

Initiatives TICE

Compte-rendu du point d'information collectif « **indexation** » du 13/02/18

Animation : Hélène

Participants : Loïc, Iris, Nathalie, Guy, Claire, Anne, Hélène, Marc, Florestan, Alain P, Corinne, Aude

Absents : Michel, Roger

Introduction :

Pas une formation mais un point d'infos

Se pose la question de l'indexation notamment pour la question des vidéos

Autant il est possible pour un logiciel d'aller chercher dans un texte, mais dans une vidéo qu'on peut assimiler à une "boite noire" (impossible de voir de quoi cela parle).

Donc essentiel de se poser la question quand beaucoup de vidéos

Définitions :

Catalogage : description de la ressource : auteur, titre, nbr pages....(correspond à la bibliographie)
objectif

Indexation : description du contenu

le fait de qualifier le contenu d'une ressource. = travail intellectuel, pas d'objectivité, en fonction du public

2 personnes n'indexent pas de la même manière...

Ex catalogage vidéo : auteur, réalisateur, durée, pal, secam

Ex indexation vidéo : qu'est-ce que je mets en avant ? qu'est-ce qui intéresse NOTRE public ? (Dordogne, élevage, posture, bio, homme, femme ?)

Donc nécessaire, obligatoire, de préciser dès le début son public !

Quels sont les objectifs d'indexation sur le web ? Aujourd'hui, on n'en a pas au niveau de SAF.

Ex Supagro : on s'adresse à la communauté éducative

Référencement : s'adapter, non plus à son public, mais à la façon dont les algorithmes des moteurs de recherche nous mettent en avant (ex : Google...) (nombre de vues, nombre de liens menant vers la vidéo)

NB : du "sans titre" et "florac2"... ben non, ça ne dit rien à personne...

Lien avec la qualité de la stratégie de comm : l'indexation n'est pas aboutie, si ce n'est pas bien indexée, autant ne pas le montrer... (lien caché, rendre invisible)

NB : vos projets ont des technologies différentes et des niveaux d'avancement différents.

Qui va indexer ?

Soit 1 personne seule pendant 1 mois à temps plein

Soit 1 équipe avec des règles

Préconisations :

Des "allers / retours" pour acquérir peu à peu une vision d'ensemble du corpus...

Commencer à indexer, quitte à revenir si besoin de normaliser par la suite : quand on a assez de recul, on revient en arrière, et on réinterroge.

Quand plusieurs personnes produisent des ressources, si les règles sont trop fermes, cela empêche la créativité, donc il faut de la souplesse.

ex : si le même concept est exprimé de plusieurs manières différentes (ex : élevage - zootechnique) : mettre en place un moteur de recherche pour rassembler le résultat de ce type de recherche

Donc

soit 1 personne, avantage : cohérence de l'indexation !

Soit plusieurs personnes

La personne la mieux placée : celle qui a décidé de la réalisation de cette séquence *soit / et* celui qui a filmé

NB : Pas d'accès séquentiel sur une vidéo : on ne peut pas avancer dedans comme dans un article avec titre, sous-titre, sauf si chapitre (en piochant dans titre, intro, termes récurrents dans le texte ...)

L'idéal : mettre dans le budget qqn dédié à l'indexation (ou prévoir temps de travail)

Sinon : réussir à partager le travail entre ceux qui ont participé (sans avoir à se retaper la vidéo)

Dernier recours : ceux qui n'ont pas vu la vidéo...

NB : si des perspectives d'audio-description, avec des sous-titres non incrustés (fichier à part), alors cela permettra de faciliter l'indexation.

Indexation de texte :

Google sait rechercher dans un texte.

Il existe des logiciels destinés aux sciences sociales peut être utiles pour des enquêtes "*text mining*" : je lui donne mon corpus, il me dit de quoi cela cause.

Quand faire l'indexation ?

Mieux vaut le faire au fil de l'eau pour diminuer la charge de travail

Timing : ça dépend du groupe, trouver le curseur entre ce que la communauté va faire et ce que l'opérateur va faire

Mais commencer léger : mots-clés dans dailymotion, et arriver à une indexation dans le site si nécessité. Boulot de copier-coller chiant

Indexer c'est pas rigolo :(

Ca peut être problématique si on a les vidéos avant la plateforme

Si je demande à des personnes de le faire, quelles consignes je peux leur donner ? dépend de leur temps dispo, de leur niveau d'implication, mais cela apporte une plus grande plus-value.

Ex : dépôt sur dailymotion au début : possibilité de mettre des mots-clés à côté, mais pas possible de les "récupérer" pour aller vers autres plateformes

A quel moment on sent qu'on passe une marche importante dans l'indexation ?

à définir à priori : les licences, l'indexation des ressources

Comment sentir l'implication du groupe : gens habitués à travailler ensemble etc, cf animation de projet.

Comment on indexe ?

Réfléchir entre ce que va faire l'animateur et les partenaires, le chef de projet et les autres participants...

Folksonomie : indexation libre = tout le monde met les mots-clés qu'il veut = a priori le plus simple, je n'ai pas à apprendre un langage, ni me conformer à une liste (principe : comme on est dans la même

communauté (ceux qui indexent et ceux qui chercheront la ressource), a priori, même langage).
Demande de repasser derrière pour uniformiser, toletter

- Remarque sur la saisie : - é ' Synonmie et orthographe, accentuation.... dépend du moteur et des possibilités de celui-ci
- Cela dépend du logiciel (ex : au CDI, si tu dis "poules", il est capable de chercher "volailles", et google de te repérer tes fautes d'orthographe)
- Yeswiki / bazar : non
- Bête et méchant : juste ce que j'ai saisi
- Certains sont capables de chercher la **troncature** par défaut ex : poule -> poules - > poulette -> pouailler
- Certains ont des capacités **sémantiques** ex : poules -> volailles
- Attention, passer d'un niveau à l'autre est un gros boulot !

NB : possibilité d'acheter la / les "**technologie(s)**" du **moteur de recherche google** et de l'intégrer sur mon outil, mais en échange, je lui donne mes données ! et il cherche selon ses algorithmes, et pas forcément en lien avec ma stratégie d'indexation

Ici, il vaut mieux avoir des nos moyens techniques, on aura ou pas les moyens pour récupérer les mots-clés, et le modifier à la marge (ex : substitution de termes proches, passage du singulier au pluriel... de manière automatique).

Donc niveaux :

NB : curseur sur les axes suivantes :

le système s'adapte aux gens

les gens s'adaptent au système

1) Folksonomie + outil informatique puissant

2) **Folksonomie + Pas d'outil informatique** qui aide à la recherche et la modification de l'indexation, donc dans ce cas, donc besoin de qqn dédié

3) **Folksonomie guidée + Communauté motivée/puissante impliquée** avec possibilité de définir des consignes et que ça ne les décourage pas : les gens sont capables de se mettre d'accord sur des règles (singulier, pluriel, accent...), ils s'impliquent

Processus itératif : qu'est-ce qui fait que ça ne marche pas ? où j'ai mis la barre trop haute ?

Soit je baisse le cran de participation

Soit j'essaie de me faire aider par des outils informatiques qui vont les aider eux (corriger automatiquement l'orthographe) : ex recherche par lot (remplacer toutes les poules par volailles)

4) Liste d'autorité

soit définie en amont

soit définie en amont et on retravaille

soit on travaille un peu et on définit par la suite

A éviter : fixer un truc qui ne bouge plus

NB : le même mot-clé pour toutes les vidéos, veut dire qu'il n'est pas bon, car pas discriminant.

5) Thésaurus

Mots-clés figés et liés entre eux par une hiérarchie, une catégorie

ex : agriculture / élevage / granivores / volailles / poules...

NB : descripteurs peuvent être rajoutés au thésaurus, pour qq temps après, rajouter de nouveaux mots-clés si ça ressort

Comment précisément ?

Indexation peut se faire par mots-clés, résumé, chiffres

En liste d'autorité :

Choisir dans une liste déroulante (maxi 30 choix dans une liste déroulante)
ou dans un nuage de mots

Attention, effet "prime au vainqueur" : celui qu'on utilise le plus se voit le mieux, et donc c'est celui qu'on va le plus utiliser.

ou taper le début d'un mot, et proposition automatique

Possibilité de proposer une **indexation "par facette"** = proposer des champs (ex : localisation géographique, public, thématique, durée...)

C'est bien pour indexer, mais pas pour celui qui cherche.

Consultation : ce que cela donne pour l'utilisateur

à quel point l'utilisateur a besoin de l'info et va faire des efforts pour la chercher

Mainstream

Google avec un seul champ où tu tapes tout ce que tu veux a changé les pratiques (la recherche par champs est obsolète)

Se retrouver dans une arbo n'est plus d'actualité

Qui a le plus de boulot cognitif ? l'outil informatique ou l'utilisateur ?

NB : google a mis en place l'évitement du silence informatique, il nous donne plein de réponses, même éloignées. Si j'ai pas de réponse, je me dis qu'il n'y a rien sur la question, et pas que je dois peut-être changer de mots-clés

Donc : indexer LARGE !

Questions diverses

Question : est-ce qu'il y a un processus d'indexation itératif et participatif, où ma communauté d'utilisateurs peut en rajouter ?

Réponse : il faut être certain que ma communauté soit super impliquée, et cela ne doit pas être implémenté dès le départ, car ça peut ne pas être utilisé.

oui, avec des #hashtags, ou des mots-clés, mais il faut une "masse" de participants

Parenthèse Comment on indexe sur un projet multi-langues ?

A priori, il faut faire le travail plusieurs fois..., par les partenaires qualifiés

Avec une capacité de traduire le mot technique dans le même contexte

également problème des droits en fonction des pays

Parenthèse Droit à l'image à l'international

cf Isabelle dans le département BE qui doit avoir des références.

cf pour la propriété intellectuelle : Convention de Berne signée par différents pays européens.

Conclusion

Pour travailler sur une définition plus précise de la stratégie d'indexation de vos ressources, prenez rdv avec le binôme Hélène + Florestan (alliant ainsi le développement web associé).

On se retrouve bientôt pour un prochain d'infos Initiatives TICE. Au programme à venir :

- **les licences (creative commons)**

- **la mise en réseau, c'est quoi ?**

- **stratégie digitale réseaux sociaux et news au sein d'un réseau**