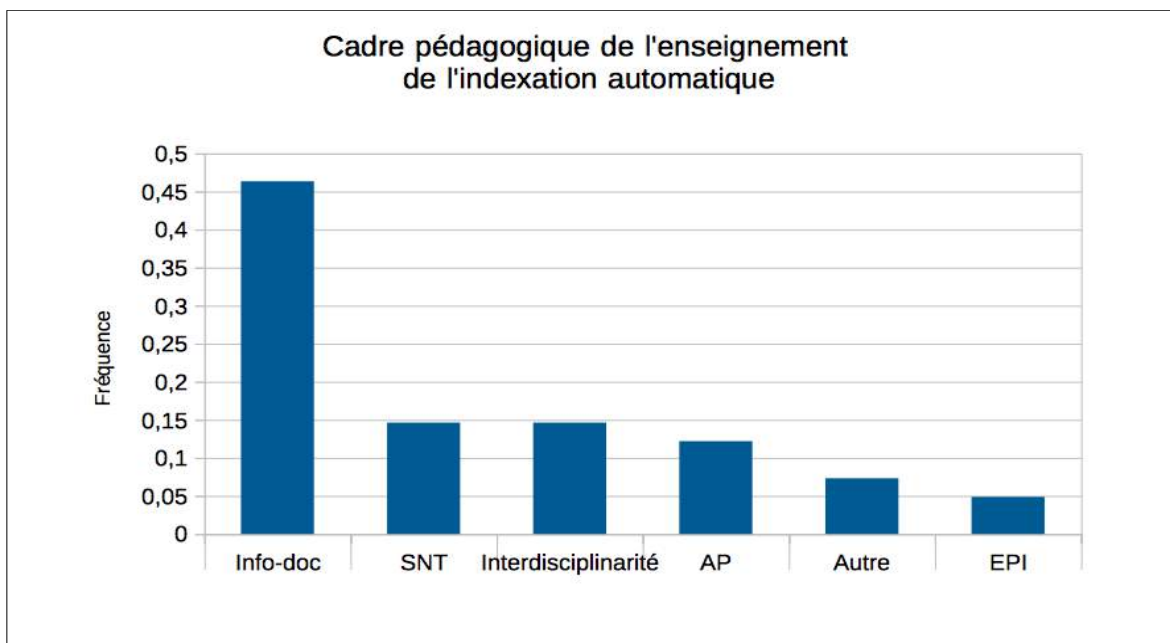


Figure 11b : Cadres pédagogiques dans lesquels l'indexation automatique est enseignée par les professeur.e.s documentalistes.

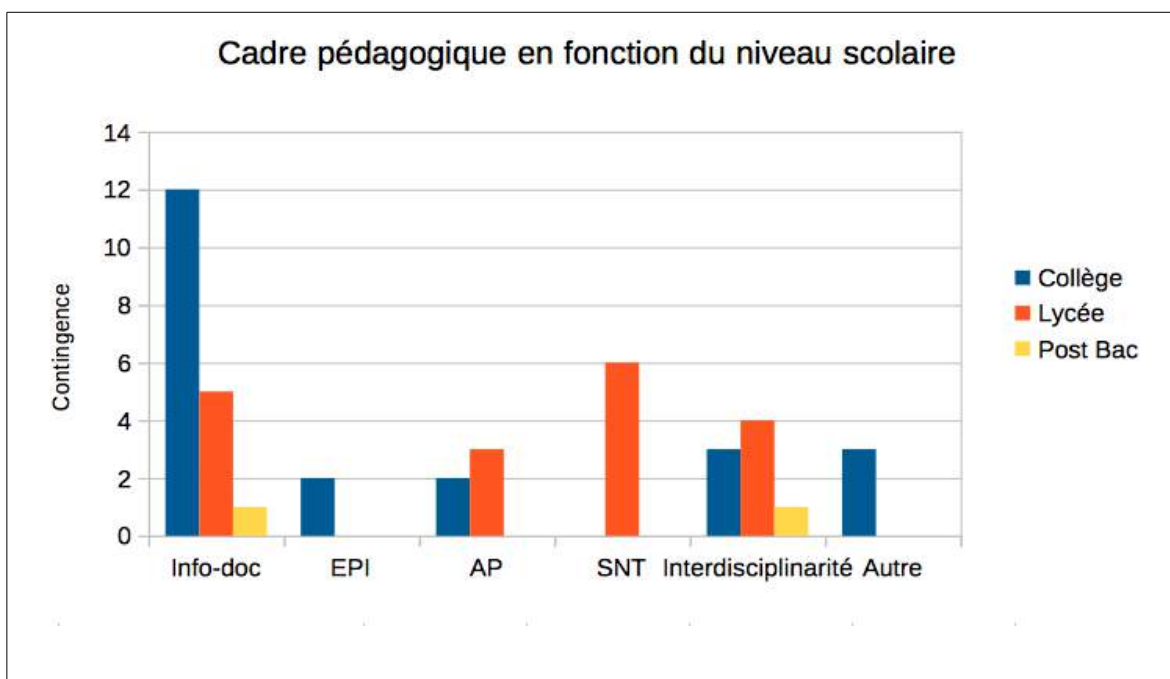
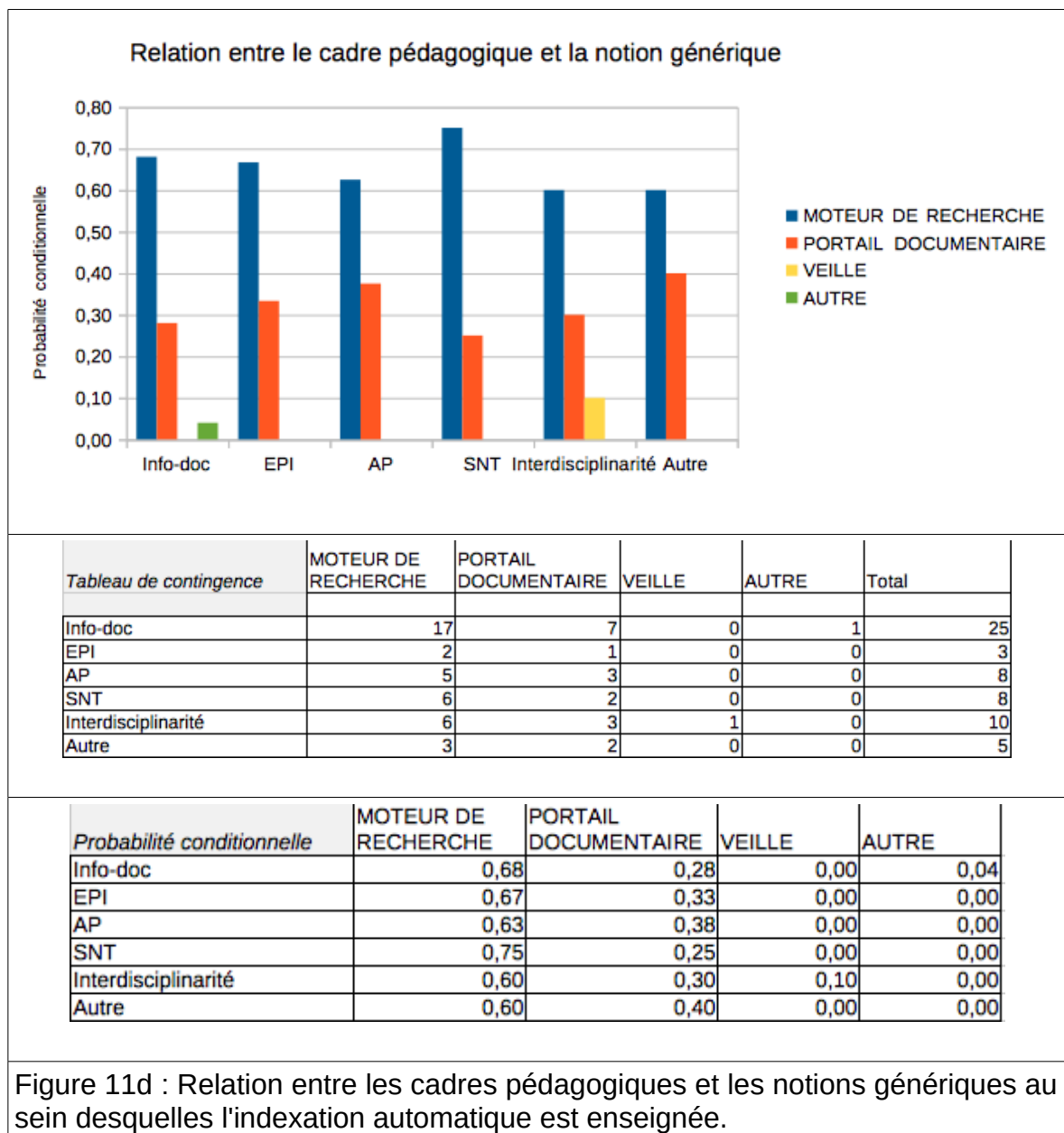


Tableau contingence Cadre pédagogique	Niveaux d'enseignement		
	Collège	Lycée	Post Bac
Info-doc	12	5	1
EPI	2	0	0
AP	2	3	0
SNT	0	6	0
Interdisciplinarité	3	4	1
Autre	3	0	0

Figure 11c : Relation entre le cadre pédagogique dans lequel l'indexation automatique est enseignée et le niveau scolaire (collège, lycée ou *post-bac*).



- Hypothèse D : Les professeur.e.s documentalistes créent leurs propres séances pédagogiques pour aborder la notion d'indexation automatique.

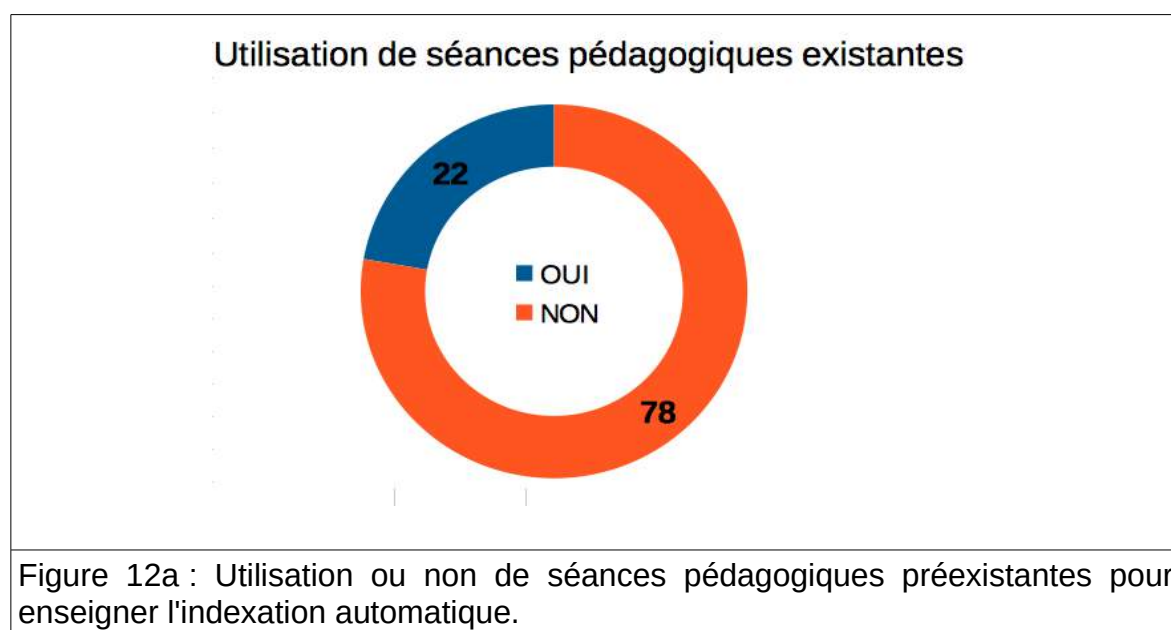
Réponse D : Environ 80% des professeur.e.s documentalistes créent leurs propres séances pédagogiques pour aborder la question de l'indexation automatique.

À la question « Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ? », 78% des professeur.e.s documentalistes ont répondu « Non »

(Figure 12a ; Annexe 5 - Q4). La raison principale évoquée est qu'il.elle.s ont une préférence pour l'utilisation de leur propre séance qui confère une meilleure maîtrise du déroulé de la séance et de la notion (*fréquence=0,32* ; Figure 12b ; Annexes 4 et 5 - Q4). La deuxième raison apportée est que l'indexation automatique est abordée de façon trop succincte pour être enseignée dans une séance à part entière (*fréquence=0,26* ; Figure 12b ; Annexes 4 et 5 - Q4). Les autres raisons évoquées sont le manque de temps pour aborder la notion, le fait qu'elle puisse ne pas être maîtrisée (voire non connue pour une réponse), et le fait qu'elle fasse le choix de ne pas être enseignée (Figure 12b ; Annexes 4 et 5 - Q4).

Pour les sept professeur.e.s documentalistes qui ont répondu utiliser des séances pédagogiques existantes pour réaliser les leurs sur l'indexation automatique, l'origine de ces ressources est majoritairement professionnelle (43%, *fréquence=0,43* ; Annexe 5 - Q4).

L'hypothèse selon laquelle il existerait une relation entre les méthodes pédagogiques et les cadres pédagogiques dans lesquels l'indexation automatique est enseignée ne peut pas fondamentalement être testée à cause du très faible nombre de données (Annexe 5 – Contingence Q1d-Q5). Il semblerait que les méthodes actives soient surtout utilisées en information-documentation dans 60% des réponses (Annexe 5 – Contingence Q1d-Q5).



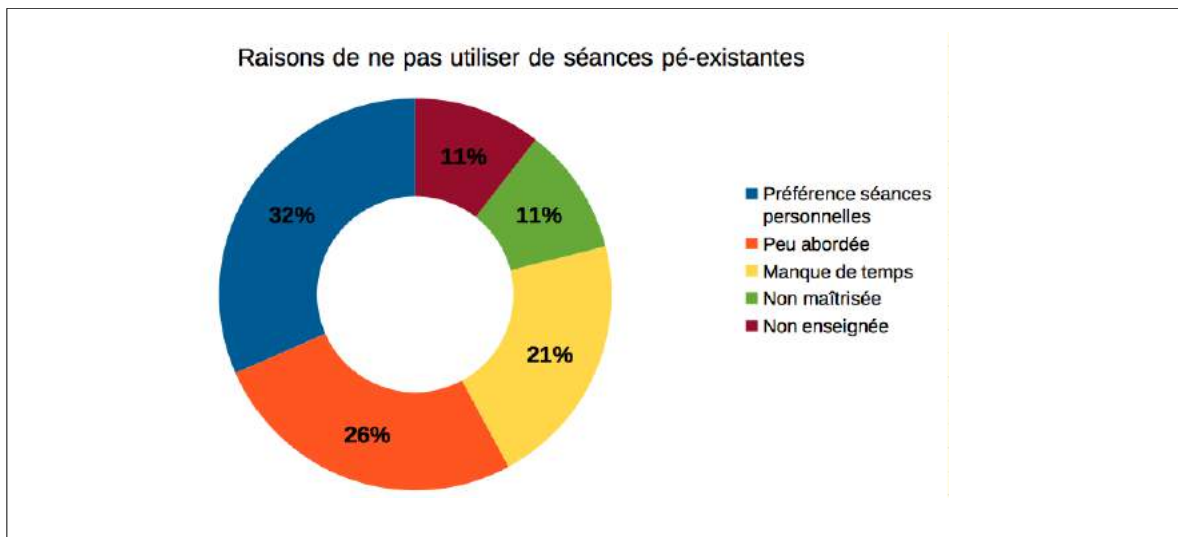


Figure 12b : principalement pour une préférence pour la création de séances personnelles. *N.B. : Les résultats ont été obtenus à partir des fréquences totales des réponses.*

- Hypothèse E: Les professeur.e.s documentalistes n'abordent pas systématiquement la notion d'indexation en amont de la notion d'indexation automatique.

Réponse E: À hauteur de 56% de leurs réponses, les professeur.e.s documentalistes n'abordent pas la notion d'indexation en amont de la notion d'indexation automatique.

En accord avec l'hypothèse initiale stipulant que les professeur.e.s documentalistes n'abordent pas systématiquement la notion d'indexation en amont de la notion d'indexation automatique, nous pouvons montrer que c'est le cas pour 56% des professeur.e.s documentalistes ayant répondu (Figure 13a ; Annexe 5 - Q2). Cependant, les raisons pour lesquelles il.elle.s ne le font pas sont principalement les mêmes que pour ceux.celles qui n'enseignaient pas l'indexation automatique au départ. Par conséquent, ce résultat est à nuancer car les personnes ayant répondu « NON » ont été probablement influencées par le fait qu'il.elle.s aient répondu « NON » à la question sur l'enseignement de l'indexation automatique initiale. Toutefois, certains répondant.e.s ont indiqué qu'il.elle.s n'enseignaient pas l'indexation en amont, mais en même temps que l'indexation automatique (Annexe 4)

Les professeur.e.s documentalistes ayant fait le choix d'enseigner l'indexation en amont de l'indexation automatique (44% ; Figure 13b ; Annexe 5 -

Q2), l'ont fait pour comprendre le processus d'indexation, qu'il soit historique et/ou fonctionnel (algorithmie) (*fréquence=0,42* ; Figure 13b ; Annexes 4 et 5 -Q2). Au même titre, il.elle.s considèrent l'intégration de la notion d'indexation automatique dans le cadre notionnel de l'indexation comme essentielle pour comprendre la notion d'indexation automatique (*fréquence=0,42* ; Figure 13b ; Annexe 5 -Q2). Dans une moindre mesure, l'indexation humaine est mise en regard de l'indexation automatique ou en lien avec l'index du livre (*fréquence=0,12* respectivement ; Figure 13b ; Annexes 4 et 5 - Q2).

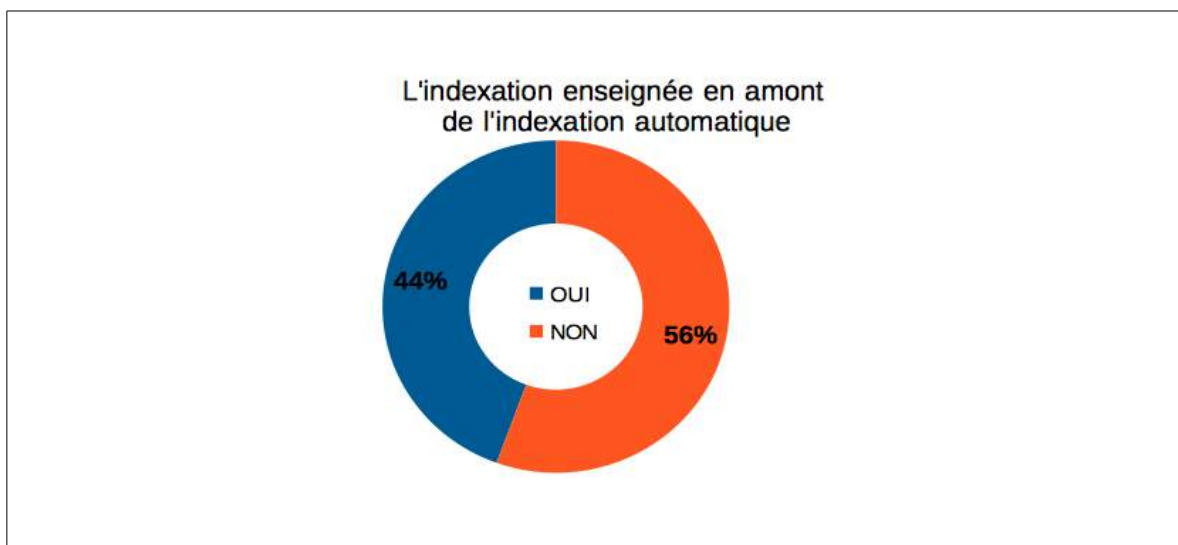


Figure 13a : Enseignement de l'indexation en amont de la notion d'indexation automatique. À 56%, les professeur.e.s documentalistes n'enseignent pas l'indexation en amont de la notion d'indexation automatique.

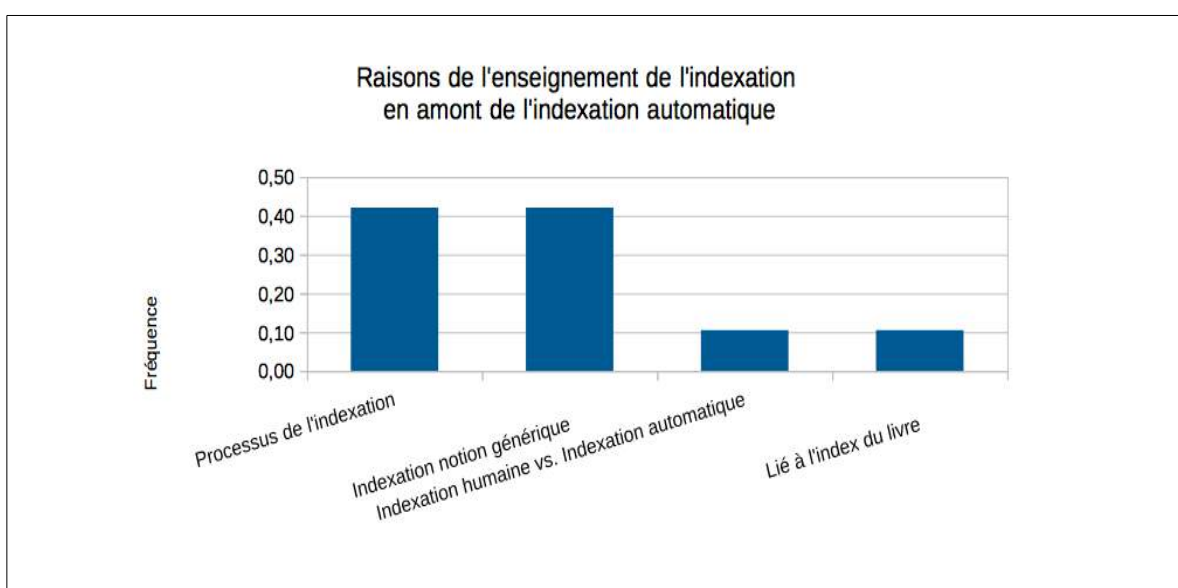


Figure 13b : Raisons pour lesquelles les professeur.e.s documentalistes enseignent l'indexation en amont de l'indexation automatique.

- Hypothèse F: Les professeur.e.s documentalistes n'utilisent pas les *infoboxes* dans leurs enseignements.

Réponse F: Bien que la quasi-totalité des professeur.e.s documentalistes ait remarqué l'existence des encarts informationnels sur une page de résultats, il.elle.s sont moins de 30% à les utiliser dans leurs enseignements.

Près de la totalité des répondant.e.s ont remarqué l'existence des encarts informationnels sur une page de résultats d'un moteur de recherche (98% ; Figure 14a ; Annexe 5 - Q3a). L'encart informationnel le plus repéré est celui de Wikipédia (45%), devant les encarts définitionnels (32% ; Figure 14b ; Annexe 5 - Q3a). Ce sont les encarts informationnels d'images ou vidéos les moins remarqués parmi les réponses données (23% ; Figure 14a ; Annexe 5 - Q3a). Malgré ce repérage massif de l'existence des *infoboxes*, les professeur.e.s documentalistes ne les utilisent que dans moins de 30% des cas (Figure 14c ; Annexe 5 - Q3b).

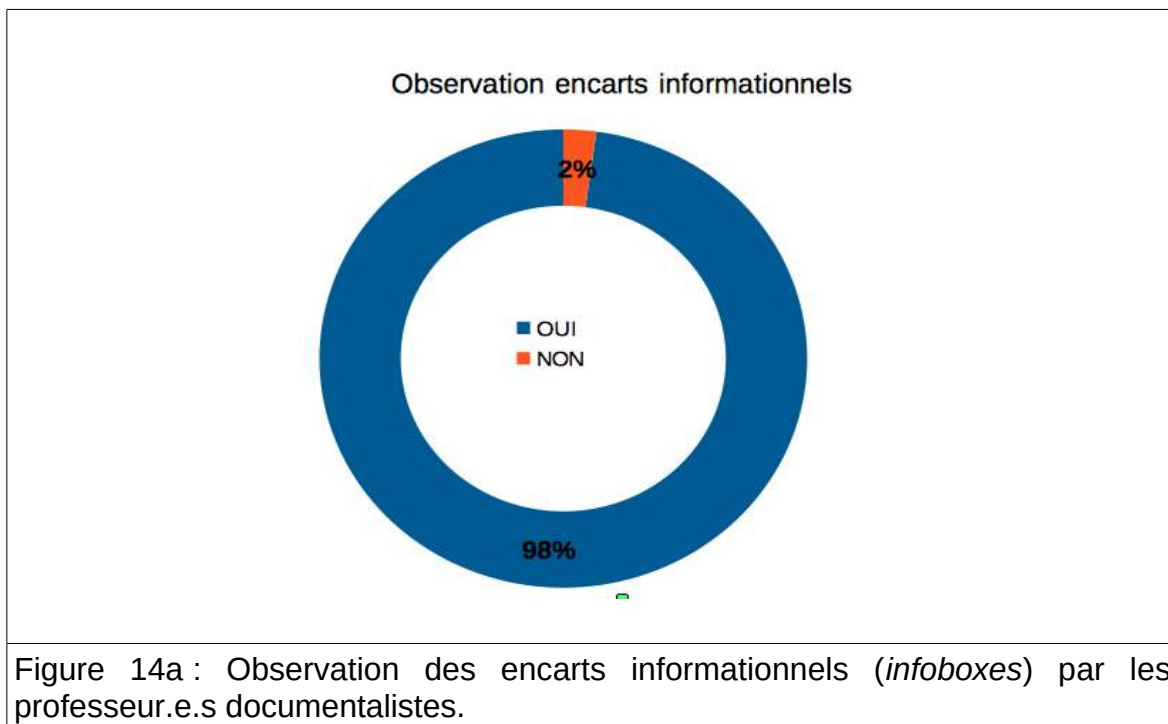


Figure 14a : Observation des encarts informationnels (*infoboxes*) par les professeur.e.s documentalistes.

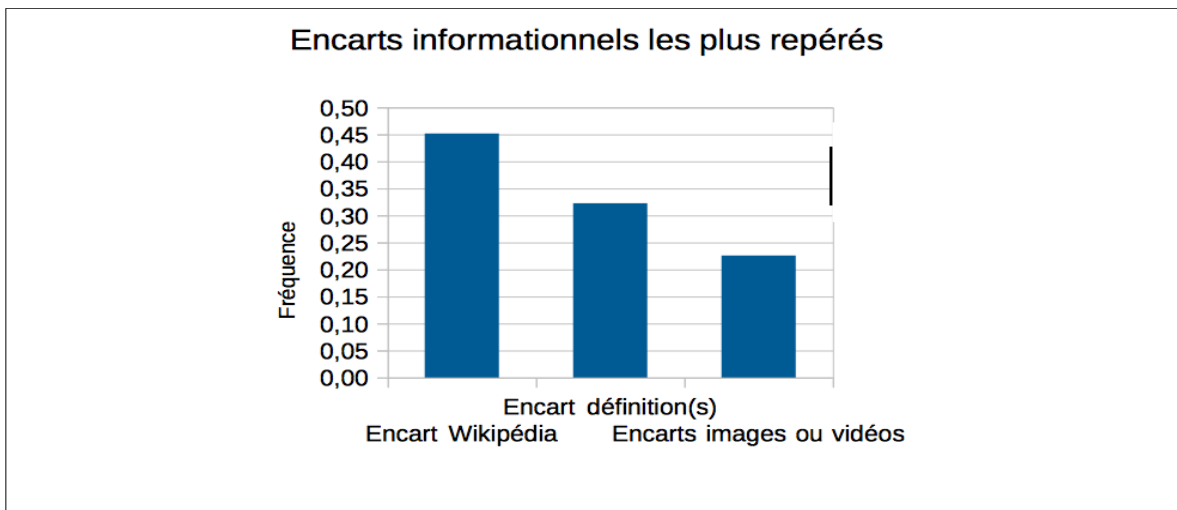


Figure 14b : Encarts informationnels les plus repérés sur une page de résultats de moteur de recherche.

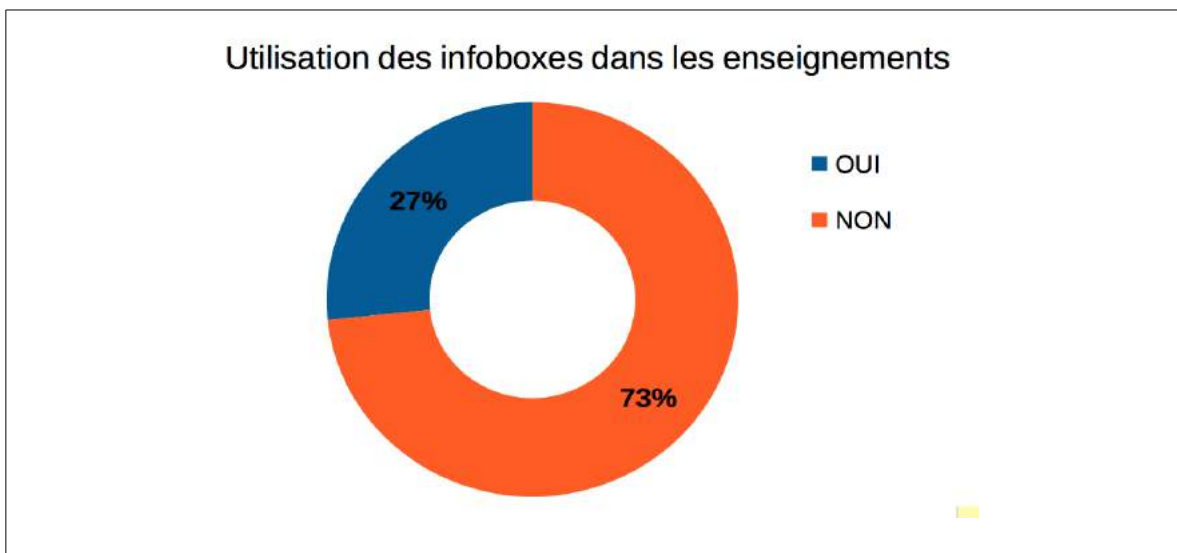


Figure 14c : Utilisation des *infoboxes* dans les enseignements par les professeur.e.s documentalistes.

- Hypothèse G : Lorsqu'ils.elles l'utilisent, ils.elles n'abordent pas la question de l'indexation automatique.

Réponse G : À travers leurs réponses, 54% des professeur.e.s documentalistes utilisant les *infoboxes* dans leurs enseignements souhaitent que les élèves retiennent des notions liées à l'évaluation de l'information. La compréhension du fonctionnement des *infoboxes* ne représente que 23% des réponses.

À la question « Que doivent retenir les élèves des *infoboxes* ? », les professeur.e.s documentalistes les utilisant dans leurs enseignements ont majoritairement proposé des notions liées à l'évaluation de l'information, comme la source, la crédibilité et/ou l'origine de l'information (ex. publicité ou non) (54% des réponses ; Figure 15a ; Annexes 4 et 5 - Q3c). À hauteur de 23% des réponses, les professeur.e.s documentalistes souhaitent que les élèves retiennent des *infoboxes* leur fonctionnement, c'est-à-dire l'algorithmie sous-jacente et/ou l'intelligence artificielle qui y est incluse (Figure 15a ; Annexe 4). Dans une même proportion (23%), les professeur.e.s documentalistes utilisent les encarts informationnels afin que les élèves comprennent le fonctionnement des outils de recherche informationnelle comme les mots-clés, le type d'information, etc. (Figure 15a ; Annexe 4).

Pour les professeur.e.s documentalistes qui n'enseignent pas les *infoboxes*, 61% d'entre eux.elles ne le font pas par manque de temps, mais 11% ont envisagé l'éventualité de le faire, et ce même après avoir répondu à ce questionnaire ! (Figure 15b ; Annexe 4). Cela montre l'intérêt que pourraient porter les professeur.e.s documentalistes à cette question même s'ils ne le font pas dans l'immédiat. Dans une moindre mesure, les professeur.e.s documentalistes n'intègrent pas les *infoboxes* dans leurs enseignements car cela ne fait pas partie des projets des collègues avec lesquels il.elle.s enseignent (25% ; Figure 15b ; Annexes 4 et 5 - Q3c). Enfin, le fait que l'indexation automatique puisse être un sujet complexe pour les élèves, et même pour certain.e.s professeur.e.s documentalistes eux.elles-mêmes est une des raisons limitantes pour enseigner avec les *infoboxes* dans 7% des cas (Figure 15b ; Annexes 4 et 5 - Q3c).

Contrairement à l'enseignement de l'indexation automatique, il n'existe pas de relation entre l'enseignement des *infoboxes* et l'expérience (Figure 15c ; Annexe 5 – Contingence Q3b-Q6). Les *infoboxes* sont abordées aussi bien par les professeur.e.s documentalistes très expérimentés et expérimentés que par ceux.celles moins expérimentés (Figure 15c ; Annexe 5 – Contingence Q3b-Q6).

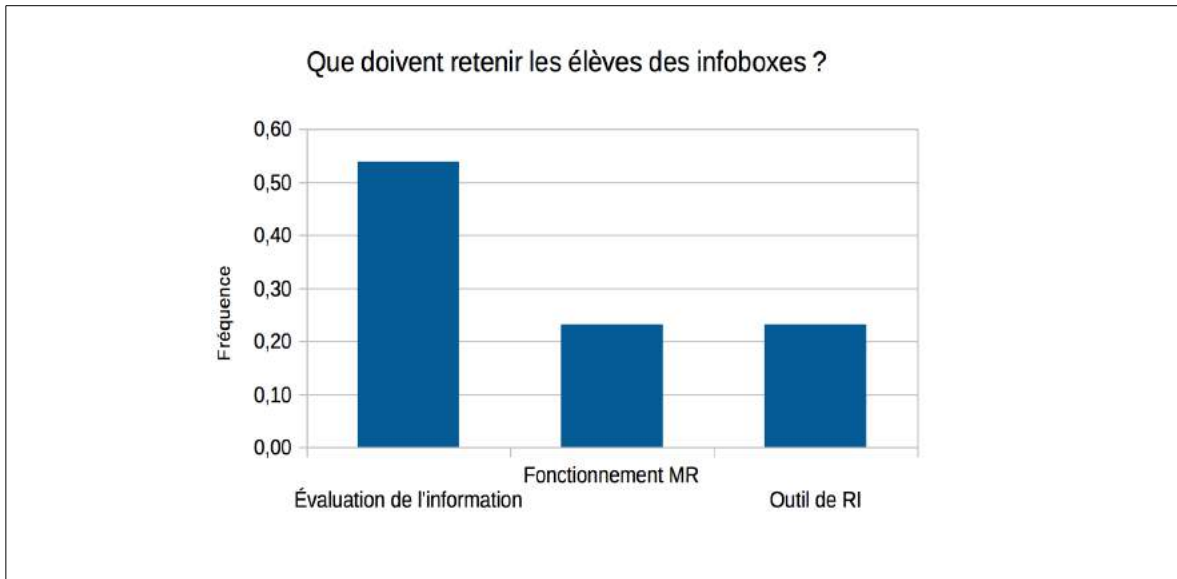


Figure 15a : Notions que doivent retenir les élèves des *infoboxes*.

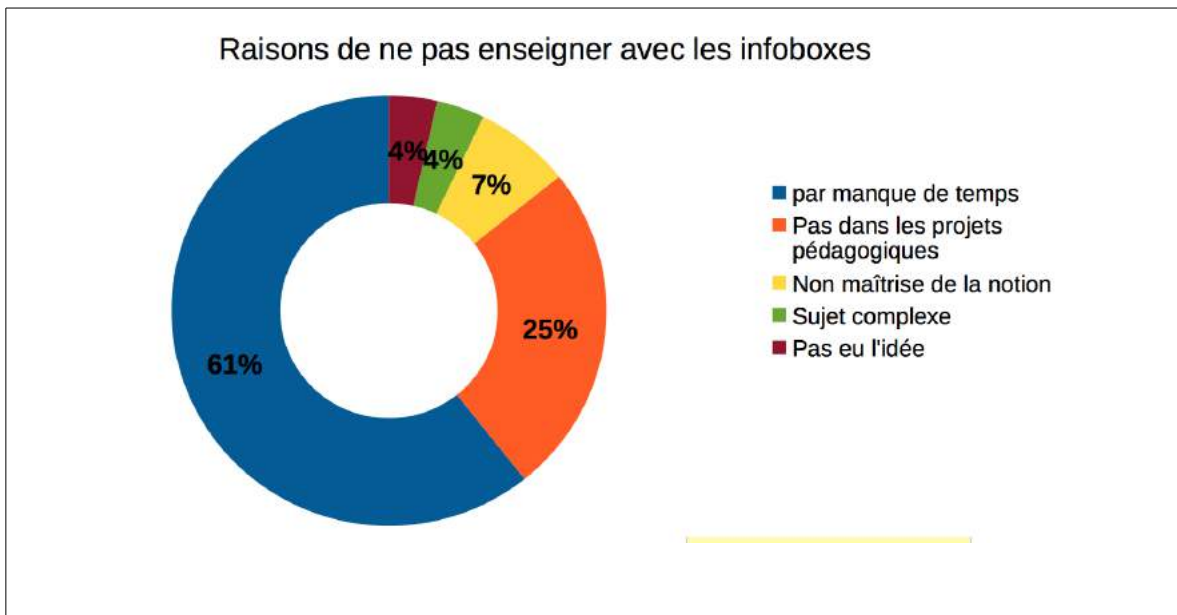
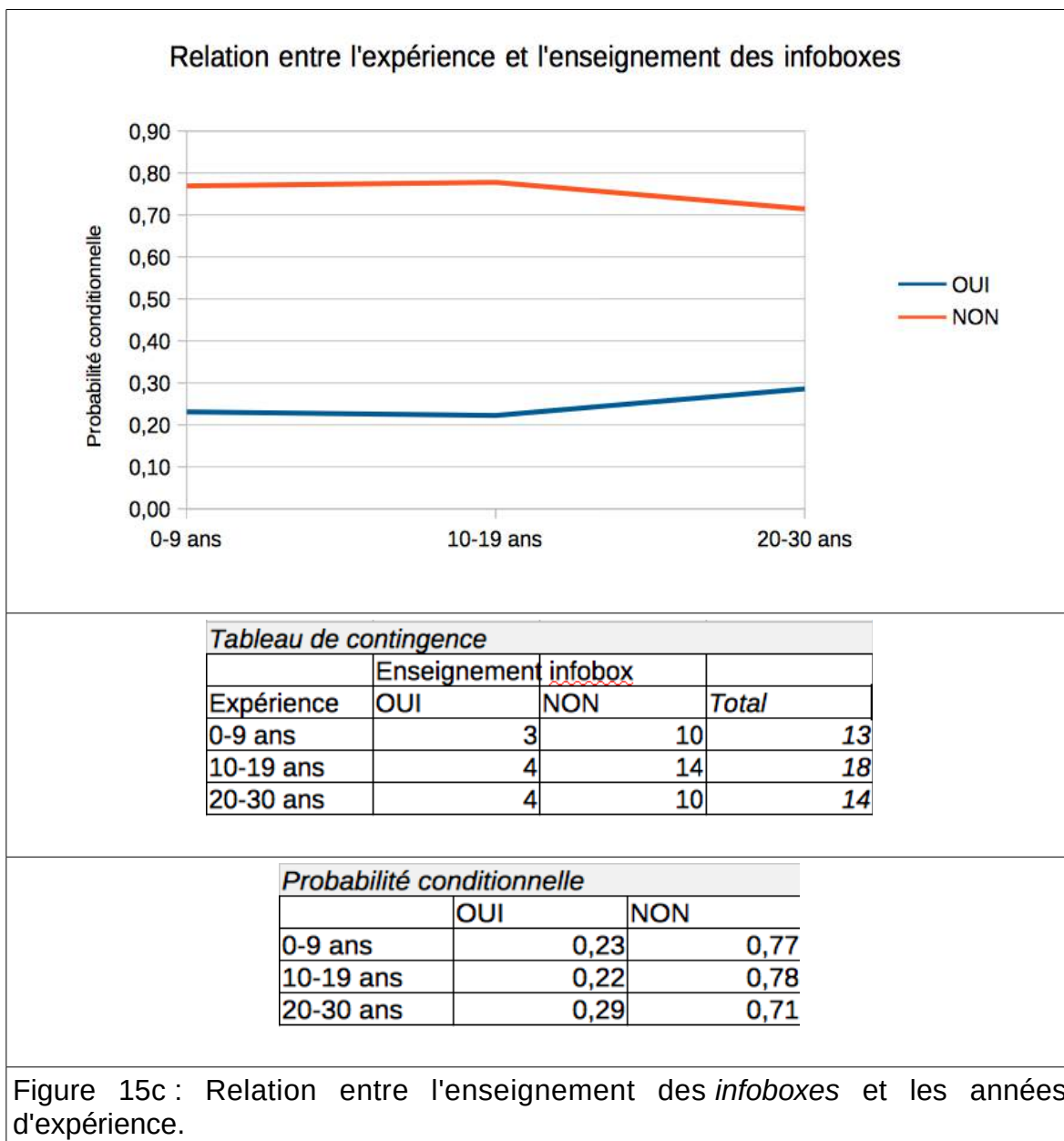


Figure 15b : Raisons pour lesquelles les professeur.e.s documentalistes n'enseignent pas avec les *infoboxes*.



3.2 Discussion des résultats

Historiquement, l'indexation automatique est un procédé documentaire avant d'être une partie intégrante du fonctionnement des moteurs de recherche. Il est possible que cet aspect évolutif des procédés d'indexation explique que ce soient les professeur.e.s documentalistes les plus expérimentés qui enseignent majoritairement l'indexation automatique (20-30 ans d'expérience ; Figure 10 ; Annexe 5 - Q6). Bien que seulement deux professeur.e.s documentalistes de notre panel disent enseigner l'indexation en même temps que l'indexation automatique, 44% disent enseigner l'indexation en amont de l'indexation

automatique. (Figure 13a ; Annexe 4). 44% d'entre eux le font pour aider les élèves à comprendre le processus d'indexation (Figure 12b ; Annexe 4). C'est-à-dire, comprendre d'où vient la notion d'indexation dans son aspect historique, conceptuel et/ou fonctionnel. La dimension historique de la notion d'indexation n'est pas sans rappeler le programme de SNT en seconde expliquant l'antériorité de l'indexation humaine à l'indexation automatique. Dès lors, est-il préférable d'enseigner l'indexation en amont, ou de façon synchrone, de l'indexation automatique ? Les résultats de notre étude montrent que la majorité des professeur.e.s documentalistes ne l'estiment pas nécessaire à 56% (Figure 13a ; Annexes 4 et 5 - Q2). Cependant, il est parfois important de replacer la notion dans son contexte historique ou notionnel pour la faire comprendre aux élèves. Cela dépend sûrement aussi des élèves eux-mêmes. Quoi qu'il en soit, nous observons dans notre étude que les professeur.e.s documentalistes utilisent majoritairement l'entrée des moteurs de recherche pour aborder la notion d'indexation automatique (68% ; Figure 11a ; Annexes 4 et 5 - Q1c), et ce, dans tous les cadres pédagogiques du secondaire (Figure 11d ; Annexe 5 – Contingence Q1c-Q1d). Ce résultat s'explique par le fait que le moteur de recherche est un outil familier des élèves qui permet de les sensibiliser, voire les mobiliser et faciliter l'émergence des représentations de l'indexation automatique. C'est la base pour aller vers une situation-problème dans la pédagogie active (Astolfi, 2008a). Bien que l'entrée par le portail documentaire soit un outil utilisé par les professeur.e.s documentalistes pour aborder la notion d'indexation automatique, elle reste minoritaire par rapport à l'entrée par les moteurs de recherche (23% ; Figure 11a ; Annexe 4). Pour 20% des réponses, l'enseignement des moteurs de recherche est connexe à celui du portail documentaire, notamment pour la différentier de l'indexation automatique de l'indexation humaine (Annexe 4).

Le fait que nous n'ayons pas trouvé beaucoup de séances pédagogiques en ligne évoquant l'indexation automatique (Annexe 2), n'est pas révélateur de son enseignement par les professeur.e.s documentalistes. En effet, dans notre panel, 57% des professeur.e.s documentalistes enseignent l'indexation automatique (Figure 9a ; Annexe 5 - Q1a) et 78% d'entre eux n'utilisent pas de séances pédagogiques au préalable (Figure 12a ; Annexe 4), et ce, principalement par préférence pour la construction de séances pédagogiques

personnelles (32% des réponses ; Figure 11b ; Annexe 5 -Q1d). L'autre raison évoquée est que l'indexation automatique est rarement enseignée pour elle-même (27% des réponses ; Figure 12b ; Annexe 4), donc elle ne justifie pas la création de séances pédagogiques à part entière lorsqu'elle est simplement évoquée. Nous pouvons supposer que les personnes qui enseignent l'indexation automatique ont davantage répondu au questionnaire. Cependant, la courte majorité des professeur.e.s documentalistes qui enseignent l'indexation automatique ayant répondu au questionnaire (57% ; Figure 9a ; Annexe 5 - Q1a), nous montre un relatif équilibre entre ceux.celles qui l'enseignent et ceux.celles qui ne l'enseignent pas. Par conséquent, malgré un nombre très important de séances pédagogiques traitant des moteurs de recherche (cf. Rousseau, 2015), le faible nombre de séances pédagogiques en ligne traitant de l'indexation automatique n'est pas le reflet de son enseignement par les professeur.e.s documentalistes. Ainsi, notre étude montre qu'il existe une réelle implication des professeur.e.s documentalistes pour l'enseignement de cette notion.

Il est à noter aujourd'hui que l'indexation automatique est majoritairement enseignée en seconde, et pour ce niveau, en sciences numérique et technologie (SNT) et en information-documentation (Figures 9b et 11c ; Annexe 5 – Q1b et Q1d). Dans les séances pédagogiques que nous avons récoltées, qui datent du début des années 2000 et qui intègrent la notion d'indexation automatique pour le niveau seconde, le cadre pédagogique utilisé était souvent l'accompagnement personnalisé (AP). Il semblerait que ce soit moins le cas aujourd'hui. Notre étude montre que l'AP est le troisième cadre pédagogique au sein duquel l'indexation automatique est enseignée, derrière les SNT et l'information-documentation (Figure 11c ; Annexe 5 -Q1d). Même si une majorité de professeur.e.s documentalistes en lycée n'enseignent pas en SNT (54,7% ; APDEN, 2019), et malgré son entrée à la rentrée 2019, nous observons que les professeur.e.s documentalistes de notre panel se sont emparé.e.s de ce cadre pédagogique, et il.elle.s y enseignent l'indexation automatique à travers les moteurs de recherche (Figure 11c ; Annexe 5 - Q1d). Le potentiel de ce cadre pédagogique pour l'enseignement de l'indexation automatique par les professeur.e.s documentalistes est indéniable. Il ressort de l'étude de l'APDEN que ce sont aux professeur.e.s documentalistes d'être force de propositions de séances pédagogiques pour

faciliter leur intégration dans ce cadre pédagogique au lycée général (APDEN, 2019).

Au collège, même si l'indexation automatique y est moins enseignée, elle est majoritairement enseignée en 6^e (Figure 9b ; Annexe 5 - Q1b). Nous avons mis en évidence, à partir des travaux de Pascal Duplessis (2019a), que pour le cycle 3, l'indexation automatique pouvait être enseignée à neuf moments du programme à travers la notion organisatrice de la recherche informationnelle. De plus, parmi les séances pédagogiques que nous avons récoltées, plusieurs relèvent du niveau 6^e (Annexe 2). Il est possible également que les professeur.e.s documentalistes qui enseignent au collège soient plus impliqués au niveau 6^e, notamment à travers l'accueil des élèves, puis leur intégration dans l'établissement par la suite. Les situations sont très disparates de l'une à l'autre, mais quoi qu'il en soit, s'il est un niveau à privilégier au collège pour enseigner l'indexation automatique, c'est le niveau 6^e. Nos résultats indiquent que le cadre pédagogique de l'information-documentation est privilégié pour enseigner l'indexation automatique au collège (Figure 11c ; Annexe 5 - Q1d). Le texte « Vers un *curriculum* en information-documentation » de l'APDEN préconise une approche de l'indexation via la recherche informationnelle au collège, et notamment par la comparaison entre l'indexation documentaire (contrôlée ou libre) et l'indexation automatique (APDEN, 2016a). De même, l'étude de l'emprise des notions info-documentaires des cycles 3 et 4 réalisée par Pascal Duplessis (2019), montre que la notion organisatrice de recherche informationnelle est privilégiée (15 fois pour les deux cycles). Notre étude mentionne quelques raisons pour lesquelles les professeur.e.s documentalistes enseignent l'indexation automatique, au collège et au lycée, notamment, à travers les moteurs de recherche ou le portail documentaire. Cependant, nous pouvons montrer que l'étude des *infoboxes* est une entrée possible, mais pas majoritaire, pour aborder les outils de la recherche informationnelle, comme les mots-clés par exemple (Figure 15a ; Annexe 4). De plus, le fait de replacer la notion d'indexation en amont, ou de façon synchrone, de l'indexation automatique permet d'entrer dans cette notion via la recherche informationnelle. En effet, parmi les raisons pour lesquelles les professeur.e.s documentalistes enseignent l'indexation en amont de l'indexation automatique, il ressort de la catégorie majoritaire « Processus d'indexation », l'étude par mots-clés notamment (Figure 13b ; Annexes 4 et 5 - Q2). Dans une

moindre mesure, la comparaison entre l'indexation humaine et l'indexation automatique est évoquée, de même que l'entrée par l'index du livre (Figure 13b ; Annexes 4 et 5 - Q2). Cependant, la comparaison entre index du livre, indexation humaine et indexation automatique est un moyen simple d'évoquer la notion d'indexation automatique avec des élèves de collège pour comprendre « *l'intérêt d'indexer* » comme l'a écrit un.e répondant.e à notre questionnaire (Annexe 4). Dans une séance pédagogique publiée par Marie Guillet en 2013 à destination des classes de 6^e, elle évoque l'intérêt conceptuel de la comparaison entre indexation humaine et indexation automatique (Guillet, 2013). Ainsi, l'enseignement de l'indexation automatique au collège est principalement préconisé via la recherche informationnelle à travers l'étude des mots-clés, et de la comparaison entre l'indexation humaine et automatique pour en comprendre le fonctionnement.

Avec les nouvelles avancées du *Web*, le fonctionnement des moteurs de recherche est devenu de plus en plus perfectionné et le processus d'indexation automatique également. Les encarts informationnels, ou *infoboxes*, sont le reflet de ces dernières évolutions. Bien que les professeur.e.s documentalistes aient, pour la grande majorité (98% ; Figure 14a), observé ces *infoboxes* sur la page de résultats des moteurs de recherche, seuls 27% d'entre eux enseignent à partir de ces encarts informationnels (Figure 14c). Nous avons noté que les *infoboxes* représentaient un intérêt en information-documentation pour mieux comprendre le fonctionnement des moteurs de recherche, pour faire prendre conscience aux élèves l'origine et l'évaluation de l'information, ainsi que pour apprendre la notion d'indexation automatique pour elle-même. Il est intéressant de noter que les professeur.e.s documentalistes qui enseignent les *infoboxes* souhaitent également que les élèves retiennent des notions liées à l'origine et à l'évaluation de l'information comme la source et la crédibilité (54% des réponses, soit 7/13 répondant.e.s ; Figure 15a ; Annexes 4 et 5 - Q3c), mais aussi qu'ils comprennent le fonctionnement des moteurs de recherche (algorithmie, intelligence artificielle) (23% des réponses ; Figure 15a ; Annexes 4 et 5 - Q3c). Nous n'avons pas envisagé que l'enseignement des *infoboxes* puisse être utilisé pour mettre en avant les outils liés à la recherche d'information (mots-clés, comparaison des types d'information, etc.) (23% des réponses respectivement ; Figure 15a ;

Annexes 4 et 5 - Q3c). En même temps, ces outils, comme les mots-clés, sont des notions corrélées à l'indexation, et donc à l'indexation automatique.

Il apparaît dans notre étude que la raison principale pour laquelle les professeur.e.s documentalistes n'enseignent pas l'indexation automatique (38% des réponses ; Figure 9c ; Annexe 5 - Q1b(suite)), ni les *infoboxes* (61% des réponses ; Figure 15b ; Annexe 5 - Q3c), est le manque de temps, même si certain.e.s d'entre eux.elles aimeraient le faire. La deuxième raison pour laquelle les professeur.e.s documentalistes n'enseignent pas l'indexation automatique est la complexité de la notion, et surtout au niveau collège (33% des réponses ; Figure 9c ; Annexe 4 et 5 - Q1b(suite)). De plus, pour 21% des réponses, les professeur.e.s documentalistes avouent ne pas maîtriser la notion eux.elles-mêmes. Nous espérons que notre travail de vulgarisation avec l'infographie (Figure 1), et de didactisation de la notion (Tableau 4) pourra être utile aux professeur.e.s ne se sentant pas à l'aise avec cette notion et qui souhaiteraient l'enseigner. Pour les professeur.e.s documentalistes qui considèrent, à juste titre, que l'indexation automatique est une notion difficile à aborder avec les élèves, et notamment au niveau collège, nous proposons un exemple de progression d'enseignement des *infoboxes* pour comprendre le fonctionnement de l'indexation automatique en comparaison avec l'indexation humaine.

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

Dans un premier temps il est important de connaître les représentations des élèves sur l'origine de l'information dans ces encarts informationnels, ou *infoboxes*. D'où vient l'information dans un encart Wikipédia de la page de résultats ? Il est peu probable que les élèves répondent de « la page Wikipédia », mais plutôt de « Google » ou d'un autre moteur de recherche, d'« Internet » ou du « Web ». Or, l'origine de l'information dans un encart Wikipédia de la page de résultats, ou *infobox* d'ontologie, est principalement issue de la page Wikipédia elle-même, mais également de la centralisation des données issues d'autres sites (Tableau 1). En analysant avec les élèves l'*infobox* d'ontologie, et en comparant avec la page originelle de Wikipédia, ils se rendront compte que l'information n'est pas exactement la même. Par conséquent, les élèves se retrouveront face à une

situation-problème, ou un conflit cognitif, dans laquelle leur représentation initiale paraissant évidente ne l'est pas autant qu'ils auraient pu le penser (Minder, 2007).

Le conflit cognitif désigne le déséquilibre entre ce que l'on sait et ce que l'on constate (Minder, 2007). Pour Jean-Pierre Astolfi, ancien didacticien des sciences, le conflit cognitif est une pédagogie incontournable car la tension qu'il crée est le « *moteur de l'élaboration de nouvelles structures cognitives* » (Astolfi, 1997). Par conséquent, en recherchant l'origine de l'information à partir des *infoboxes* d'ontologie Wikipédia avec les élèves, il sera possible d'aborder la question de l'**indexation automatique**. Il sera même possible de comparer la **notion d'indexation manuelle/humaine**, avec celle de l'**indexation automatique**.

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 pages de résultats (Figure 16). Par simple comparaison des contenus de ces *infoboxes*, il est possible d'identifier avec les élèves, qu'elles ne sont pas construites de la même façon. Dans le cas des *infoboxes* Wikipédia, il est possible de déterminer que **l'indexation est manuelle/humaine** car dans cette encyclopédie collaborative tout le monde peut contribuer, et il existe une page expliquant la réalisation de ces *infoboxes* par complétion d'un formulaire pré-formaté au choix³⁰. En ce qui concerne les *infoboxes* des pages de résultats de Qwant et de Google, la comparaison des premières lignes introductives permet de déterminer l'origine de l'information (Figure 16). Dans le cas de l'*infobox* de la page de résultats de Qwant, les premières lignes sont directement indexées sur la page Wikipédia (Figure 16). Cette indexation correspond à une indexation automatique dite « classique » en ce sens qu'elle renvoie à la segmentation de la page Wikipédia permettant d'identifier le texte du début et de le copier dans l'*infobox* de la page de résultats. (Figure 1). Dans le cas du moteur de recherche Google, les premières lignes de l'*infobox* ne renvoient pas directement à la page Wikipédia et ne renvoient pas à une page en particulier. Par conséquent, il est possible d'en déduire que l'**indexation automatique** ici résulte d'une « **reformulation** » de types de l'analyse syntaxique basée sur les réseaux de neurones (Figure 1). Nous retiendrons avec les élèves de début de collège (niveau débutant), que dans ce cas, il y a « réécriture » du texte par les opérations informatiques du moteur de

³⁰Aide:Infobox. (2019). In *Wikipedia*. <https://fr.wikipedia.org/wiki/Aide:Infobox>

recherche. (Figure 17). Avec les élèves de niveau intermédiaire (fin de collège, début de lycée), nous retiendrons qu'il y a « reformulation » du contenu des pages web indexées par les algorithmes des moteurs de recherche (Figure 17). Pour le niveau avancé, ce type d'exemple permet d'aborder la notion d'« analyse syntaxique » dans l'**indexation automatique** et de brièvement aborder les méthodes basées sur les réseaux de neurones avec les élèves spécialisés en sciences numérique et technologie par exemple (Figure 17).

Pour comprendre le processus de centralisation des données en pointant vers d'autres pages web, il est possible d'utiliser *l'infobox* de la page de résultats de Google qui reprend une information de *l'infobox* Wikipédia (Figure 16). En effet, le fait que seule une partie de *l'infobox* Wikipédia soit reprise dans *l'infobox* 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. Cet exemple permet de montrer de façon explicite l'**indexation automatique** avec un niveau débutant, ou intermédiaire (Figure 17). Il permet également de comprendre la centralisation des données pour un niveau avancé en comparant la page Wikipédia avec les éléments descriptifs de *l'infobox* de la page de résultats comme la « météo » ou la date de l'actualisation de la population qui ne sont pas les mêmes. Les *infoboxes* des pages de résultats permettent également d'aborder la notion d'agrégation de contenu, mais c'est un processus d'indexation complexe qu'il est difficile d'aborder avec des élèves, même pour un niveau avancé (Figure 17). Cette notion peut juste être évoquée en expliquant que cela requiert d'aller « chercher » les contenus d'autres pages web en intégrant une image ou une vidéo, associé à du texte en-dessous. En ce qui concerne la carte géographique, elle résulte d'un processus de pointage vers une page en particulier qui répertorie des données géographiques. Pour le moteur de recherche Google, il « pointe » vers la page *Google maps*, pour le moteur de recherche Qwant, il « pointe » vers *Qwant maps*.

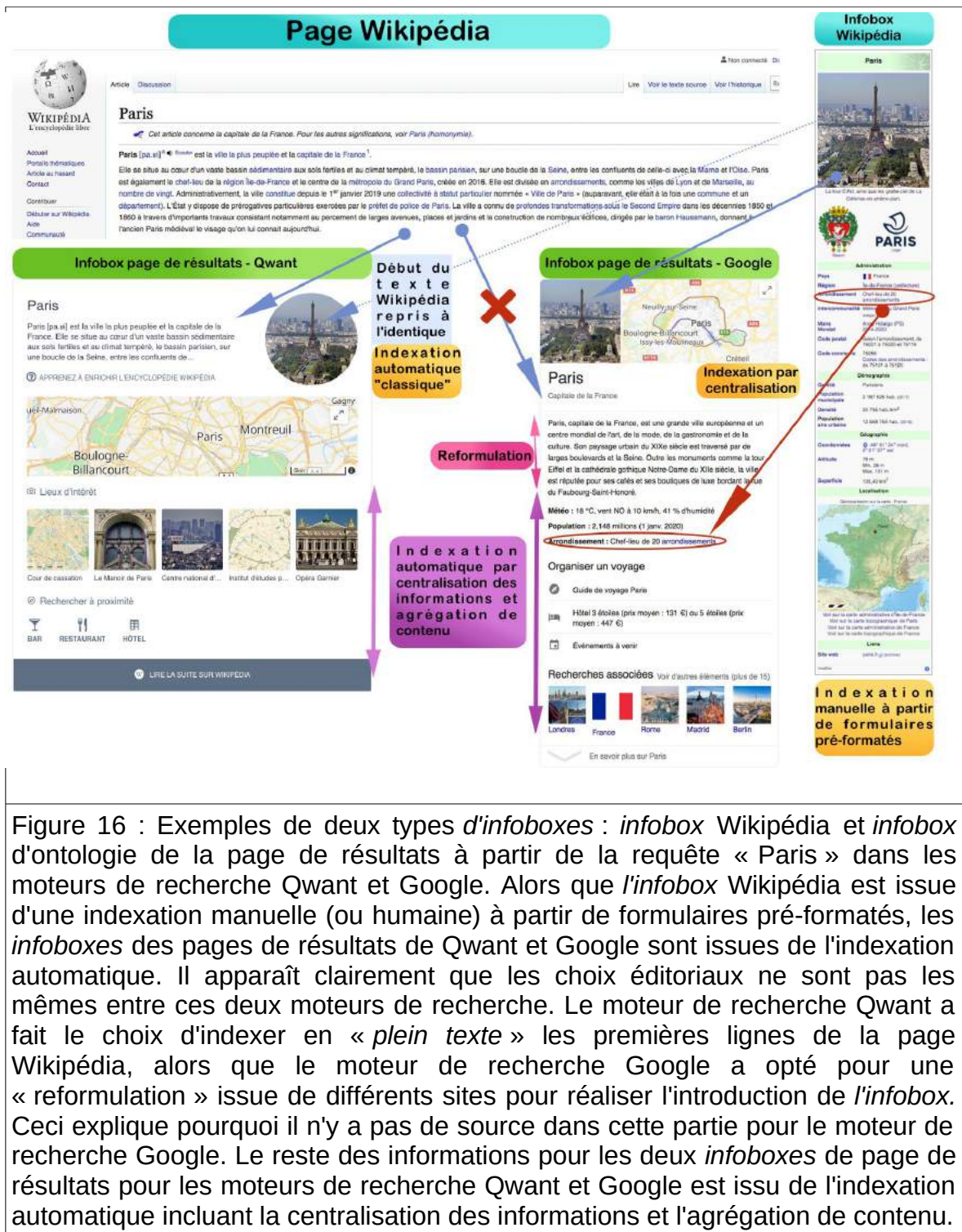


Figure 16 : Exemples de deux types d'infoboxes : infobox Wikipédia et infobox d'ontologie de la page de résultats à partir de la requête « Paris » dans les moteurs de recherche Qwant et Google. Alors que l'infobox Wikipédia est issue d'une indexation manuelle (ou humaine) à partir de formulaires pré-formatés, 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.

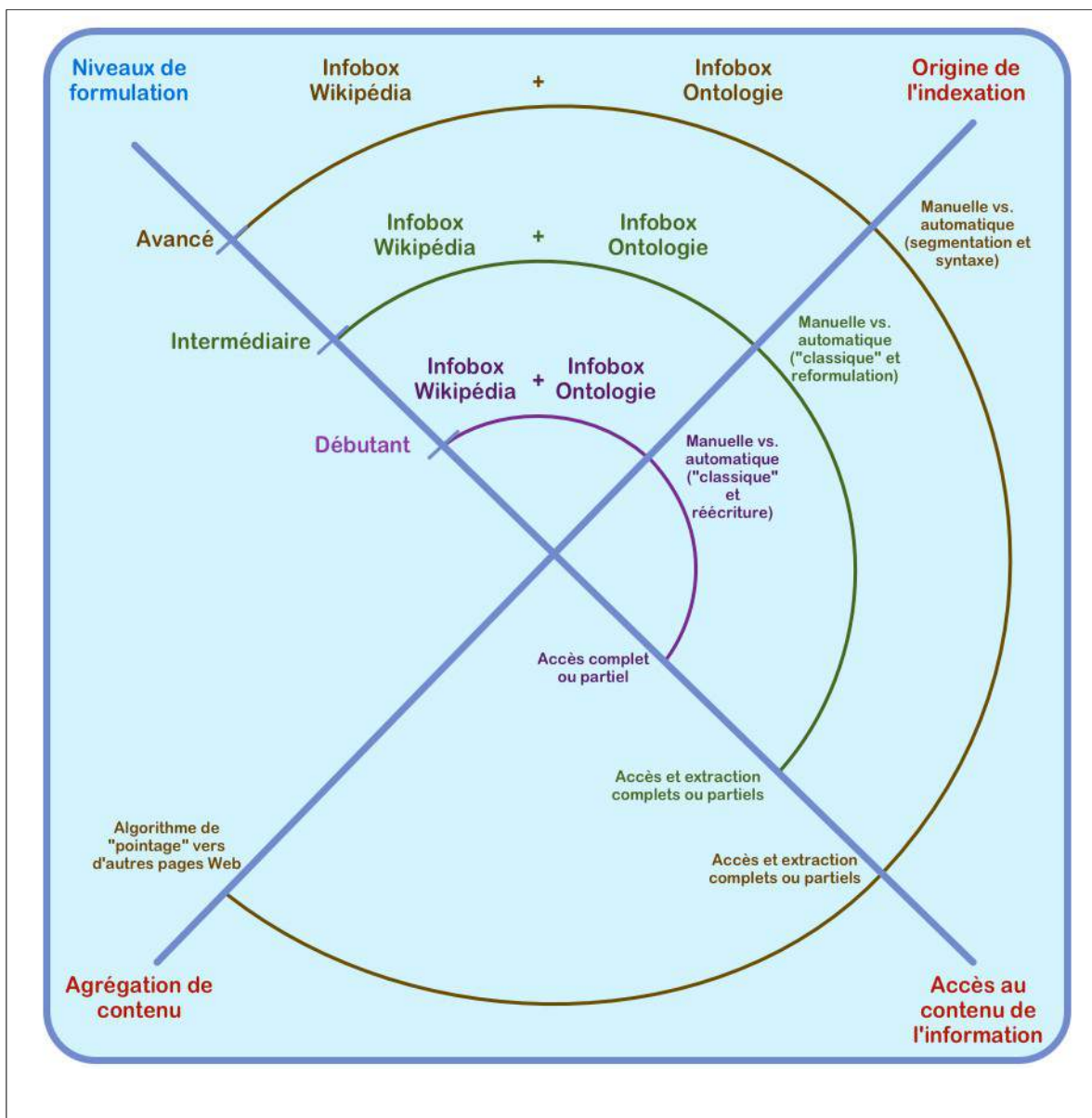


Figure 17 : Progression spiralaire des notions d'indexation manuelle et automatique du niveau débutant à avancé. Les notions sont décomposées en trois caractéristiques : l'origine de l'information, l'accès au contenu de l'information et l'agrégation de contenu. Les notions sont abordées à partir de la 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.

Par conséquent les *infoboxes* Wikipédia et de la page des résultats des moteurs de recherche sont de bons matériaux pour aborder les notions d'**indexation manuelle/humaine et automatique** avec les élèves à différents niveaux du collège au lycée et aussi avec des élèves spécialisés en science numérique et technologie. Le fait que des *infoboxes* qui se ressemblent aient des modes d'indexation différents place les élèves face à un conflit cognitif quant à l'origine de l'information. Alors que l'on pourrait supposer que toutes les *infoboxes*

sont indexées de la même façon, leur étude permet d'identifier des procédés d'indexation différents et donc de dissocier l'indexation manuelle/humaine, de l'indexation automatique au sein même des moteurs de recherche.

Cette entrée par les *infoboxes* Wikipédia en comparaison avec celles des pages de résultats des moteurs de recherche, permettra d'aller plus loin sur l'indexation automatique en général dans un moteur de recherche, mais avec une approche différente de celle des sites sur une page de résultats. Cela permettra d'envisager des enseignements autres que de renseigner une page Wikipédia, qui de l'aveu même des administrateurs de l'encyclopédie en ligne, devient de plus en plus difficile à effectuer car le nombre de pages incomplètes est de plus en plus rare.

Par ailleurs, d'autres notions info-documentaires pourront être abordées grâce à ces encarts informationnels comme la notion de source, la notion d'auteur et celle du fonctionnement d'un moteur de recherche et du *Web*. De plus, comme le font déjà certains professeur.e.s documentalistes, les *infoboxes* permettent un apport à la compréhension des algorithmes et de leur rôle dans les nouvelles formes d'indexation (existantes, naissantes et à venir).

3.4 Conclusion

La majorité des hypothèses que nous avons émises ont été confirmées par nos résultats. Celle qui a été réfutée a montré que la majorité des professeur.e.s documentalistes se sont emparés de la notion de l'indexation automatique dans leurs enseignements, et ce, à partir de leurs propres séances pédagogiques et majoritairement en classe de seconde. Ce sont surtout les professeur.e.s documentalistes expérimentés qui enseignent l'indexation automatique, majoritairement à travers les moteurs de recherche dans le cadre pédagogique des sciences numérique et technologie au lycée et l'information-documentation au collège. La majorité des professeur.e.s documentalistes n'estiment pas nécessaire d'enseigner la notion d'indexation en amont de la notion d'indexation automatique et ceux.celles qui le font souhaitent faire comprendre aux élèves le processus historique intellectuel et fonctionnel de l'indexation. Même si la quasi-totalité des professeur.e.s documentalistes, ayant répondu au questionnaire, a repéré l'existence des encarts informationnels (ou *infoboxes*) sur une page de résultats de moteur de recherche, il n'en demeure pas moins que seulement un-tiers d'entre

eux.elles les intègrent dans leurs enseignements. La majorité d'entre eux.elles répondent qu'ils souhaitent que les élèves retiennent des notions liées à l'évaluation de l'information à partir de l'étude des *infoboxes*. Moins d'un tiers des réponses données évoquent les *infoboxes* comme matériaux pour aborder la question de l'indexation automatique à travers le fonctionnement des moteurs de recherche. La majorité des professeur.e.s documentalistes qui n'enseignent ni l'indexation automatique, ni les *infoboxes*, évoquent le manque de temps majoritairement, mais aussi la complexité de la notion pour les élèves et pour eux.elles mêmes. Par conséquent, la didactisation des processus d'indexation automatique à travers les *infoboxes* peut permettre aux élèves de comprendre l'accès à l'information et son origine sur le *Web*. Cependant, l'étude des *infoboxes* a vocation à se généraliser car c'est une entrée peu utilisée par les professeur.e.s documentalistes à la date de notre étude.

3.5 Discussion sur les biais de l'étude

3.5.1 Point de vue critique

Les résultats que nous avons obtenus pour notre étude confirment la majorité de nos hypothèses, cependant la question de leur représentativité pour l'ensemble de la profession se pose. En effet, nous avons vu que la majorité des répondant.e.s (73%) enseignent dans l'académie de Nantes, ce qui représente 33 personnes. Les données de l'APDEN parues en 2019 évoquent 440 professeur.e.s documentalistes dans cette académie. Par conséquent notre étude est représentative de seulement 7,5% des professeur.e.s documentalistes de cette académie. Cependant nous avons noté un certain équilibre entre le nombre de professeur.e.s documentalistes ayant répondu sur le fait qu'ils enseignaient ou non l'indexation automatique. S'il la majorité des professeur.e.s documentalistes avaient répondu oui à cette question, cela aurait remis trop grandement en question l'hypothèse selon laquelle l'indexation automatique n'est pas enseignée par les professeur.e.s documentalistes. Pour savoir si notre étude est véritablement représentative, il faudrait connaître la raison pour laquelle les professeur.e.s documentalistes qui n'ont pas répondu à ce questionnaire ne l'ont pas fait. Il est difficile d'avoir accès à cette information.

Nous n'avons pas obtenu beaucoup de réponses à la précision demandée après la réponse à la question sur l'enseignement ou non de l'indexation

automatique (7 réponses sur 26 répondant.e.s). En effet, le fait de demander une précision n'appelle pas nécessairement à une réponse. Il aurait été préférable de demander pourquoi il.elle.s enseignaient l'indexation automatique et ainsi nous aurions probablement obtenus plus de réponses. Cela nous aurait permis d'appréhender de façon plus précise les raisons pour lesquelles les professeur.e.s documentalistes enseignent l'indexation automatique. Cependant, avec la question « Que souhaiteriez-vous que les élèves retiennent des encarts informationnels », il y a des réponses qui ont trait au fonctionnement des moteurs de recherche, et donc à l'indexation automatique.

Lorsque nous avons demandé pourquoi les professeur.e.s documentalistes n'enseignaient pas l'indexation en amont de l'indexation automatique, beaucoup des réponses ont été reprises de la question sur les raisons pour lesquelles il.elle.s n'enseignaient pas l'indexation automatique. Par conséquent, le traitement de ces données aurait donné une information redondante. Soit les professeur.e.s documentalistes ont renvoyé ou répété leurs réponses précédentes (8/15 réponses ; Annexe 4), soient il.elle.s ont déclaré manquer de temps (3/15 réponses ; Annexe 4), soit il.elle.s abordent les deux en même temps (2/15 réponses ; Annexe 4), et enfin, soit il.elle.s n'abordent pas du tout ces notions (2/15 réponses ; Annexe 4). Il aurait peut-être été préférable de préciser la question en écrivant « Pourquoi n'enseignez-vous pas l'indexation en amont de l'indexation automatique ? ». Malgré cela, nous ne sommes pas sûrs que les réponses auraient été plus pertinentes.

Pour avoir une meilleure représentativité des cadres pédagogiques dans lesquels l'indexation automatique était enseignée, nous avons dû pondérer par le niveau d'enseignement (collège ou lycée). Il s'avère que nous n'avons pas demandé au préalable à quel niveau du secondaire les professeur.e.s documentalistes enseignaient actuellement. Par conséquent, certains ont répondu en fonction de leur carrière personnelle et ont indiqué tous les niveaux du secondaire, voire même de l'enseignement supérieur. Par conséquent, nous aurions pu obtenir un niveau de précision plus important pour la relation entre le cadre pédagogique et le niveau d'enseignement, et ainsi préciser les différences pour le niveau collège, si nous avons ajouté cette question au préalable.

Nous n'avons pas pu conclure sur la relation entre la méthode pédagogique et le cadre pédagogique car un nombre non négligeable de réponses (15% des réponses ; Annexe 4) à la question « Quelle méthode pédagogique vous paraît la plus pertinente pour enseigner l'indexation automatique, les moteurs de recherche ou les encarts informationnels ? » étaient hors sujet. Par exemple, certaines réponses étaient « *moteurs* » ou bien « *encarts* ». Pour éviter ce type d'écueil, il aurait été préférable de proposer un questionnaire à choix multiple parmi les méthodes pédagogiques. Cependant, il est probable également que les personnes ayant répondu hors-sujet à la question, aient répondu au hasard au choix multiple. De plus, le fait qu'il y avait peu de données dans les cadres pédagogiques en dehors de l'information-documentation, le croisement des données n'était pas pertinent. En effet, le tableau de contingence n'a d'intérêt que lorsque la quantité de données dans chacune des variables testées est relativement équivalente.

3.5.2 Ouvertures et enjeux professionnels

Nous avons vu que, malgré une mise en place récente des sciences numérique et technologie en classe de seconde à la rentrée 2019, certain.e.s professeur.e.s documentalistes se sont emparé.e.s de ce cadre pédagogique pour y enseigner l'indexation automatique à travers le fonctionnement des moteurs de recherche. Nous pouvons supposer que si l'étude était de nouveau réalisée dans quelques années, nous pourrions obtenir un vivier plus important de séances pédagogiques en SNT à analyser. Pour le lycée, les SNT représentent un cadre pédagogique intéressant pour enseigner l'indexation automatique pour les professeur.e.s documentalistes.

Nous pensons qu'il est important de prendre en compte le fait que certain.e.s professeur.e.s documentalistes (21%), n'enseignent pas l'indexation automatique car il.elle.s ne se sentent pas assez compétent.e.s. Une des réponses obtenue stipule « *Zéro formation sur le sujet (vieille doc de plus de 40 ans) : je vois bien ce que c'est maintenant qu'on me le dit mais avant, en paraphasant M. Jourdain "j'en vois sans le savoir" »*. Il est intéressant de noter qu'elle se considère comme « vieille », alors qu'elle est dans la moyenne d'âge des professeur.e.s documentalistes en France (APDEN, 2019). Mais c'est la question de la formation que nous souhaiterions noter ici. Il est probable qu'une formation professionnelle sur l'évolution des moteurs de recherche et de

l'indexation automatique, notamment au travers de l'étude des *infoboxes*, permettrait aux professeur.e.s documentalistes de s'emparer de ces questions dans leurs enseignements.

3.6 Conclusion générale

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, et notre étude le montre, l'indexation automatique est enseignée à travers l'étude de la page de résultats pour comprendre le fonctionnement des moteurs de recherche. L'autre approche utilisée par les professeur.e.s documentalistes est la comparaison entre indexation humaine et indexation automatique à travers le fonctionnement du portail documentaire et les 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 celles liées à la recherche informationnelle. Il permet une autre entrée que la page de résultats plus traditionnelle pour comprendre le fonctionnement des moteurs de recherche. De même, il offre une alternative à l'étude du portail documentaire pour enseigner la comparaison entre l'indexation humaine et l'indexation automatique. Nous montrons que les professeur.e.s documentalistes, et *a fortiori* les élèves, ont remarqué l'existence des différents types d'encarts informationnels. Les *infoboxes* sont d'autant plus remarquables qu'ils sont souvent situés en haut de la page de résultat.s Que pensent les élèves de ces *infoboxes* ? Comment les utilisent-ils ? Quels sont leurs représentations ? sont autant de questions à leur poser, soit dans une autre étude, soit en classe afin de les mettre face à une situation-problème.

Bien que peu de professeur.e.s documentalistes enseignent les *infoboxes* avec leurs élèves, beaucoup d'entre eux.elles souhaiteraient le faire, mais souvent le manque de temps les en empêche. Pour d'autres, le manque de compétences personnelles est un facteur limitant pour l'enseignement de l'indexation automatique. C'est pourquoi nous proposons ici une révision de la didactisation de la notion d'indexation automatique et une progression possible d'enseignement

des *infoboxes* dans le but de différencier l'indexation humaine de l'indexation automatique, et ce, pour tous les niveaux du secondaire. Dans la séance que nous proposons, nous comparons une *infobox* Wikipédia (de la page Wikipédia), à une *infobox* d'ontologie (de la page de résultats). L'objectif est ici de repérer l'origine de l'information de ces *infoboxes* et de comparer l'indexation humaine, de l'indexation automatique. De l'aveu même de l'équipe Wikipédia, l'exercice qui consistait à créer et/ou compléter des pages Wikipédia avec les élèves est aujourd'hui limité car la majorité des pages, du niveau des élèves, sont complètes. Par conséquent, l'étude des *infoboxes* laisse la possibilité de renouveler l'utilisation de cette encyclopédie en ligne dans les enseignements. Il nous semble également nécessaire qu'une formation continue soit disponible pour les professeur.e.s documentalistes qui le souhaitent, car même certain.e.s professeur.e.s documentalistes qui enseignent l'indexation automatique reconnaissent la difficulté de sa compréhension et de son enseignement.

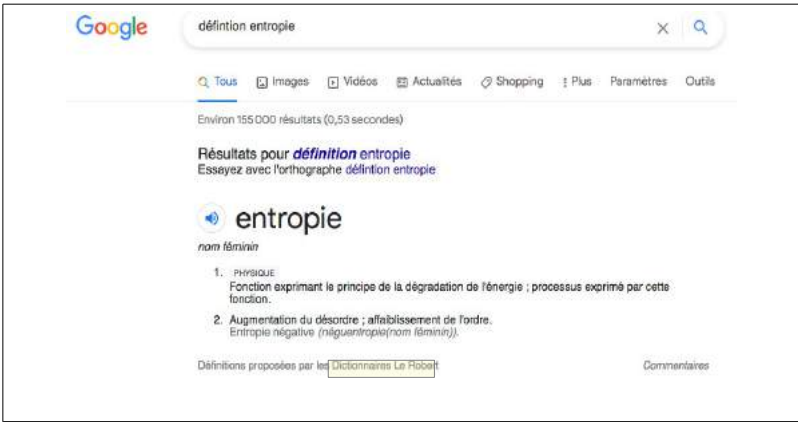

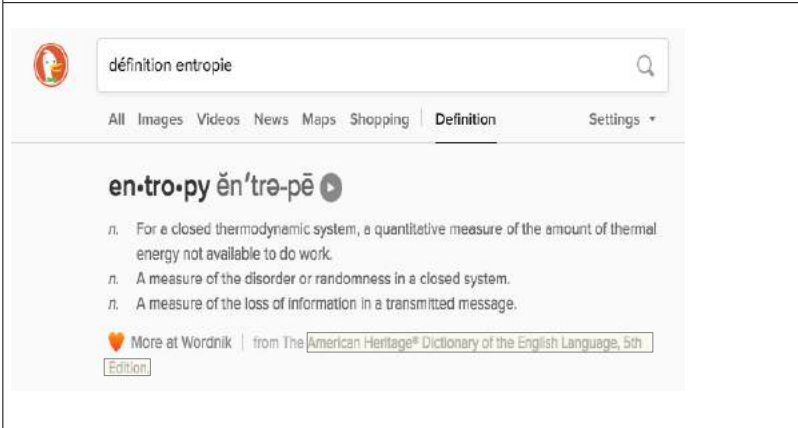
La didactisation de l'indexation automatique à travers les *infoboxes* s'intègre parfaitement dans les programmes du collège et du lycée et autorise une potentialité nouvelle d'enseignement pour les professeur.e.s documentalistes. Nous avons vu dans notre étude que l'indexation automatique est principalement enseignée dans le cadre pédagogique de l'information-documentation au collège et dans le cadre des SNT de seconde générale au lycée. Au lycée général, bien que les professeur.e.s documentalistes ne soient pas encore majoritairement impliqués dans l'enseignement des SNT, sa récente mise en place (rentrée 2019), présage d'une évolution possible vers une implication plus importante. D'autant plus, si les professeur.e.s documentalistes sont en mesure de proposer des séances « clés-en-main » comme démontré dans l'étude de l'APDEN en 2019.

Par conséquent, les *infoboxes*, ou encarts informationnels, peuvent être vus comme un nouveau matériau dont les professeur.e.s documentalistes, mais aussi les autres professeurs, peuvent s'emparer dans leurs enseignements pour répondre aux programmes institutionnels tout en enseignant l'information-documentation. Ils offrent, par ailleurs, une porte d'entrée privilégiée pour enseigner l'indexation automatique, notamment, par une approche de situation-problème avec les élèves. En effet, leur présence croissante sur les pages de résultats des moteurs de recherche permet de questionner les élèves sur l'origine de l'information et de leur fonctionnement.

Les moteurs de recherche aujourd'hui utilisent les dernières fonctionnalités algorithmiques en termes d'intelligence artificielle pour apporter, non plus une liste de résultats, mais une réponse à une requête. Les procédés d'indexation automatique n'échappent pas à cette règle. Les *infoboxes* représentent donc un condensé d'informations permettant de comprendre les dernières fonctionnalités du Web.

Addendum

Au moment de la publication de ce présent mémoire en Avril 2021, les moteurs de recherche Google, Bing et désormais Duckduck go, ont ajouté la source des définitions dans les encarts définitionnels. Ainsi, la précision de l'origine de l'information dans les encarts définitionnels (Tableau 1, page 38) est désormais une pratique adoptée par ces deux moteurs de recherche. Cependant, le moteur de recherche Bing, présente d'abord un encart Wikipédia, puis des exemples de questions avant de proposer l'encart définitionnel. Quant aux moteurs de recherche Ecosia et Qwant. ils ne disposent toujours pas de ce service.

	<p>Encart définitionnel du mot « entropie » pour le moteur de recherche Google au 2/03/2021. La source est encadrée en jaune.</p>
	<p>Encart définitionnel du mot « entropie » pour le moteur de recherche Bing au 2/03/2021. La source est encadrée en jaune.</p>
	<p>Encart définitionnel du mot « entropie » pour le moteur de recherche Duckduck go au 2/03/2021. La source est encadrée en jaune.</p>

Bibliographie

- Académie de Poitiers. (2019). Comprendre le fonctionnement des moteurs de recherche. *Doc'Poitiers - Le site des professeurs documentalistes*. Récupéré le 10 avril 2020 depuis le site de l'académie de Poitiers : <http://ww2.ac-poitiers.fr/doc/spip.php?article855>
- Accart, J.-P. (2015). *Le métier de documentaliste*. Éditions du Cercle de la Librairie.
- Aide:Infobox. (2019). In *Wikipedia*. <https://fr.wikipedia.org/wiki/Aide:Infobox>
- Amar, Muriel. (2000). *Les fondements théoriques de l'indexation : Une approche linguistique*. Paris: ADBS Éditions.
- Amar, Muriel, & Menon, Bruno. (2011). Bienvenue dans la « gigantesque base de données » . *Documentaliste-Sciences de l'Information*, 48 (4), 22-23. <https://doi.org/10.3917/docsi.484.0022>
- Association des bibliothécaires de France. (2013). *Le métier de bibliothécaire* (13è). Éditions du Cercle de la Librairie.
- Association des professeurs documentalistes de l'éducation nationale (APDEN). (2016a). *Vers un curriculum en information-documentation—Chapitre 8—Les programmes de l'information-documentation en collège*. Récupéré le 5 mai 2020 sur le site de l'APDEN : <http://www.apden.org/Vers-un-curriculum-en-information.html>
- Association des professeurs documentalistes de l'éducation nationale (APDEN). (2016b). *Vers un curriculum en information-documentation—Chapitre 9 : Les bases communes de l'information-documentation au lycée*. Récupéré le 5 mai 2020 sur le site de l'APDEN : <http://www.apden.org/Vers-un-curriculum-en-information.html>
- Association des professeurs documentalistes de l'éducation nationale (APDEN). (2017). *Présentation du projet*. Wikinotions. Récupéré le 4 avril 2020 sur le site Wikinotions de l'APDEN : http://Wikinotions.apden.org/index.php?title=Pr%C3%A9sentation_du_projet
- APDEN. (2019). *Données statistiques. Edition 2019*. PROFESSION : PROF DOC. Récupéré le 24 mai 2020 sur le site de l'APDEN : <http://professionprofdoc.educapass.fr/788-2-2/donnees-statistiques/>
- Astolfi, Jean-Pierre. (1997). *L'erreur, un outil pour enseigner*. Paris: ESF Sciences humaines.
- Astolfi, J., Darot, É., Ginsburger-Vogel, Y. & Toussaint, J. (2008). Chapitre 6. Didactique. Dans J. Astolfi, É. Darot, Y. Ginsburger-Vogel & J. Toussaint (dir.), *Mots-clés de la didactique des sciences: Repère, définitions, bibliographies* (p. 67-76). Louvain-la-Neuve, Belgique: De Boeck Supérieur.
- Astolfi, J.-P., Darot, É., Ginsburger-Vogel, Y., & Toussaint, J. (2008b). Chapitre 13. Pratique sociale de référence. Dans J. Astolfi, É. Darot, Y. Ginsburger-Vogel & J. Toussaint (dir.), *Mots-clés de la didactique des sciences: Repère, définitions, bibliographies* (p. 131-136). De Boeck Supérieur.

- Balmisse, G. (2015). *Digital workplace et gestion des connaissances*. ENI.
- Barati, M., Bai, Q., & Liu, Q. (2019). Automated Class Correction and Enrichment in the Semantic Web. *Journal of Web Semantics*, 59, 100533. <https://doi.org/10.1016/j.websem.2019.100533>
- Bellier, N., & Estéoule, B. (1988). Histoires d'OPACs. *Bulletin des Bibliothèques de France (BBF)*, 3, 224-226.
- Berners-Lee, T., Hendler, James, & Lassila, Ora. (2001). The Semantic Web. *Scientific American*, Feature article, 4. <https://www.scientificamerican.com/article/the-semantic-web/>
- Bissan, A. (2014). *Reformulation sémantique des requêtes pour la recherche d'information ad hoc sur le Web* (Thèse de doctorat, École Nationale Supérieure des Mines, Saint-Étienne, France). Récupéré de HAL archives ouvertes, système de dépôt de documents universitaires en ligne : <https://tel.archives-ouvertes.fr/tel-01126932/document>
- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia—A crystallization point for the Web of Data. *Journal of Web Semantics*, 7(3), 154-165. <https://doi.org/10.1016/j.websem.2009.07.002>
- Boulogne, A. (2004). *Vocabulaire de la documentation*. Paris: ADBS Éditions.
- Broudoux, E. (2012). Vers l'objet documentaire (re)contextualisé. Objets documentaires numériques ?. *9^e 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.
- Cardon, D. (2015). *À quoi rêvent les algorithmes. Nos vies à l'heure des big data*. Paris: Éditions du Seuil.
- Cardon, D. (2019). *Culture numérique*. Paris: Presses de Sciences Po.
- Cayrel, J. (2020). *Moteurs de recherche et bulle de filtre*. Récupéré le 11 avril 2020 sur le site des personnels de l'éducation nationale, Éduscol : https://disciplines.ac-toulouse.fr/documentation/sites/documentation/files/fichiers/2-_moteurs_de_recherche_et_bulles_de_filtres.pdf.
- Chambaud, P., Serret, M., Le Seven, I., & Delecour, B. (2020, 19 mars). *Les professeurs documentalistes et l'enseignement de SNT*. Récupéré le 11 avril 2020 sur le site des professionnels de l'éducation nationale, Éduscol : <https://eduscol.education.fr/cdi/anim/actions-mutuali/le-professeur-documentaliste-dans-le-continuum-de-formation-des-eleves-du-cycle-3-a-lentree-dans-le-superieur-2019-2020/documents-realises-par-les-referentes-traam/les-professeurs-documentalistes-et-lenseignement-de-snt>
- Chaumier, Jacques. (2006). *Document et numérisation : Enjeux techniques, économiques, culturels et sociaux*. Paris: ADBS Éditions.
- Chevallard, Yves. (1985). *La transposition didactique : Du savoir savant au savoir enseigné*. Grenoble: La Pensée Sauvage.

- Chevallard, Yves, & Joshua, Marie-Alberte. (1991). *La transposition didactique : Du savoir savant au savoir enseigné*. Grenoble: La Pensée Sauvage.
- Dehaene, Stanislas. (2018). *Apprendre ! Les talents du cerveau, le défi des machines*. Paris: Odile Jacob.
- Delestre, Nicolas, & Malandain, Nicolas. (2017). *Du web des documents au web sémantique*. Bois-Guillaume: KLOG Éditions.
- Dublin Core. (2020). Dans *Wikipedia*. https://fr.wikipedia.org/wiki/Dublin_Core
- Duffez, O. (2020). *Parts de marché moteurs de recherche (mai 2020) France, USA, monde*. Récupéré le 31 mai 2020 sur le site WebRankInfo : <https://www.webrankinfo.com/dossiers/etudes/parts-marche-moteurs>
- Duplessis, P. (2014). *Les méthodes pédagogiques en information-documentation*. Les trois couronnes. <http://lestroiscouronnes.esmeree.fr/outils/les-methodes-pedagogiques-en-information-documentation>
- Duplessis, P. (2016). *L'information-documentation en dix tableaux : Ce qui est réellement enseigné par les professeurs documentalistes*. Récupéré le 10 avril 2020 sur le site Les Trois Couronnes - Didactique de l'Information Documentation : <http://lestroiscouronnes.esmeree.fr/didactique-information/l-information-documentation-en-dix-tableaux>
- Duplessis, P. (2017). *Enseigner l'information-documentation : Un triple défi didactique pour le professeur documentaliste*. Récupéré le 15 avril 2020 sur le site Les Trois Couronnes - Didactique de l'Information Documentation : <http://lestroiscouronnes.esmeree.fr/didactique-information/enseigner-l-information-documentation-un-triple-defi-didactique-pour-le-professeur-documentaliste>
- Duplessis, P. (2019a). *Emprise de l'information-documentation dans les programmes du cycle 3*. Récupéré le 10 avril 2020 sur le site Les Trois Couronnes - Didactique de l'Information Documentation : <http://lestroiscouronnes.esmeree.fr/didactique-information/emprise-de-l-information-documentation-dans-les-programmes-du-cycle-3>
- Duplessis, P. (2019b). *Emprise de l'information-documentation dans les programmes du cycle 4*. Récupéré le 10 avril 2020 sur le site Les Trois Couronnes - Didactique de l'Information Documentation : <http://lestroiscouronnes.esmeree.fr/didactique-information/emprise-de-l-information-documentation-dans-les-programmes-du-cycle-4>
- El Hachani, M. (1997). *L'indexation automatique* (Mémoire de DEA, ENSSIB, Lyon). Récupéré de la Bibliothèque de l'ENSSIB : <http://enssibal.enssib.fr/bibliotheque/documents/dea/elhachani.pdf>
- Ertzscheid, O. (2017). *L'appétit des géants : Pouvoir des algorithmes, ambitions des plateformes*. Caen: C&F éditions.
- Fleury, B. (Productrice) et Billon, E. (Réalisateur). (2010). *Entretien avec Jean-Pierre Astolfi : Retrouver la saveur des savoirs à l'école*. [DVD]. Dijon: Éducagri.
- Frisch, M. (2008, 28-30 mars). Didactisation de l'information-documentation : Un enjeu de professionnalisation. 8^e congrès des enseignants documentalistes de l'Éducation nationale, ENS Sciences, Lyon, 22-29. <https://hal.archives-ouvertes.fr/hal-00687743/document>

- Frisch, M. (2016). Des modèles et des concepts en question pour une didactique de l'Information-Documentation et proposition d'une matrice curriculaire dynamique. *Tréma*, 45, 43-57. <https://doi.org/10.4000/trema.3492>
- Gardiès, C. (2008). *L'éducation à l'information : Guide d'accompagnement pour les professeurs documentalistes*. Educagri Editions.
- Guillet, M. (2013, 19 juin). Séance 7 : Culture Net'—Classer l'information sur le web (Initiation) (techniques documentaires 6ème). *Doc à bord*. <https://docabord.wordpress.com/2013/06/19/seance-7-culture-net-classer-linformation-sur-le-web-initiation-techniques-documentaires-6eme/>
- Ibáñez, L. D. (2015). *Towards a Read/Write Web of Linked Data* (Thèse de doctorat, Université Nantes-Angers-Le Mans). Récupéré sur le site des archives du service commun de la documentation de l'Université de Nantes : <http://archive.bu.univ-nantes.fr/pollux/show.action?id=9089939a-874b-44e1-a049-86a4c5c5d0e6>
- 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.
- Krichen, I., Kopliku, A., Pinel-Sauvagnat, K., & Boughanem, M. (2012). Une approche de recherche d'attributs pertinents pour l'agrégation d'information. *Document numérique*, 15(1), 9-32. <https://doi.org/10.3166/dn.15.1.9-32>
- Le Deuff, O. (2014). *La documentation dans le numérique*. Lyon: Les Presses de l'Enssib.
- Lemelin, A. (2004). Chapitre 4-1 : L'analyse des tableaux de contingence. In *Méthodes quantitatives des sciences sociales appliquées aux études urbaines et régionales* (p. 1-39). Canada: INRS-UCS. http://www.lemelin-metho.ucs.inrs.ca/wp-content/uploads/4_1.pdf
- Martinand, J.-L. (2014). Point de vue V - Didactique des sciences et techniques, didactique du curriculum. *Éducation et didactique*, 8(8-1), 65-76. doi.org/10.4000/educationdidactique.1886
- Ministère de l'Éducation nationale et de la jeunesse (2019). *Arrêté du 17-1-2019 -J.O. du 20-1-2019 relatif au programme d'enseignement de sciences numériques et technologie de la classe de seconde générale et technologique, Pub. L. No. MENE1901641A*. Récupéré le 21 avril 2020 sur le site du ministère de l'éducation nationale : <https://www.education.gouv.fr/bo/19/Special1/MENE1901641A.htm>
- Ministère de l'Éducation nationale et de la jeunesse (2019, 19 juillet). *Arrêté du BO spécial n°8 du 19-7-2019 relatif au programme des enseignements de spécialité de la classe de terminale conduisant au baccalauréat technologique série sciences et technologies du management et de la gestion (STMG), Pub. L. No. MENE1921262A*. Récupéré le 21 avril 2020 sur le site du ministère de l'éducation nationale : <https://www.education.gouv.fr/bo/19/Special8/MENE1921262A.htm>
- Ministère de l'Éducation nationale et de la jeunesse (2019, 30 août). *Décret n° 2019-919 du 30 août 2019 relatif au développement des compétences numériques dans l'enseignement scolaire, dans l'enseignement supérieur et*

- par la formation continue, et au cadre de référence des compétences numériques, Pub. L. No. MENE1915146D, 2019-919. Récupéré le 21 avril 2020 sur le site du ministère de l'éducation nationale : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000039005162&categorieLien=id>
- Mesguich, V., & Thomas, A. (2010). *Net recherche 2010. Le guide pratique pour mieux trouver l'information utile et surveiller le web* (4^e éd.). Paris: ADBS éditions.
- Minder, M. (2007). *Didactique fonctionnelle : Objectifs, stratégies, évaluation, le cognitivisme opérant* (9^e éd.). Louvain-la-Neuve, Belgique: De Boeck université.
- Montoya, G. (2016). *Answering SPARQL Queries using Views* (Thèse de doctorat, Université de Bretagne Loire). Récupéré sur le site des archives du service commun de la documentation de l'Université de Nantes <http://archive.bu.univ-nantes.fr/pollux/show.action?id=9fe6e29f-d2fe-4249-b44d-eb4dbe472a99>
- Mulot, H. (2013, 21 juin). *Un réseau social pour un défi lecture : Présence numérique assumée*. Récupéré le 13 avril 2020 sur le blog L'odyssée d'Ln : je tisse m@ toile : <http://odysseedln.verblog.com/reseau-social-defi-lecture>
- Murdock, V., & Lalmas, M. (2008). *Workshop on aggregated search*. ACM SIGIR Forum, 2, 80-83. <http://dx.doi.org/10.1145/1480506.1480520>
- Ontologie (Informatique). (2019). Dans *Wikipedia*. [https://fr.wikipedia.org/wiki/Ontologie_\(informatique\)](https://fr.wikipedia.org/wiki/Ontologie_(informatique))
- Planté, M. (2006). *Extraction automatique de connaissances pour la décision multicritère* (Id: tel-00353770) [Ecole Nationale Supérieure des Mines]. HAL. <https://tel.archives-ouvertes.fr/tel-00353770>
- Queval, S. (2010). Pour une progression spiralaire de l'apprentissage du philosophe. *Diotime - Revue internationale de didactique de la philosophie*, 46. <http://www.educ-revues.fr/DIOTIME/AffichageDocument.aspx?iddoc=39409&pos=15>
- 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>
- Reuter, Y., Cohen-Azria, C., Daunay, B., Delcambre, I., & Lahanier-Reuter, D. (2013). Didactiques. Dans *Dictionnaire des concepts fondamentaux des didactiques: 3e éd.* (p. 65-69). Louvain-la-Neuve, Belgique: De Boeck Supérieur. doi:10.3917/dbu.reute.2013.01.
- Reuter, Y., Cohen-Azria, C., Daunay, B., Delcambre, I. & Lahanier-Reuter, D. (2013). Didactiques. Dans : Y. Reuter, C. Cohen-Azria, B. Daunay, I. Delcambre & D. Lahanier-Reuter (dir.), *Dictionnaire des concepts fondamentaux des didactiques* (p. 65-69). Louvain-la-Neuve, Belgique: De Boeck Supérieur.
- Rey, A. (2016a). Sémantique. Dans *Dictionnaire historique de la langue française* (5^e éd., Vol. 3). Le Robert.
- Rey, A. (2016b). Web. Dans *Dictionnaire historique de la langue française* (5^e éd., Vol. 3). Le Robert.

- Rey, A. (2019). Ontologie. Dans *Dictionnaire historique de la langue française* (5^e éd., Vol. 2). Le Robert.
- Reynaud, F. (2014). Espaces et documents du CDI. Récupéré le 10 avril 2020 sur le site de l'APDEN, Wikinotions : http://Wikinotions.apden.org/index.php?title=Espaces_et_documents_du_CDI
- Reynaud, F. (2017). Indexation automatisée. Récupéré le 01 juin 2020 sur le site de l'APDEN, Wikinotions : http://Wikinotions.apden.org/index.php?title=Indexation_automatis%C3%A9e
- 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>
- Schneuwly, B. (2008). 3. De l'utilité de la « transposition didactique ». In *Didactique du français : Fondements d'une discipline*, 8^e éd. (p. 47-59). Louvain-la-Neuve, Belgique: De Boeck Supérieur. <https://www-cairn-info.budistant.univ-nantes.fr/didactique-du-francais--9782804158699-page-47.htm>
- Simmonot, B. (2012). *L'accès à l'information en ligne*. Cachan: Hermès - Lavoisier.
- Simmonot, B. (2015). Moteurs de recherche. Dans *Encyclopaedia Universalis*. <http://www.universalis-edu.com/encyclopedie/moteurs-de-recherche/>
- 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/>
- Soulier, E. (2005). Le système de gestion des connaissances pour soutenir le storytelling dans l'entreprise. *Revue française de gestion*, no 159(6), 247-264.
- Suominen, O. (2019). *Annif, l'indexation automatique à la Bibliothèque nationale de Finlande*. Arabesque, Agence bibliographique de l'enseignement supérieur. <https://publications-prairial.fr/arabesques/index.php?id=605>
- Timimi, I., & Kovacs, S. (2006). *Indice, index, indexation : Actes du colloque organisé par les laboratoires CERSATES et GERICO de l'université Lille-3 des 3 et 4 novembre 2005*. Paris: ADBS Éditions.
- Web sémantique. (2013). Dans C. Jackson (dir.), *Le dictionnaire de l'Enssib*. Bulletin des Bibliothèques de France, Les Presses de l'Enssib. <https://www.enssib.fr/le-dictionnaire/web-semantique>
- Wikipédia/Wikipédia dans la classe—Wikilivres. (2018). Wikibooks. Récupéré le 18 avril 2020 sur le site Wikipédia : [https://fr.wikibooks.org/wiki/Wikip%C3%A9dia/Wikip%C3%A9dia_dans_la_classe#Exemples_et_suggestions_de_projets_p%C3%A9dagogiques_\(scolaires_ou_%C3%A9tudiants\)](https://fr.wikibooks.org/wiki/Wikip%C3%A9dia/Wikip%C3%A9dia_dans_la_classe#Exemples_et_suggestions_de_projets_p%C3%A9dagogiques_(scolaires_ou_%C3%A9tudiants))
- Zacklad, M. (2007). Une théorisation communicationnelle et documentaire des TIC. *Humanités numériques 2 : socio-informatique et démocratie cognitive*, 20-35. Récupéré sur le site des archives-ouvertes HAL : https://archivesic.ccsd.cnrs.fr/sic_00180186

ANNEXES

Annexe 1 : Exemples d'infoboxes de définition

Exemples d'infoboxes de définitions issus des moteurs de recherche Google et Bing.

Les sources issues des définitions surlignées en bleu sont exposées sous les captures d'écran des encarts de définitions.

On pourra remarquer que les sources semblent plus souvent issues de sites crédibles pour le moteur de recherche Bing, que pour le moteur de recherche Google.

MOTEUR DE RECHERCHE	
GOOGLE	BING
<p>information</p> <p><i>nom féminin</i></p> <ol style="list-style-type: none"> DROIT Enquête pour établir la preuve d'une infraction, pour en découvrir les auteurs. <small>Ouvrir une information contre X.</small> Renseignement (sur qqn, sur qqch.). Des informations confidentielles. Action de s'informer. Une réunion d'information. Renseignement ou événement qu'on porte à la connaissance d'une personne, d'un public. Une information exclusive (anglicisme scoop). - L'information action d'informer le public, l'opinion (par les médias). SCIENCES Ce qui peut être transmis par un signal ou une combinaison de signaux (message) selon un code commun et par un canal ; ce qui est transmis (objet de connaissance, de mémoire). <small>Traitement informatique de l'information.</small> <p>Traduire "information" en <input type="text" value="Sélectionnez une langue"/></p> <p>Moins</p>	<p>information</p> <p>[information] </p> <p>NOM FEMININ</p> <ol style="list-style-type: none"> Ensemble de renseignements sur quelqu'un, sur quelque chose. "Recueillir de l'information sur un sujet." - [more] Évènement porté à la connaissance d'un public. "Une information de dernière heure (et non un "scoop)." <i>au pluriel</i> Actualités radiodiffusées ou télévisées. "Écouter le bulletin d'Informations à la radio." Ensemble des activités qui ont pour objet la collecte, le traitement et la diffusion des nouvelles auprès du public. "Le traitement de l'information." - [more] <p>More definitions, origin and translations</p>
<p>cest-isic.fandom.com · wiki · Histoire_de_l'ISIC ▾</p> <p>Histoire de l'ISIC Wiki C'est ISIC Fandom</p> <p>Ronald Aylmer Fisher. © Wikipedia affirmant que moins une observation est probable, plus est porteuse d'information. De plus, un élément ou une ...</p> <p>FANDOM est une marque média dédiée aux fans de divertissement à l'échelle internationale, à fournissons une plate-forme pour explorer, contribuer à et célébrer l'univers de la pop cultu constitue la source des fans de divertissement par excellence. Que vous cherchiez des infor approfondies sur vos fandoms favoris ou les dernières actualités, FANDOM satisfait les curios chacun en matière de pop culture à travers le savoir de fans experts et un contenu multiplat original, attractif et géré avec soin.</p>	<p>INFORMATION : Définition de INFORMATION Translate this page</p> <p>https://cnrtl.fr/definition/information -</p> <p>Ensemble des activités qui ont pour objet la collecte, le traitement et la diffusion des nouvelles auprès du public. Une information honnête, libre, manipulée, objective; filtrer l'information; les grands moyens d'information; le marché mondial de l'information; le Ministère de l'Information. [Simon] approuvait ces projets et voyait un avenir secondaire s'ouvrir à lui dans la ...</p> <p>Définition pour INFORMATION subst. fém. — Le ... Translate this page</p> <p>https://www.le-tresor-de-la-langue.fr/definition/information -</p> <p>Ensemble des activités qui ont pour objet la collecte, le traitement et la diffusion des nouvelles auprès du public. Une information honnête, libre, manipulée, objective; filtrer l'information; les grands moyens d'information; le marché mondial de l'information; le Ministère de l'Information. [Simon] approuvait ces projets et voyait un avenir secondaire s'ouvrir à lui dans la ...</p>

communication

nom féminin

1. Le fait de communiquer, d'établir une relation avec (qqn, qqch).
Être en communication avec un correspondant.
• SCIENCES
Relation dynamique qui intervient dans un fonctionnement ; échange de signes, de message entre un émetteur et un récepteur.
Etude des signes et de la communication.
Synonymes : information
- INFORMATIQUE
Échange d'informations entre organes divers (matériels, logiciels).
Communication de données.
2. Action de communiquer qqch. à qqn ; résultat de cette action.
Demander communication d'un dossier.
Synonymes : information
3. Moyen technique par lequel des personnes communiquent.
Communication par Internet.
Synonymes : transmission
4. Ensemble des techniques médiatiques d'information et de publicité.
Service de communication d'une entreprise (relations publiques).
5. Ce qui permet de communiquer dans l'espace ; liaison d'un lieu à un autre.
Porte de communication.

communication

[communication]

NOM FEMININ

1. Action de communiquer à l'aide de paroles, de gestes ou de signes.
"Une communication téléphonique." · "La qualité de la communication."
2. Information, message.
"J'ai une communication à vous transmettre."
3. Conférence.
"Ce scientifique fera une communication sur un nouveau traitement."
4. Passage d'un lieu à un autre.
"Une porte, un couloir de communication."

dictionnaire.lerobert.com · définition · communication ▾

communication - Définitions, synonymes, conjugaison ...

28 févr. 2020 - Sciences Relation dynamique qui intervient dans un fonctionnement ; échange de signes, de messages entre un émetteur et un récepteur.

ivoiremobiles.blogspot.com · 2010/09 · la-communication-c-est-le-fait-de

ENTREPRISE IVOIREMOBILES - IvoireMobiles

24 sept. 2010 - ... relation dynamique qui intervient dans un fonctionnement échange de signes de messages entre un emetteur et un récepteur c est aussi l ...

docplayer.fr · 20514594-Universite-de-nantes-faculte-de-medecine-e...

Université de Nantes Faculté de Médecine Ecole d ...

Relation dynamique qui intervient dans un fonctionnement ; échange de signes, de messages entre un émetteur et un récepteur. a.2. Action de communiquer ...

Technologies de l'information Flashcards | Quizlet Translate this page

<https://quizlet.com/189362178/technologies-de-linformation-flash-cards> ▾

Action de communiquer à l'aide de paroles, de gestes ou de signes. 2. Information, message. la fuite d'information. La déperdition ou la divulgation inopportune de données. informatique. Relatif au traitement automatisé de l'information, l'intelligence artificielle ...

7 Qui suis-je : mes qualités et mes défis

meta-phare.centrefora.on.ca/mtph/themes/connaissance_de_soi/qui_suis... · PDF file

Action de communiquer à l'aide de paroles, de gestes ou de signes 5 Écoute Action d'écouter, d'être attentif à ce qui se dit 9 Optimisme Disposition à voir les bons côtés de la réalité 10 Bonnes manières Façons habituelles d'agir en société 11 Loyauté Honnêteté, droiture 12 Coopération Collaboration 13 Diplomatie Habileté, tact 14 Ponctualité Exactitude, être à l ...

COMMUNICATION : Définition de COMMUNICATION Translate this page

<https://www.cnrtl.fr/definition/communication> ▾

1. - Action de communiquer quelque chose à quelqu'un; le résultat de cette action (cf. communiquer 1).
Avoir, demander, donner, ordonner, prendre, recevoir, refuser (la) communication d'une affaire, d'un avis, d'un document, d'un renseignement; salle de communication. Si vous le [le commencement, formant préambule] trouvez bon, mais seulement dans ce cas, imprimez-le ...

entropie

nom féminin

1. PHYSIQUE
Fonction exprimant le principe de la dégradation de l'énergie ; processus exprimé par cette fonction.
2. Augmentation du désordre ; affaiblissement de l'ordre.
Entropie négative (*neg-entropie*(nom féminin)).

Traductions et autres définitions

entropie

[entropie]

NOM FEMININ

Dégradation de l'énergie.

theatredublog.unblog.fr · 2018/10/30 · entropie-chroniques-parisiennes...

Entropie, chroniques parisiennes, de Eric Da Silva, Henri ...

30 oct. 2018 - Entropie n.f. En physique : fonction exprimant le principe de la dégradation de l'énergie ; processus exprimé par cette fonction. Augmentation ...
Vous avez consulté cette page 3 fois. Dernière visite : 17/10/19

dictionnaire.lerobert.com · définition · entropie ▾

entropie - Définitions, synonymes, conjugaison, exemples | Le ...

27 févr. 2020 - Physique Fonction exprimant le principe de la dégradation de l'énergie ; processus exprimé par cette fonction. Augmentation du désordre ...

XIX, il est devenu évident que le processus de dégradation de l'énergie, à savoir le travail utile qui peut être obtenu dans un système isolé l'énergie pas tout est conservé, mais il y a une partie de cette énergie qui se dégrade (**entropie, deuxième principe de la thermodynamique**).

Définition de l'énergie

energie-nucleaire.net/definitions/energie.html

Was this helpful?

Dégradation de l'énergie | Le site de la Fondation La ...

Translate this page

<https://www.fondation-lamap.org/fr/page/11971/degredation-de-lenergie> ▾

Les machines thermiques sont les seules machines à pouvoir transformer de la chaleur en travail. L'étude des machines thermiques montre que la chaleur reçue par celles-ci (de la part d'un mélange combustible-air, par exemple) ne peut pas être entièrement transformée en travail.

Annexe 2 : Récolte des séances pédagogiques de l'indexation automatique

Niveau	Source	Titre + url séance	Auteur, Année	Notion(s) abordée(s)
Lycée - 2nde GT	Wikinotions	Une année avec Google (suite) – Peut-on enseigner Google http://documentation.spip.ac-rouen.fr/?Une-annee-avec-Google-suite Analyse ID'Base http://idbase.esmeree.fr/notice/e54fincw/24-une-annee-avec-google-suite	Frédéric Rabat, 2008	Moteur de recherche Indexation automatique Mots-clés
Lycée - 2nde GT	Wikinotions	Comment fonctionne un moteur de recherche : l'exemple de Google http://beaumont-redon.fr/wp/cultureinfomedias/2013/11/20/comment-fonctionne-un-moteur-de-recherche-l'exemple-de-google/	Gaëlle Sogliuzzo, 2013	Moteur de recherche Requête Mots-clés Base de données Index Evaluation de l'information Popularité
Lycée 2nde G - SNT	Édubase Professeurs documentalistes et CDI	Le Web https://doc.dis.ac-guyane.fr/SNT-Le-WEB.html	Perrine Chambaud, 2020	Moteur de recherche Moteur d'indexation
Lycée - 2nde GT	Eduscol	Moteurs de recherche et bulle de filtre https://disciplines.ac-toulouse.fr/documentation/sites/documentation/files/fichiers/2-_moteurs_de_recherche_et_bulles_de_filtres.pdf	Jane Cayrel, 2019/2020	Moteur de recherche Indexation automatique
Collège - 6ème	ID Base Professeurs documentalistes de l'académie de Toulouse	Classer les informations sur le Web http://idbase.esmeree.fr/notice/pbrsax3a/38-classer-linformation-sur-le-web-initiation https://docabord.wordpress.com/2013/06/19/seance-7-culture-net-classer-linformation-sur-le-web-initiation-techniques-documentaires-6eme/	Marie Guillet, 2013	Classement Indexation humaine Indexation automatique
Collège - 3ème	Wikinotions	Espace et document au CDI http://Wikinotions.apden.org/index.php?title=Espaces_et_documents_du_CDI	Florian Reynaud, 2014	Indexation Indexation documentaire
Collège – 5ème - EMI	Edubase Doc'Poitiers	Comprendre le fonctionnement des moteurs de recherche http://ww2.ac-poitiers.fr/doc/spip.php?article855	2019	Moteur de recherche Moteur d'indexation
Collège	Blog professionnel Marthe au CDI	Moteur de recherche http://martheaucdi.overblog.com/ Présentation prezi https://prezi.com/6mq76ky-g1mf/moteur-de-recherche/	2012	Moteur de recherche Moteur d'indexation
Collège	Blog professionnel L'Odysée d'LN : je tisse ma toile	Réseau social (Babélio) – défi lecture http://odysseedln.overblog.com/reseau-social-defi-lecture	Hélène Mulot, 2013	Indexation humaine libre (tags, folksonomie)
Collège - 6ème	Blog professionnel Fenêtre sur	Comment comprendre le fonctionnement d'un moteur de recherche – Vidéo 1m30 https://fenetresur.wordpress.com/?s=indexation	Claire Cassaigne, 2013	Moteur de recherche Moteur d'indexation Mots-clés

Annexe 3 : 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
	AUTRE (précisez)

Précision

- **Q1d - Dans quel cadre pédagogique ?**

	INFORMATION-DOCUMENTATION
	EPI
	AP
	SNT
	INTERDISCIPLINARITÉ
	AUTRE (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 ?

<input type="checkbox"/>	ENCART WIKIPÉDIA
<input type="checkbox"/>	ENCART IMAGES OU VIDÉOS
<input type="checkbox"/>	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 ?

Q 4 – 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 ?

Annexe 4 : Données brutes du questionnaire et catégorisation des verbatims

Indexation automatique et niveaux d'enseignement

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1b ... pour quel(s) niveau(x) d'enseignement ?	Collège				Lycée			Post-Bac	... si non pourquoi ?	Catégories (pourquoi Non)
				6è	5è	4è	3è	2nde	1ère	Tale	Post-Bac		
1	1	OUI	6è, 5è	1	1								
2	3667809	OUI	6è, 5è-4è	1	1	1							
3	3695407	OUI	6e 5e	1	1								
4	3699309	OUI	Abordé en sixième – Mis en question en troisième	1			1						
5	3728296	NON										Depuis la réforme du collège, le module info-documentation destiné au 6e a été réduit d'une séance par quinzaine sur l'année à 6 séances en tout... Concernant les autres niveaux, je vois certaines notions très ponctuelles d'EMI avec quelques collègues. Mais la réflexion sur les moteurs de recherche n'est quasiment jamais abordée par manque de temps.	Manque de temps
6	3825041	OUI	Niveau 4ème			1							
7	3842152	OUI	2nde					1					
8	3842699	NON										Zero formation sur le sujet (vieille doc de plus de 40 ans) : j'vois bien ce que c'est maintenant qu'on me le dit mais avant, en paraphrasant M. Jourdain "j'en vois sans le savoir" et de toute façon pas suffisamment de créneaux horaires avec les mêmes élèves pour mettre en place dans de bonnes conditions (ne serait-ce qu'en temps de préparation) une vraie progression des apprentissages documentaires depuis la 6e jusqu'à la terminale où éventuellement un tel sujet (pointu quand même) pourrait trouver sa place.	Manque de temps Sujet complexe Non maîtrise personnelle De la notion

Indexation automatique et niveaux d'enseignement

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1b ... pour quel(s) niveau(x) d'enseignement ?	Collège				Lycée			Post-Bac	... si non pourquoi ?	Catégories (pourquoi Non)
				6è	5è	4è	3è	2nde	1ère	Tale	Post-Bac		
9	3842736	NON										Je maîtrise moi-même assez mal ce concept... et je ne m'y suis jamais intéressée	Non maîtrise personnelle De la notion
10	3843204	NON										Je n'ai pas beaucoup de séances avec les élèves donc je vais au plus utile...	Manque de temps
11	3843245	NON											
12	3844210	OUI	collège	1	1	1	1						
13	3844587	OUI	sixième	1									
14	3845747	OUI	6e	1									
15	3846066	OUI	2nde LGT 2nde LP					1					
16	3846331	OUI	2de					1					
17	3847078	OUI	seconde					1					
18	3847180	NON											
19	3847417	NON										Thème n'étant pas dans les programmes, et pas utilisé non plus durant la semaine de la presse à l'école.	Pas dans les programmes Pas dans l'EMI
20	3847759	OUI	Seconde					1					
21	3847795	NON										Parce que c'est quelque chose que je connais peu et que je ne maîtrise pas	Non maîtrise personnelle De la notion
22	3847993	NON										trop complexe en 6eme	Sujet complexe au collège

Indexation automatique et niveaux d'enseignement

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1b ... pour quel(s) niveau(x) d'enseignement ?	Collège				Lycée			Post-Bac	... si non pourquoi ?	Catégories (pourquoi Non)
				6è	5è	4è	3è	2nde	1ère	Tale	Post-Bac		
23	3848738	NON										Je n'ai pas vraiment l'occasion d'aborder cette notion avec les élèves. En 6ème, c'est encore un peu complexe. Pour le Cycle 4, je n'ai malheureusement pas l'occasion d'avoir des séances avec les élèves, ou en tout cas, je n'aborde pas ce thème, mais davantage les images et la manipulation, les recherches documentaires...	Sujet complexe au collège Manque de temps
24	3849530	NON										Par manque de temps	Manque de temps
25	3850259	OUI	COLLEGE - LYCEE - ENSEIGNEMENT SUPERIEUR	1	1	1	1	1	1	1	1		
26	3851255	NON										Trop compliqué pour des élèves de 6ème, seul niveau que j'ai en EMI. Je l'aborde à travers le fonctionnement d'un moteur de recherche et de la notion de mot clé mais sans trop rentrer dans les détails.	Sujet complexe au collège
27	3861256	NON										Je ne sais pas ce que c'est.	Non maîtrise personnelle De la notion
28	3866343	OUI	6°	1									

Indexation automatique et niveaux d'enseignement

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1b ... pour quel(s) niveau(x) d'enseignement ?	Collège				Lycée			Post-Bac	... si non pourquoi ?	Catégories (pourquoi Non)
				6è	5è	4è	3è	2nde	1ère	Tale	Post-Bac		
29	3874475	NON										Les élèves viennent peu au CDI pour faire des séances de recherches documentaires. Nous essayons de proposer des séances à toutes les 11 classes de secondes à minima 1 ou 2 fois dans l'année, mais c'est trop peu pour aborder l'indexation automatique. Les élèves travaillent la question des moteurs de recherche en SNT. Pour cette première année de mise en place de la discipline SNT, nous avons axé notre travail sur les questions de culture numérique et de médias sociaux. A voir pour les années à venir...	Manque de temps
30	3883589	OUI	Tout le collège	1	1	1	1						
31	3883990	OUI	je l'ai dit au dessus								1		
32	3885350	OUI	2nde					1					
33	3885411	OUI											
34	3885522	OUI	Tous les niveaux d'enseignement au collège (6è-3è)	1	1	1	1						
35	3885594	OUI	2nde					1					
36	3885695	OUI	5ème		1								
37	3887292	OUI	de la 6e à la terminale	1	1	1	1	1	1	1			

Indexation automatique et niveaux d'enseignement

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1b ... pour quel(s) niveau(x) d'enseignement ?	Collège				Lycée			Post-Bac	... si non pourquoi ?	Catégories (pourquoi Non)
				6è	5è	4è	3è	2nde	1ère	Tale	Post-Bac		
38	3888417	NON										pas de possibilité d'avoir des heures de cours dédiées (je me bats déjà beaucoup pour avoir des heures info-documentaires de base), et notion trop difficile à aborder en une heure sans préparation avec les élèves et sans suivi.	Manque de temps Sujet complexe au collège
39	3888525	OUI	2nde					1					
40	3889198	NON										Le niveau de mes élèves et leurs besoins sont beaucoup plus basiques et ne me permettent pas d'en arriver là.	Sujet complexe au collège
41	3893876	NON	collège	1	1	1	1					Cela me semble complexe et moi-même je ne maîtrise pas le sujet !	Sujet complexe au collège Non maîtrise personnelle De la notion
42	3896242	NON										Le niveau de mes élèves et leurs besoins sont beaucoup plus basiques et ne me permettent pas d'en arriver là.	Sujet complexe au collège
43	3898313	OUI											
44	3911410	NON										Pas eu l'occasion d'intégrer ce point dans une séance	Pas l'occasion
45	4018186	OUI	6e, 5e, 4e, 3e	1	1	1	1						
46	4052318	NON										Pas d'heures pour le faire.	Manque de temps

Indexation automatique et notions génériques

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1c ... en lien avec quelles notions génériques ?				Précision
			MOTEUR DE RECHERCHE	PORTAIL DOCUMENTAIRE	VEILLE	AUTRE	
1	1	OUI	1	1			« Méthodologie e-sidoc »
2	3667809	OUI	1	1		1	« Identité numérique (Traçabilité) »
3	3695407	OUI	1				
4	3699309	OUI	1				
5	3728296	NON					
6	3825041	OUI	1				
7	3842152	OUI	1				
8	3842699	NON					
9	3842736	NON					
10	3843204	NON					
11	3843245	NON					
12	3844210	OUI	1				« Le fonctionnement des moteurs de recherche et notamment Google et le contenu des Data Centers. »
13	3844587	OUI	1				
14	3845747	OUI	1				
15	3846066	OUI	1	1			
16	3846331	OUI	1				
17	3847078	OUI	1				« Et plus précisément dans la lecture des résultats et de la mise en évidence des processus qui y conduise »
18	3847180	NON					
19	3847417	NON					
20	3847759	OUI	1	1			
21	3847795	NON					
22	3847993	NON					
23	3848738	NON					
24	3849530	NON					

Indexation automatique et notions génériques

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1c ... en lien avec quelles notions génériques ?				Précision
			MOTEUR DE RECHERCHE	PORTAIL DOCUMENTAIRE	VEILLE	AUTRE	
25	3850259	OUI	1				
26	3851255	NON					
27	3861256	NON					
28	3866343	OUI	1				
29	3874475	NON					
30	3883589	OUI	1				
31	3883990	OUI	1				travailler l'algorithme de google
32	3885350	OUI	1				
33	3885411	OUI	1				
34	3885522	OUI		1			
35	3885594	OUI	1				
36	3885695	OUI	1	1			
37	3887292	OUI	1	1	1		
38	3888417	NON					
39	3888525	OUI	1	1			« Lorsqu'on aborde le fonctionnement des moteurs de recherche, on peut comparer avec E-sidoc. »
40	3889198	NON					
41	3893876	NON					
42	3896242	NON					
43	3898313	OUI	1				
44	3911410	NON					
42	3896242	NON					
43	3898313	OUI	1			1	
44	3911410	NON					

Indexation automatique et notions génériques

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1c ... en lien avec quelles notions génériques ?				Précision
			MOTEUR DE RECHERCHE	PORTAIL DOCUMENTAIRE	VEILLE	AUTRE	
45	4018186	OUI	1	1			« Observer la pertinence des résultats d'une recherche sur Google par exemple. Apprendre les différents critères de recherche avancée sur E-Sidoc Terminologie et notion complexe qui nécessitent du temps d'enseignement dont je ne dispose pas suffisamment. »
46	4052318	NON					

Indexation automatique et cadre pédagogique

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1d ... dans quel(s) cadre(s) pédagogique(s) ?					Précision
			INFORMATION-DOCUMENTATION	EPI	AP	SNT	INTERDISCIPLINARITÉ	
1	1	OUI	1					
2	3667809	OUI	1					1 « Education à l'hygiène numérique »
3	3695407	OUI	1					« Cours 6e sur les moteurs de recherche »
4	3699309	OUI	1					
5	3728296	NON						
6	3825041	OUI		1				« EPI presse et média : recherche préalable à la réalisation d'un journal radiophonique. »
7	3842152	OUI				1		
8	3842699	NON						
9	3842736	NON						
10	3843204	NON						
11	3843245	NON						
12	3844210	OUI	1					« Cours EMI en 6ème (1h par quinzaine en demi-groupe) tout au long de l'année. + Ponctuellement en fonction des projets avec les collègues des autres classes. »
13	3844587	OUI	1					
14	3845747	OUI	1					
15	3846066	OUI	1		1	1	1	
16	3846331	OUI	1			1		
17	3847078	OUI			1		1	« En lycée, il n'y a plus beaucoup d'AP, souvent fléché français ou math, donc on prend le support horaire que l'on peut »
18	3847180	NON						
19	3847417	NON						
20	3847759	OUI	1		1			
21	3847795	NON						
22	3847993	NON						
23	3848738	NON						
24	3849530	NON						
25	3850259	OUI	1				1	
26	3851255	NON						
27	3861256	NON						
28	3866343	OUI			1			
29	3874475	NON						

Indexation automatique et cadre pédagogique

Séquentiel	SID	Q1a. Abordez-vous la notion d'indexation automatique ?	Q1d ... dans quel(s) cadre(s) pédagogique(s) ?						Précision
			INFORMATION-DOCUMENTATION	EPI	AP	SNT	INTERDISCIPLINARITÉ	AUTRE	
30	3883589	OUI	1						
31	3883990	OUI					1		« maths sup, maths spé »
32	3885350	OUI	1			1			
33	3885411	OUI	1						
34	3885522	OUI	1						
35	3885594	OUI				1			
36	3885695	OUI	1						
37	3887292	OUI					1		
38	3888417	NON							
39	3888525	OUI				1			
40	3889198	NON							
41	3893876	NON	1						« J'aborde l'indexation automatique uniquement pour expliquer aux élèves qu'un moteur de recherche s'appuie sur leurs mots-clés pour trouver des pages web et cela grâce à des robots. Je le fais à la fois en 6e lors de séances dédiées aux moteurs de recherches (cadre : initiation à la recherche d'informations, séances prévues à l'année en 6e) et à la fois "à la volée" lorsque j'ai des élèves en permanence au CDI et qui sont "bloqués" lors de leur recherche d'infos (dans ce cas : élèves de la 6e à la 3e). »
42	3896242	NON							
43	3898313	OUI	1					1	
44	3911410	NON							
45	4018186	OUI	1	1	1		1	1	« EMI, divers travaux de recherche en lien avec les autres disciplines, souvent au cas par cas. »
46	4052318	NON							

Indexation en amont de l'indexation automatique

Séquentiel	SID	Q2. Abordez-vous la notion d'indexation en amont de la notion d'indexation automatique ?	Si oui, pourquoi	Catégories	Si non, pourquoi ? Cette question a été majoritairement considérée comme redondante par rapport à la justification de la réponse « non » à l'enseignement de l'indexation automatique
1		NON			En même temps
2	3667809	NON			Les deux en même temps
3	3695407	OUI	Notion d'indexation que j'évoque en lien avec l'index des clés du livre	Lié à l'index du livre	
4	3699309	OUI	Pour que les élèves questionnent l'intérêt d'indexer : A quoi sert d'indexer une page web, un site web ?	Indexation notion générique	
5	3728296	NON			Pour les mêmes raisons citées précédemment.
6	3825041	OUI	pour une meilleure compréhension de la notion.	Indexation notion générique	
7	3842152	NON			Manque de temps
8	3842699	NON			voir plus haut
9	3842736	NON			Je n'aborde pas ces notions.
10	3843204	NON			
11	3843245	NON			
12	3844210	OUI	Avec la recherche sur le portail esidoc (indexation réalisée par des humains) et sur le début des moteurs de recherche avec indexation humaine.	Indexation humaine vs. Indexation automatique	
13	3844587	OUI	recherche par mots clés sur esidoc	Processus de l'indexation	
14	3845747	OUI	J'aborde la notion via l'index d'un livre documentaire	Lié à l'index	
15	3846066	OUI	Pour expliquer la notion de manière générale d'abord et ensuite la notion plus spécifique d'indexation automatique. Pour rester clair avec les élèves et ne pas les perdre dans leur processus de compréhension.	Indexation notion générique	

Indexation en amont de l'indexation automatique

Séquentiel	SID	Q2. Abordez-vous la notion d'indexation en amont de la notion d'indexation automatique ?	Si oui, pourquoi	Catégories	Si non, pourquoi ? Cette question a été majoritairement considérée comme redondante par rapport à la justification de la réponse « non » à l'enseignement de l'indexation automatique
16	3846331	OUI	Les 2 notions sont liées. L'indexation automatique vient des pratiques d'indexation qui l'ont précédées et qui existent toujours. Il est important de replacer les notions abordées dans leur filiation historique, ainsi que dans un réseau notionnel.	Processus de l'indexation	
17	3847078	OUI	En tant que notion élémentaire que l'on décline ensuite en types d'indexation (sans trop pousser... mais en transmettant qu'il y en a plusieurs : keywords des métadonnées, tag, poids différent des occurrences des termes de la recherche qui pondère et hiérarchise les résultats...)	Indexation notion générique	
18	3847180	NON			
19	3847417	NON			Parce que ce n'est pas utilisé par les collègues dans les programmes ou projets.
20	3847759	OUI	Compréhension de la procédure de description et de référencement d'un document : mots clés, tags...	Processus de l'indexation	
21	3847795	NON			
22	3847993	NON			
23	3848738	NON			même réponse que précédemment
24	3849530				
25	3850259	OUI			
26	3851255	NON			
27	3861256	NON			Car je n'aborde pas la notion d'indexation automatique.
28	3866343	OUI	Dans le cadre de la présentation de la base documentaire et du portail E-Sidoc	Indexation humaine vs. Indexation automatique	

Indexation en amont de l'indexation automatique

Séquentiel	SID	Q2. Abordez-vous la notion d'indexation en amont de la notion d'indexation automatique ?	Si oui, pourquoi	Catégories	Si non, pourquoi ? Cette question a été majoritairement considérée comme redondante par rapport à la justification de la réponse « non » à l'enseignement de l'indexation automatique
29	3874475	NON			Pour les mêmes raisons que celles données précédemment.
30	3883589	OUI	Pitzee les livres pour comprendre le principe	Lié à l'index du livre	
31	3883990	OUI	on se rapproche des cours d'informatique	Processus de l'indexation	
32	3885350	OUI	Connaissance nécessaire du principe général	Indexation notion générique	
33	3885411	NON			
34	3885522	OUI			
35	3885594	OUI	Parce que c'est une notion complexe pour des secondes et pour introduire la notion.	Indexation notion générique	
36	3885695	OUI	je pense que c'est un pré-requis.	Indexation notion générique	
37	3887292	OUI	Afin qu'ils prennent conscience de la démarche initiale.	Processus de l'indexation	
38	3888417	NON			même chose
39	3888525	NON			Pas le temps
40	3889198				
41	3893876	NON			idem
42	3896242	NON			Même raison hélas
43	3898313	NON			manque de temps. Je l'aborde tout de même un peu lors des séances de découverte du CDI en 6e : je leur fais réaliser quelques citations de livres afin qu'ils comprennent le classement des documents. Mais c'est assez superficiel : je ne cite pas forcément le terme d'indexation devant les élèves, l'objectif étant surtout qu'ils sachent se repérer de manière autonome dans le CDI.

Indexation en amont de l'indexation automatique

Séquentiel	SID	Q2. Abordez-vous la notion d'indexation en amont de la notion d'indexation automatique ?	Si oui, pourquoi	Catégories	Si non, pourquoi ? Cette question a été majoritairement considérée comme redondante par rapport à la justification de la réponse « non » à l'enseignement de l'indexation automatique
44	3911410	NON			
45	4018186	OUI	Oui, pour expliquer les résultats de recherche sur E-SIDOC ou sur tout autre moteur de recherche. Explication souvent orale.	Processus de l'indexation	
46	4052318	NON			Manque d'heures.

Infobox remarqués

Séquentiel	SID	Q3a. Avez-vous remarqué ces encarts informationnels sur une page de résultats ?	... lesquels ?		
			Encart Wikipédia	Encarts images ou vidéos	Encart de définition(s)
1		OUI	1	1	1
2	3667809	OUI	1	1	1
3	3695407	OUI	1		1
4	3699309	OUI			1
5	3728296	OUI	1		1
6	3825041	OUI	1		
7	3842152	OUI	1	1	1
8	3842699	OUI	1		
9	3842736	OUI	1		
10	3843204	OUI	1	1	1
11	3843245	OUI	1		1
12	3844210	OUI	1		1
13	3844587	OUI	1		
14	3845747	OUI	1		1
15	3846066	OUI	1	1	1
16	3846331	OUI	1	1	1
17	3847078	OUI	1	1	1
18	3847180	OUI	1		
19	3847417	OUI	1		1
20	3847759	OUI	1	1	
21	3847795	NON			
22	3847993	OUI	1		
23	3848738	OUI	1	1	1
24	3849530	OUI	1		
25	3850259	OUI			
26	3851255	OUI	1		
27	3861256	OUI	1		1
28	3866343	OUI		1	1
29	3874475	OUI	1	1	1
30	3883589	OUI	1	1	1
31	3883990	OUI	1	1	1
32	3885350	OUI	1	1	1
33	3885411	OUI	1	1	1
34	3885522	OUI	1		
35	3885594	OUI	1	1	1
36	3885695	OUI	1	1	1
37	3887292	OUI	1	1	
38	3888417	OUI	1	1	
39	3888525	OUI	1		1
40	3889198	OUI	1		1
41	3893876	OUI	1		1
42	3896242	OUI	1		1
43	3898313	OUI	1	1	
44	3911410	OUI	1		1
45	4018186	OUI	1		1
46	4052318	OUI	1	1	

Encarts informationnels

Sé- quen- tiel	SID	Q3b. Abor- dez-vous la question des encarts in- formationnels avec les élèves ?	Si oui, que souhaiteriez-vous que les élèves retiennent de l'étude des encarts informationnels ?	Catégories	Si non, pourquoi ?	Catégories
1		NON				
2	3667809	OUI	Qu'ils relèvent d'un processus automatique, de l'intelligence artificielle	Fonctionnement		
3	3695407	NON			Je ne me suis jamais posée la question de ces encarts	Pas eu l'idée
4	3699309	NON				
5	3728296	NON			Idem	Manque de temps
6	3825041	OUI	je souhaite attirer leur attention sur les pistes de recherches qu'ils peuvent ainsi affiner.	Outil de recherche d'information		
7	3842152	NON				
8	3842699	NON			voir plus haut	Manque de temps Sujet complexe Non maîtrise de la notion
9	3842736	OUI	L'ordre d'apparition / le contenu	Evaluation de l'in- formation		
10	3843204	NON			Je n'ai pas eu encore le temps d'en parler ni l'occasion.	Manque de temps
11	3843245	NON				
12	3844210	OUI	Que c'est un bon moyen de trouver les mots-clés, les infos essentielles présentées en langage "télégraphique".	Outil de recherche d'information		
13	3844587	NON			manque de temps	Manque de temps
14	3845747	NON			Parce que je n'ai pas encore eu le temps (ou pris le temps) de didactiser cette question ! Mais c'est en réflexion.	En cours de réflexion

Encarts informationnels

Sé- quen- tiel	SID	Q3b. Abor- dez-vous la question des encarts in- formationnels avec les élèves ?	Si oui, que souhaiteriez-vous que les élèves retiennent de l'étude des encarts informationnels ?	Catégories	Si non, pourquoi ?	Catégories
15	3846066	NON			L'occasion ne s'est pas encore présentée et le temps nécessaire pour mener à bien les 3 axes de notre circulaire est déjà très précieux. Toutefois, la réflexion que votre questionnaire initie chez moi, engendre l'envie de m'intéresser de plus près à cette question et l'étudier pour la proposer aux collègues de discipline. Je vous en remercie.	En cours de réflexion
16	3846331	NON			Le contexte des séquences ne s'y prêtait pas.	Pas dans les projets pédagogiques
17	3847078	OUI	Notion intéressante pour l'évaluation de la page de résultats et pour décrypter la mécanique de la boîte noire des algorithmes.	Évaluation de l'in- formation Fonctionnement		
18	3847180	NON			Pas le temps	Manque de temps
19	3847417	NON			Pas encore abordé en projet.	Pas dans les projets pédagogiques
20	3847759	NON			Ce n'était pas un objectif d'apprentissage pour l'instant	Pas dans les projets pédagogiques
21	3847795					
22	3847993	NON			pas encore mais je vais l'intégrer à ma pratique	En cours de réflexion
23	3848738	NON			même réponse que précédemment	Non maîtrise de la notion
24	3849530	NON			par manque de temps	Manque de temps
25	3850259	OUI				
26	3851255	NON			Trop de notions à faire passer . Pas assez de temps .	Manque de temps

Encarts informationnels

Sé- quen- tiel	SID	Q3b. Abor- dez-vous la question des encarts in- formationnels avec les élèves ?	Si oui, que souhaiteriez-vous que les élèves retiennent de l'étude des encarts informationnels ?	Catégories	Si non, pourquoi ?	Catégories
27	3861256	OUI	Qu'il ne faut pas prendre la réponse pour argent comptant. Qu'il faut repérer la source de l'information.	Évaluation de l'in- formation		
28	3866343	NON			Manque de temps...	Manque de temps
29	3874475	NON			Toujours pour les mêmes raisons.	Manque de temps
30	3883589	NON			Pas eu l'occasion	Pas dans les projets péda- gogiques
31	3883990	OUI	d'éviter les encarts informationnels ou de pub et de se concentrer sur l'essentiel	Évaluation de l'in- formation		
32	3885350	OUI	Qu'ils ne sont pas une source unique de renseignements Qu'ils peuvent permettre de réorienter la recherche vers des concepts non envisagés qu'ils les différencient de la pub	Évaluation de l'in- formation		
33	3885411	NON				
34	3885522	NON				
35	3885594	OUI	La subjectivité des résultats obtenus, la différence avec les SEM, la complexité des algorithmes des moteurs de recherche.	Évaluation de l'in- formation Fonctionnement		
36	3885695	OUI	Info en condensée, accessible mais pas développée. C'est le début, avant approfondissement.	Outil de recherche d'information		
37	3887292	NON			Manque de temps pour aborder ce sujet	Manque de temps
38	3888417	NON			manque de temps et pas d'emploi du temps dédié des élèves	Manque de temps
39	3888525	NON			Pas l'occasion	Pas dans les projets péda- gogiques
40	3889198	OUI	Sources utilisées dans l'encart	Évaluation de l'in- formation		
41	3893876	NON				

Encarts informationnels

Sé- quen- tiel	SID	Q3b. Abor- dez-vous la question des encarts in- formationnels avec les élèves ?	Si oui, que souhaiteriez-vous que les élèves retiennent de l'étude des encarts informationnels ?	Catégories	Si non, pourquoi ?	Catégories
42	3896242	NON			Même raison (je suis en lycée profession- nel du bâtiment et notre problème primor- dial est d'amener les élèves au travail quel qu'il soit ; ceci étant, je retiens l'idée !)	Sujet complexe
43	3898313	NON			Je ne les aborde pas dans le cadre de séances pédagogiques car le contexte ne s'y prête pas en général (je parle de séances en co-animation avec des col- lègues de discipline). Je ne le fais pas non plus durant les séances d'EMI avec les 6e par manque de temps. En revanche j'en parle avec les élèves au cas par cas, lorsque je constate qu'ils prennent leurs infos dans ces infobox : je leur demande alors d'identifier le site source et je les incite à consulter le site et ne pas s'arrêter à ce résumé (contexte : lors de séances de recherches ou lors de recherches en autonomie sur des heures de permanence)	Manque de temps Pas dans les projets péda- gogiques Source (ponctuel)
44	3911410	NON			Même raison. En revanche, ce point a été abordé mais pas analysé.	Pas dans les projets péda- gogiques Notion complémentaire
45	4018186	NON			Pas encore et par manque de temps.	Manque de temps
46	4052318	NON			Manque d'heures.	Manque de temps

Séances pédagogiques

Séquentiel	SID	Q4. Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ?	Si oui, lesquelles	Catégories	Si non, pourquoi ?	Catégories
1		OUI				
2	3667809	OUI	<p>Pour le cours sur les moteurs de recherche en 6e = Adaptation de la séance de Claire Cassaigne (https://fenetresur.wordpress.com/2014/04/16/info-doc-les-moteurs-de-recherche-seance-9/) avec notamment la vidéo (https://www.youtube.com/watch?v=pMywV9ZLS4M&feature=youtu.be)</p> <p>En 6e, pour la comparaison catalogue informatisé du CDI / moteur de recherche et aboutir à la différenciation entre indexation humaine et indexation automatique, adaptation de la séance de Marie Nallathamby (https://docabord.wordpress.com/2013/06/19/seance-7-culture-net-classer-linformation-sur-le-web-initiation-techniques-documentaires-6eme/)</p>	Ressources professionnelles		
3	3695407	OUI				
4	3699309	OUI	Des schémas du classeur génération 5.	Editeur privé		
5	3728296	NON				
6	3825041	NON			L'indexation automatique est un point que j'aborde très succinctement.	Peu abordée
7	3842152	OUI	Celles de collègues prof doc	Ressources professionnelles		
8	3842699	NON			voir plus haut réponse 1	Non maîtrisée

Séances pédagogiques

Séquentiel	SID	Q4. Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ?	Si oui, lesquelles	Catégories	Si non, pourquoi ?	Catégories
9	3842736	NON				
10	3843204	NON				
11	3843245	NON				
12	3844210	NON			parce que je n'y passe pas beaucoup de temps, je l'aborde assez sommairement.	Peu abordée
13	3844587	OUI				
14	3845747	NON			Je ne comprends pas bien la question	
15	3846066	NON			J'ai besoin de créer moi-même mes outils pour m'approprier mes séances et les mener en tentant de les maîtriser au mieux parce que j'y ai beaucoup réfléchi en amont.	Préférence pour Séances personnelles
16	3846331	NON			Je n'en ai pas trouvé et n'ai pas réellement pris le temps d'en chercher. Il est souvent plus simple de créer ses propres séquences, celles existantes peuvent nous inspirer mais demandent beaucoup de travail d'adaptation pour les faire correspondre à nos objectifs.	Préférence pour Séances personnelles
17	3847078	NON			Manque de temps	Manque de temps
18	3847180	NON			Trop complexe pour les élèves et les enseignants, problème de temps	Non maîtrisée
19	3847417	NON			Je n'aborde pas l'indexation.	Non enseignée
20	3847759	OUI	L'accès aux sources d'information et l'info média	Ressources institutionnelles		

Séances pédagogiques

Séquentiel	SID	Q4. Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ?	Si oui, lesquelles	Catégories	Si non, pourquoi ?	Catégories
21	3847795	NON				
22	3847993	NON			pas encore...	Non enseignée
23	3848738	NON			idem	Manque de temps
24	3849530	NON				
25	3850259	NON				
26	3851255	NON				
27	3861256	NON			Je ne sais toujours pas ce qu'est l'indexation automatique.	Non maîtrisée
28	3866343	NON				
29	3874475	NON			Je n'utilise jamais de séances pédagogiques toutes faites.	Préférence pour Séances personnelles
30	3883589	OUI	Recherches	Moteur de recherche généraliste		
31	3883990	NON			parce qu'elles sont nulles et non avenues	Préférence pour Séances personnelles
32	3885350	NON			Je crée mes propres séances	Préférence pour Séances personnelles
33	3885411	NON				
34	3885522	NON				

Séances pédagogiques

Séquentiel	SID	Q4. Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ?	Si oui, lesquelles	Catégories	Si non, pourquoi ?	Catégories
35	3885594	NON			C'est une notion qui est juste abordée, elle ne fait pas vraiment l'objet d'une séance.	Peu abordée
36	3885695	OUI			Je ne comprends pas bien la question. Je mets en permanence à jour le travail que je propose aux élèves. De fait, ce sont mes séances pédagogiques existantes que j'adapte en fonction des évolutions des outils de recherche.	Préférence pour Séances personnelles
37	3887292	NON			Je l'évoque sans avoir le temps d'approfondir	Peu abordée
38	3888417	NON				
39	3888525	OUI	Celles sur les moteurs	Moteur de recherche généraliste		
40	3889198					
41	3893876	NON				
42	3896242	NON				
43	3898313	OUI	oui mais je ne sais plus lesquelles ! je pioche régulièrement dans les idées des collègues profs doc pour élaborer mes séances	Ressources professionnelles		
44	3911410	NON				
45	4018186	NON			Je ne fais pas une séance uniquement consacrée à l'indexation automatique.	Peu abordée
46	4052318	NON			Manque d'heures.	Manque de temps

Méthodes pédagogiques

Séquentiel	SID	Q5. Quelle.s méthode.s pédagogique.s vous semblent la.les plus appropriée.s pour enseigner l'indexation automatique, les moteurs de recherche ou les encarts informationnels	Catégories
1		Méthode active → Situation problème	Active
2	3667809	Méthodes actives	Active
3	3695407	Scénarios pédagogiques sous forme de situation problème	Active
4	3699309	Méthode active de type observationnel	Active
5	3728296	Dans le cadre d'une recherche documentaire prendre un temps sur les résultats dans un moteur de recherche serait intéressant et même essentiel pour développer leur esprit critique.	Interrogative
6	3825041	une comparaison des deux me parait pertinente.	Active
7	3842152		
8	3842699		
9	3842736		
10	3843204		
11	3843245		
12	3844210	les moteurs	HORS-SUJET
13	3844587		
14	3845747		
15	3846066	Ne m'étant pas encore penché sur les encarts informationnels, je crains que ma réponse ne soit biaisée. J'en suis désolée. De fait, je dirai par mon expérience les moteurs de recherche.	HORS-SUJET
16	3846331	Toutes les démarches didactiques restent possibles. Souvent, ce qui est intéressant c'est de mixer différentes démarches dans le déroulement d'une séquence ou séance.	Mixte
17	3847078	Sous un mode TP avec l'évaluation et la validation de l'information. Base pour une autonomie dans les stratégies de recherche d'information.	Expositive
18	3847180	Suivre les élèves dans un projet de recherche sur un temps long, quasiment plus possible avec la réforme du lycée !! Manque d'heures pour ce genre de projet	HORS-SUJET
19	3847417		
20	3847759		
21	3847795		
22	3847993	encarts	HORS-SUJET
23	3848738		
24	3849530		
25	3850259		
26	3851255		

Méthodes pédagogiques

Séquentiel	SID	Q5. Quelle.s méthode.s pédagogique.s vous semblent la.les plus appropriée.s pour enseigner l'indexation automatique, les moteurs de recherche ou les encarts informationnels	Catégories
27	3861256	Qu'est-ce qu'une méthode pédagogique ?	HORS-SUJET
28	3866343	Situations de recherches concrètes par plusieurs groupe d'élèves puis analyse et comparaison;	Active
29	3874475	Nous abordons ces thématiques à travers des sujets de recherche, idéalement intégrés dans un cours, pour que les séances soient cohérentes et constructives.	Expositive
30	3883589	La pratique avec d vraies recherches	Expositive
31	3883990	on travaille dans le dur, on met la main dans le cambouis on étudie le code les algo et les notions mathématiques afférentes	Expositive
32	3885350	recherche guidée	Interrogative
33	3885411		
34	3885522	Les pédagogies actives.	Active
35	3885594		
36	3885695	Recherches libres.	HORS-SUJET
37	3887292		
38	3888417		
39	3888525	Méthode inductive	Active
40	3889198		
41	3893876	j'utilise une courte vidéo et un schéma pour expliquer le fonctionnement d'un moteur de recherche	Expositive
42	3896242	La mise en situation pratique, en partant des observations des élèves sur les différents résultats proposés par différents moteurs de recherche (ceux qui sont mis en avant, pourquoi...) ou sur les encarts informationnels.	Active
43	3898313	L'apprentissage par l'expérience et la déduction. J'ai expérimenté différents types de séances avec des 6e et voici ce qui a le mieux fonctionné selon moi (en termes d'apprentissages des élèves mais aussi de motivation / implication de leur part) : - faire réaliser aux élèves des articles sur un blog et les faire "tagger" leurs propres articles - faire réaliser une recherche sur Google aux élèves et leur demander de repérer les "mots" en gras dans les snippet des résultats : les faire réfléchir en groupe afin qu'ils déduisent, par eux-même, qu'il s'agit des mots-clés qu'ils ont tapé dans le moteur de recherches.	Active
44	3911410	Dans le cadre d'une recherche et de l'organisation d'une page web	Expositive
45	4018186	Pour l'instant, les moteurs de recherche.	HORS-SUJET
46	4052318	Dans le cadre d'une recherche informationnelle.	HORS-SUJET

Expérience et académie

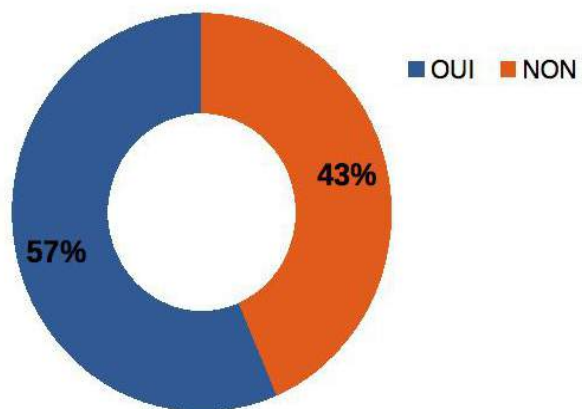
Séquentiel	SID	Q6. Depuis combien de temps exercez-vous en tant que professeur.e documentaliste ?	Q7. Dans quelle académie exercez-vous ?
1		24	NANTES
2	3667809	26	NANTES
3	3695407	10	NANTES
4	3699309	7	NANTES
5	3728296	8	NANTES
6	3825041	20	NANTES
7	3842152	16	NANTES
8	3842699	20	NANTES
9	3842736	13	NANTES
10	3843204	12	NANTES
11	3843245	3	NANTES
12	3844210	25	NANTES
13	3844587	18	NANTES
14	3845747	15	NANTES
15	3846066	2	NANTES
16	3846331	20	NANTES
17	3847078	27	NANTES
18	3847180	18	NANTES
19	3847417	10	NANTES
20	3847759	12	NANTES
21	3847795	12	NANTES
22	3847993	20	NANTES
23	3848738	10	NANTES
24	3849530	28	NANTES
25	3850259	15	NANTES
26	3851255	10	NANTES
27	3861256	3	NANTES
28	3866343	24	NANTES
29	3874475	5	NANTES
30	3883589	12	CLERMONT
31	3883990	30	RENNES
32	3885350	19	DIJON
33	3885411	20	GRENOBLE
34	3885522	4	LILLE
35	3885594	15	CAEN
36	3885695	7	NANCY-METZ
37	3887292	20	VERSAILLES
38	3888417	7	NANTES
39	3888525	14	GRENOBLE
40	3889198	4	NANTES
41	3893876	20	NANTES
42	3896242	11	NICE
43	3898313	2	RENNES
44	3911410	0,5	NANTES
45	4018186	28	NANTES
46	4052318	6	CRÉTEIL

Annexe 5 : Traitement analytique des données

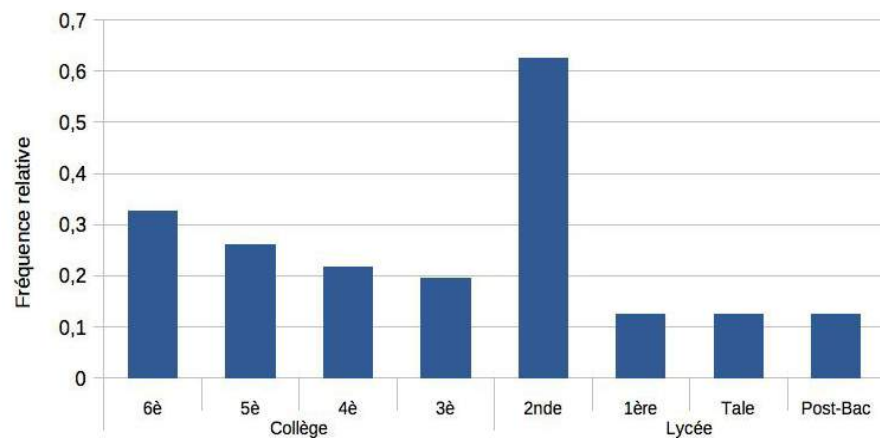
Annexe 5 - Traitement analytique

Q1a. Abordez-vous la notion d'indexation automatique ?			Q1b ... pour quel(s) niveau(x) d'enseignement ?				
	Total	Pourcentage			Total	Fréquence pondérée	Fréquence
OUI	26	58%	Collège (total=46)	6è	15	0,33	0,24
NON	20	44%		5è	12	0,26	0,19
				4è	10	0,22	0,16
				3è	9	0,20	0,15
			Lycée (total=16)	2nde	10	0,63	0,16
				1ère	2	0,13	0,03
				Tale	2	0,13	0,03
				Post-Bac	2	0,13	0,03

Enseignement de l'indexation automatique

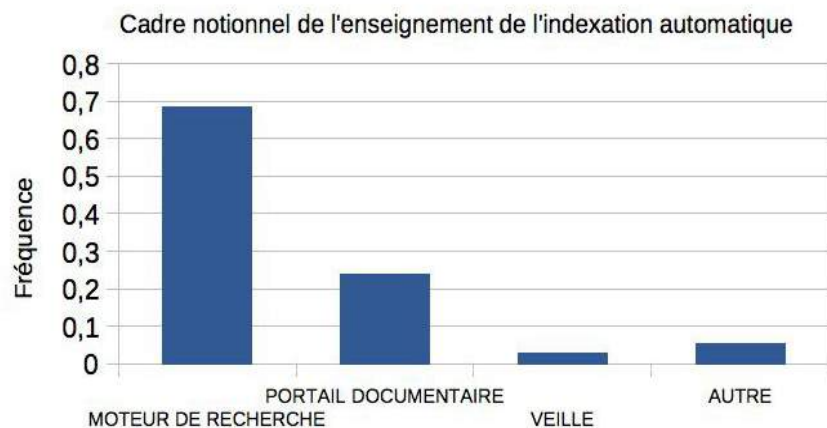


Niveaux d'enseignement de l'indexation automatique

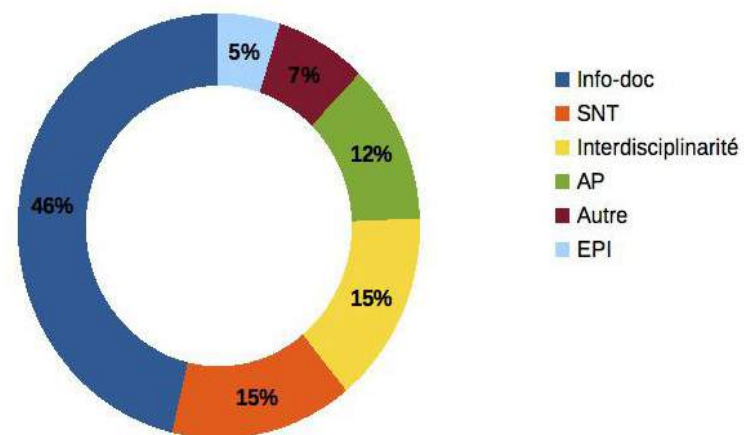


Annexe 5 - Traitement analytique

Q1c ... en lien avec quelles notions génériques ?			Q1d ... dans quel(s) cadre(s) pédagogique(s) ?		
	Total	Fréquence		Total	Fréquence
MOTEUR DE RECHERCHE	26	0,68	Info-doc	19	0,46
PORTAIL DOCUMENTAIRE	9	0,24	SNT	6	0,15
VEILLE	1	0,03	Interdisciplinarité	6	0,15
AUTRE	2	0,05	AP	5	0,12
Moteur+portail	8	0,21	Autre	3	0,07
			EPI	2	0,05



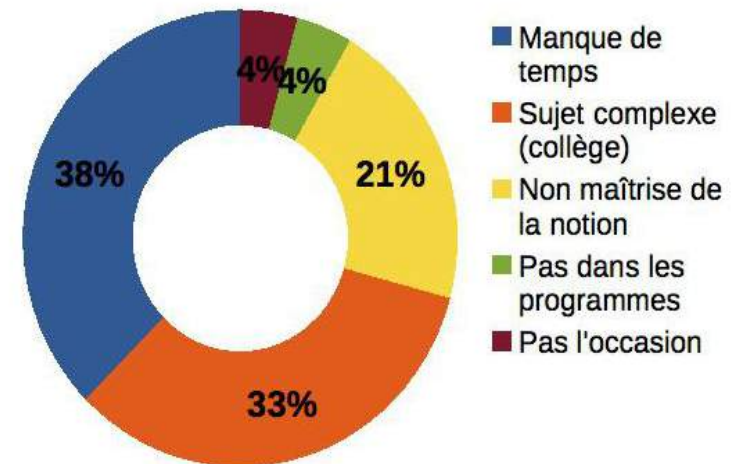
Cadre pédagogique de l'enseignement de l'indexation automatique



Annexe 5 - Traitement analytique

... si non, pourquoi ?		
Catégories	Total	Fréquence
Manque de temps	9	0,38
Sujet complexe (collège)	8	0,33
Non maîtrise de la notion	5	0,21
Pas dans les programmes	1	0,04
Pas l'occasion	1	0,04

Pourquoi l'indexation automatique n'est pas enseignée ?

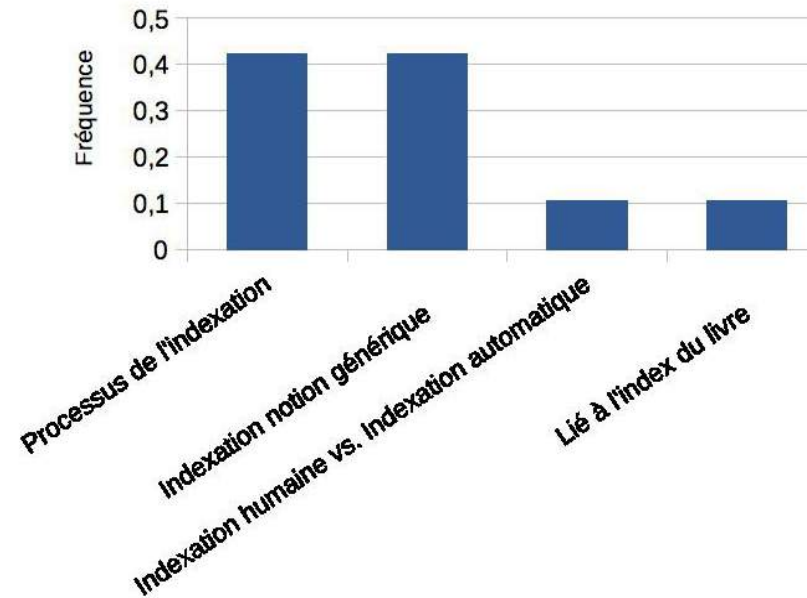
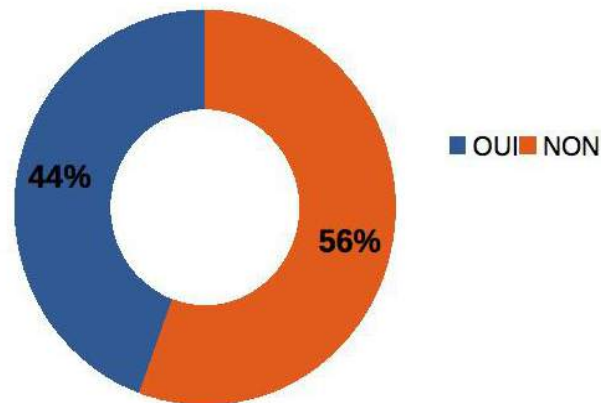


Annexe 5 - Traitement analytique

Q2. Abordez-vous la notion d'indexation en amont de la notion d'indexation automatique ?			Si oui, pourquoi ?		
	Total	Pourcentage	Catégories	Total	Fréquence
OUI	20	44%	Processus de l'indexation	8	0,42
NON	25	56%	Indexation notion générique	8	0,42
			Indexation humaine vs. Indexation automatique	2	0,11
			Lié à l'index du livre	2	0,11

Raisons de l'enseignement de l'indexation en amont de l'indexation automatique

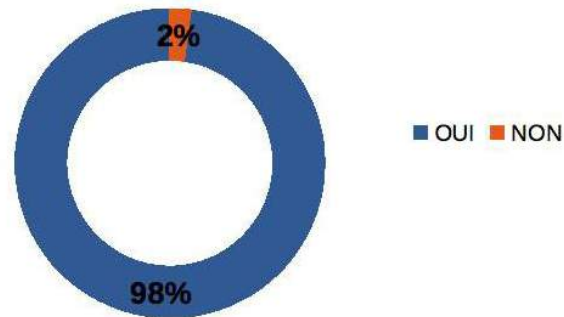
L'indexation enseignée en amont de l'indexation automatique



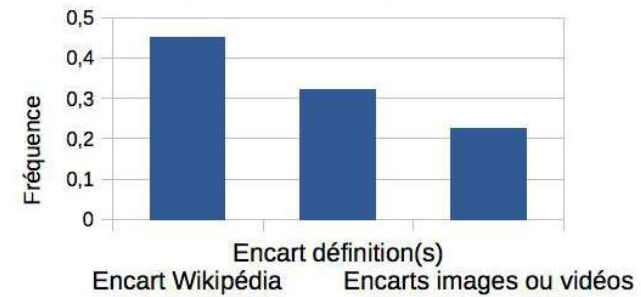
Annexe 5 - Traitement analytique

Q3a. Avez-vous remarqué ces encarts informationnels sur une page de résultats ?			... lesquels ?		
	Total	Pourcentage		Total	Fréquence
OUI	45	92	Encart Wikipédia	42	0,45
NON	1	2	Encart définition(s)	30	0,32
			Encarts images ou vidéos	21	0,23

Observation encarts informationnels

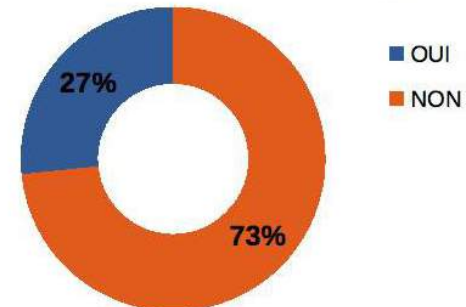


Encarts informationnels les plus repérés



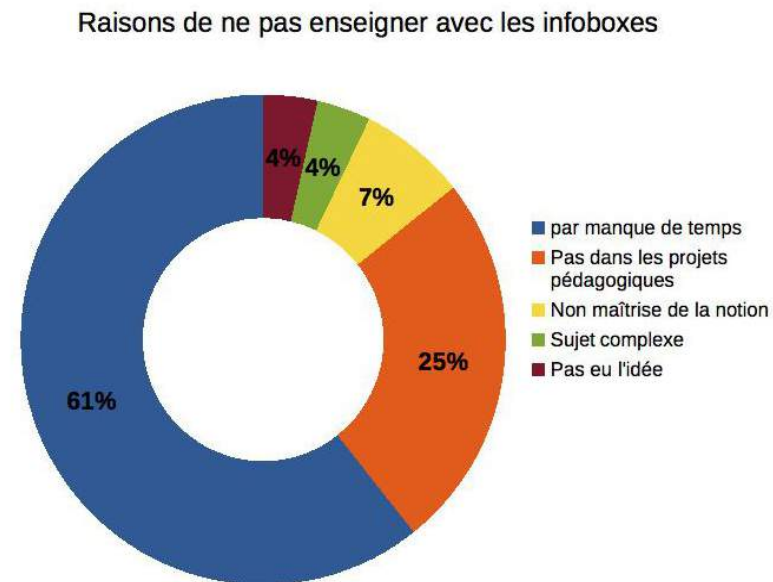
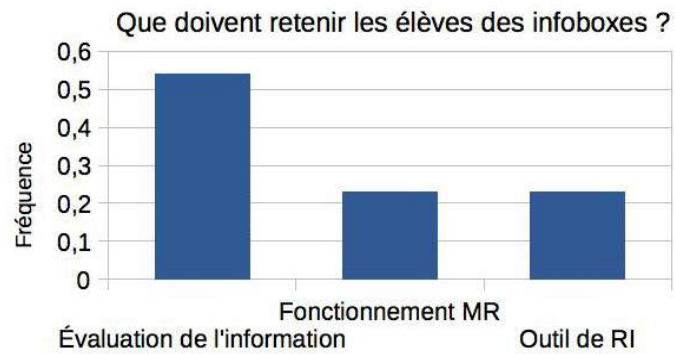
Q3b ... abordez-vous la question des encarts informationnels avec les élèves ?		
	Total	Fréquence
OUI	12	27%
NON	33	73%

Utilisation des infoboxes dans les enseignements



Annexe 5 - Traitement analytique

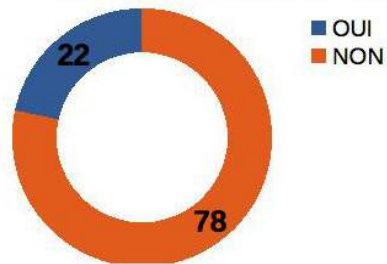
Q3c. Si oui, que souhaiteriez-vous que les élèves retiennent de l'étude des encarts informationnels ?			Si non, pourquoi ?		
Catégories	Total	Fréquence	Catégories	Total	Fréquence
Évaluation de l'information	7	0,54	Par manque de temps	17	0,61
Fonctionnement MR	3	0,23	Pas dans les projets pédagogiques	7	0,25
Outil de RI	3	0,23	Non maîtrise de la notion	2	0,07
			Sujet complexe	1	0,04
			Pas eu l'idée	1	0,04
			<i>En cours de réflexion</i>	3	0,11



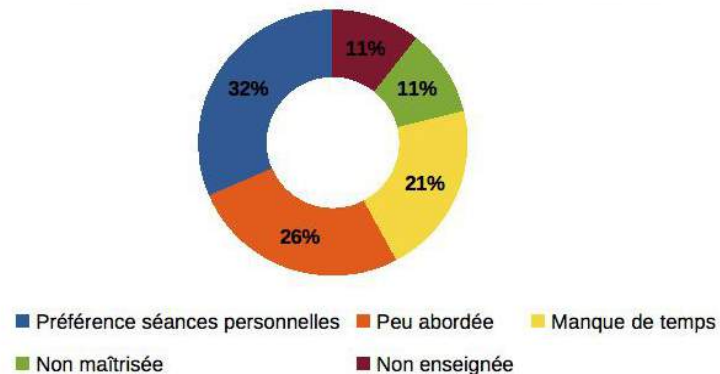
Annexe 5 - Traitement analytique

Q4. Utilisez-vous ou adaptez-vous des séances pédagogiques déjà existantes pour réaliser vos séances qui abordent l'indexation automatique ?			Si oui, lesquelles ?		
	Total	Pourcentage	Catégories	Total	Fréquence
OUI	10	22	Ressources professionnelles	3	0,43
NON	36	78	Moteurs de recherche généraliste	2	0,29
			Ressources institutionnelles	1	0,14
			Éditeur privé	1	0,14

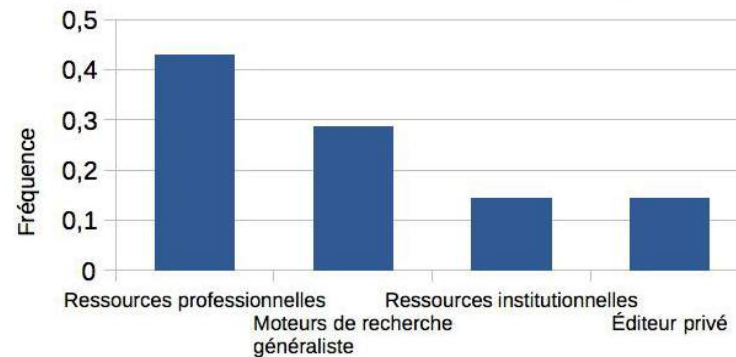
Utilisation de séances pédagogiques existantes



Raisons de ne pas utiliser de séances pé-existantes



Origine des ressources pour les séances pédagogiques



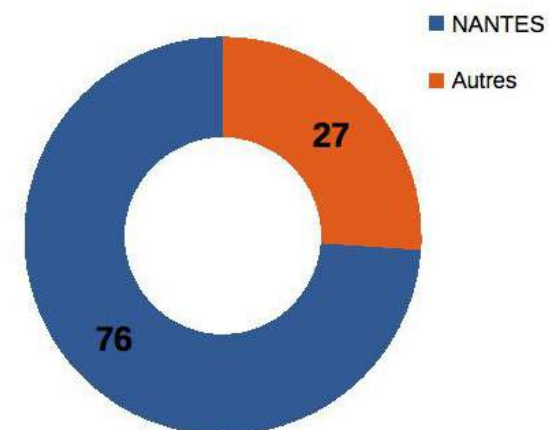
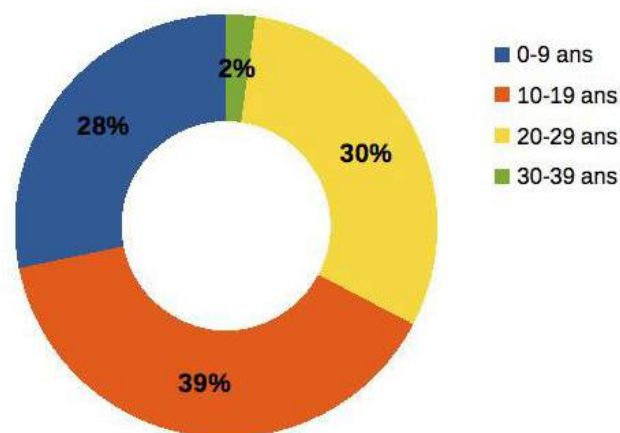
Si non, pourquoi ?		
Catégories	Total	Fréquence
Préférence séances personnelles	6	0,32
Peu abordée	5	0,26
Manque de temps	4	0,21
Non maîtrisée	2	0,11
Non enseignée	2	0,11

Annexe 5 - Traitement analytique

Q6. Depuis combien de temps exercez-vous en tant que professeur.e documentaliste ?			7. Dans quelle académie exercez-vous ?		
	Total	Pourcentage		Total	Pourcentage
0-9 ans	13	28,26	NANTES	34	75,56
10-19 ans	18	39,13	Autres	12	26,67
20-29 ans	14	30,43			
30-39 ans	1	2,17			
Moyenne	14 ans				

Origine académique des sondés

Années d'expérience



Annexe 5 - Traitement analytique

Contingence Q1a-Q6 : Relation entre l'expérience et l'enseignement de l'indexation automatique

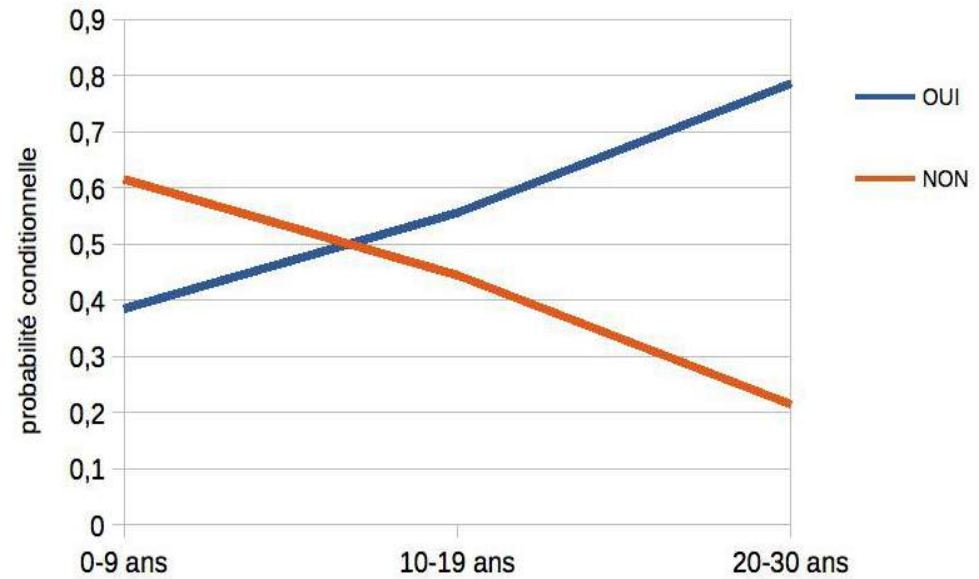
Tableau de contingence

Expérience	Indexation automatique		Total
	OUI	NON	
0-9 ans	5	8	13
10-19 ans	10	8	18
20-30 ans	11	3	14

Probabilité conditionnelle

Expérience	Indexation automatique	
	OUI	NON
0-9 ans	0,38	0,62
10-19 ans	0,56	0,44
20-30 ans	0,79	0,21

Enseignement de l'indexation automatique en fonction de l'expérience



Annexe 5 - Traitement analytique

Contingence Q2-Q6 : Relation entre l'expérience et le pré-requis de l'indexation

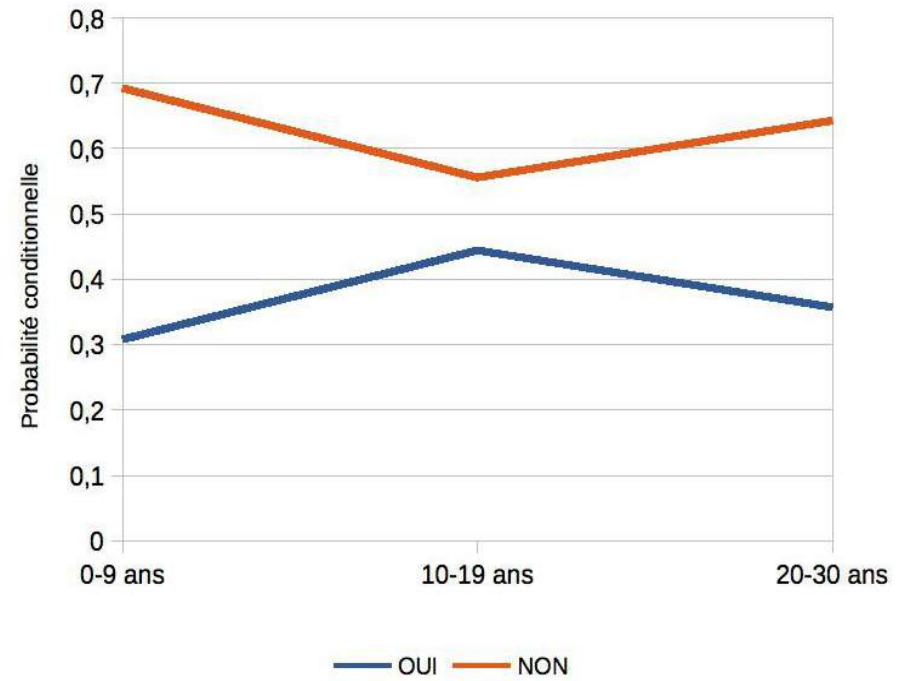
Tableau de contingence

Expérience	Indexation en amont		Total
	OUI	NON	
0-9 ans	4	9	13
10-19 ans	8	10	18
20-30 ans	5	9	14

Probabilité conditionnelle

Expérience	Indexation en amont	
	OUI	NON
0-9 ans	0,31	0,69
10-19 ans	0,44	0,56
20-30 ans	0,36	0,64

Pré-requis de l'indexation en fonction de l'expérience



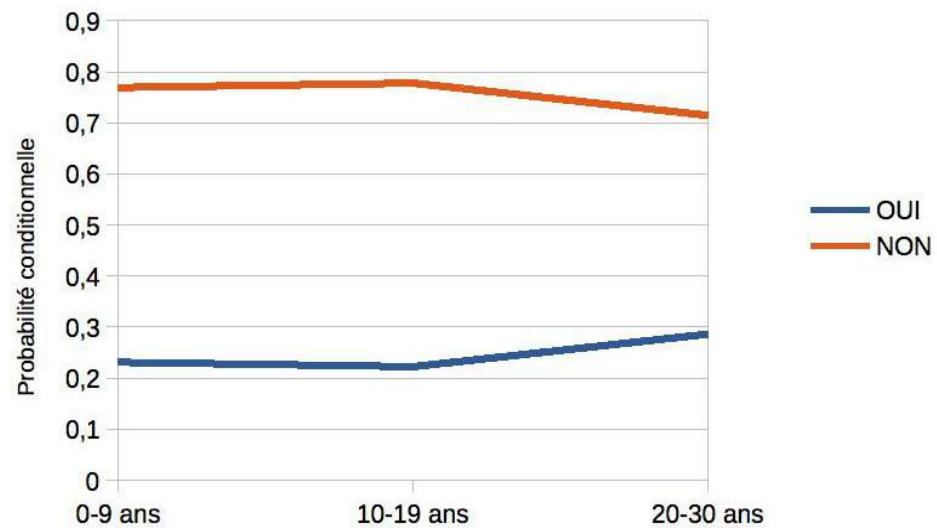
Annexe 5 - Traitement analytique

Contingence Q3b-Q6 : Relation entre l'expérience et l'enseignement des *infoboxes*

Tableau de contingence			
Expérience	Enseignement infobox		Total
	OUI	NON	
0-9 ans	3	10	13
10-19 ans	4	14	18
20-30 ans	4	10	14

Probabilité conditionnelle		
Expérience	Enseignement infobox	
	OUI	NON
0-9 ans	0,23	0,77
10-19 ans	0,22	0,78
20-30 ans	0,29	0,71

Relation entre l'expérience et l'enseignement des infoboxes



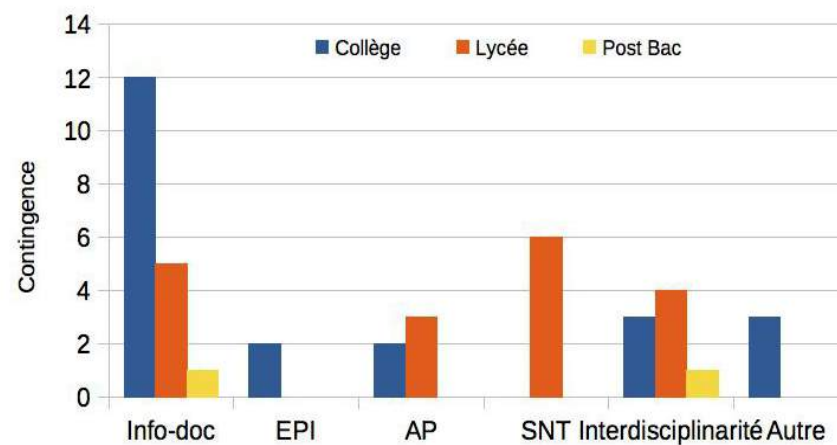
Annexe 5 - Traitement analytique

Contingence Q1b-Q1d : Relation entre le cadre pédagogique et le niveau scolaire

Tableau de contingence

Cadre pédagogique	Niveaux d'enseignement			Total
	Collège	Lycée	Post Bac	
Info-doc	12	5	1	18
EPI	2	0	0	2
AP	2	3	0	5
SNT	0	6	0	6
Interdisciplinarité	3	4	1	8
Autre	3	0	0	3

Cadre pédagogique en fonction du niveau scolaire



Annexe 5 - Traitement analytique

Contingence Q1d-Q1c : Relation entre le cadre pédagogique et les notions génériques

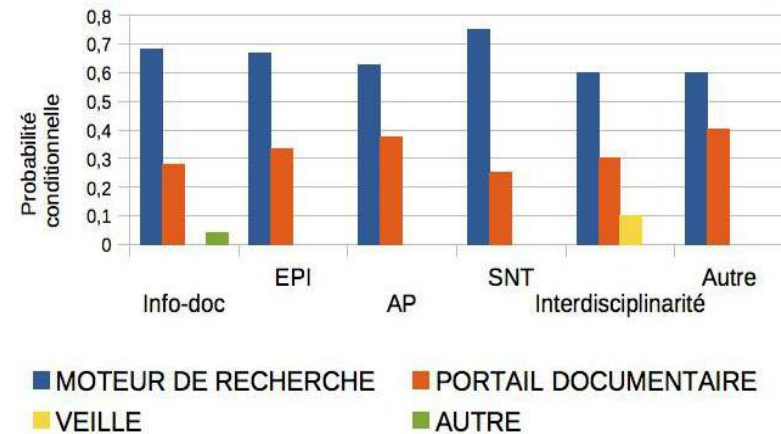
Tableau de contingence

Cadre pédagogique	Notions génériques				Total
	MOTEUR DE RECHERCHE	PORTAIL DOCUMENTAIRE	VEILLE	AUTRE	
Info-doc	17	7	0	1	25
EPI	2	1	0	0	3
AP	5	3	0	0	8
SNT	6	2	0	0	8
Interdisciplinarité	6	3	1	0	10
Autre	3	2	0	0	5

Probabilité conditionnelle

Cadre pédagogique	Notions génériques			
	MOTEUR DE RECHERCHE	PORTAIL DOCUMENTAIRE	VEILLE	AUTRE
Info-doc	0,68	0,28	0,00	0,04
EPI	0,67	0,33	0,00	0,00
AP	0,63	0,38	0,00	0,00
SNT	0,75	0,25	0,00	0,00
Interdisciplinarité	0,60	0,30	0,10	0,00
Autre	0,60	0,40	0,00	0,00

Relation entre le cadre pédagogique et la notion générique



Annexe 5 - Traitement analytique

Contingence Q1d-Q5 : Relation entre le cadre pédagogique et les méthodes pédagogiques

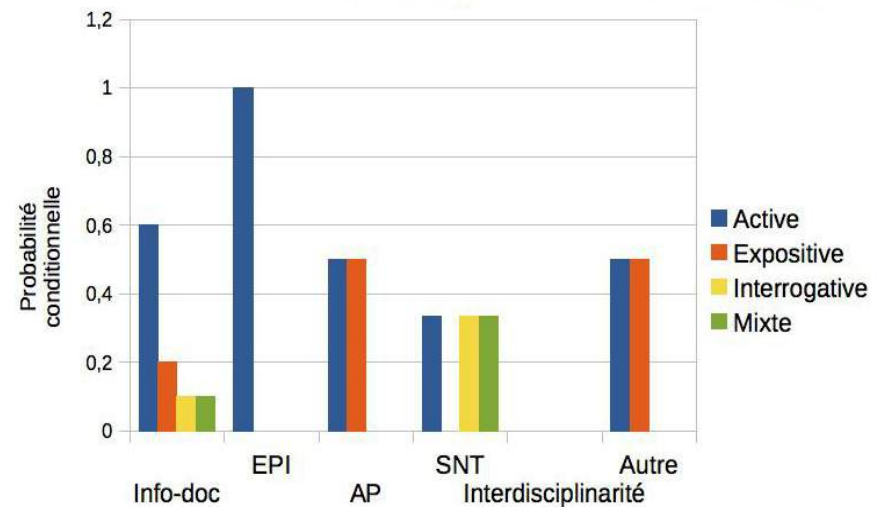
Tableau de contingence

Cadre pédagogique	Méthodes pédagogiques				Total
	Active	Expositive	Interrogative	Mixte	
Info-doc	6	2	1	1	10
EPI	1	0	0	0	1
AP	1	1	0	0	2
SNT	1	0	1	1	3
Interdisciplinarité	0	0	0	0	0
Autre	2	2	0	0	4

Probabilité conditionnelle

Cadre pédagogique	Méthodes pédagogiques			
	Active	Expositive	Interrogative	Mixte
Info-doc	0,60	0,20	0,10	0,10
EPI	1,00	0,00	0,00	0,00
AP	0,50	0,50	0,00	0,00
SNT	0,33	0,00	0,33	0,33
Interdisciplinarité	0	0	0	0
Autre	0,50	0,50	0,00	0,00

Relation entre le cadre pédagogique et les méthodes pédagogiques



4^{ème} de couverture

5 Mots clés : Indexation automatique, encart informationnel, infobox, didactique, moteur de recherche.

Résumé en Français (10 lignes):

L'indexation automatique est un procédé documentaire qui a évolué avec les nouvelles technologies du *Web* et des moteurs de recherche. Parmi ces nouveautés, les encarts informationnels, ou *infoboxes*, proposent aux internautes des condensés d'information. 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 tous observé ces *infoboxes*, il.elle.s les étudient peu avec les élèves. Notre étude propose une mise à jour de la didactisation de la notion d'indexation automatique, avec une progression originale de son enseignement à travers les *infoboxes* pour tout le secondaire.

Résumé en Anglais (10 lignes):

Automatic indexing is a documentary process that evolved with Web new technologies and web browsers. From these new technologies, infoboxes provide condensed information for web surfers. Our survey shows that the majority of school librarians teach automatic indexing. However, while most of them observed infoboxes, they rarely study them with their students. Our study yield an update of the automatic indexing didactization associated with an original teaching progression from infoboxes for every secondary degree.