Image not found or type unknown

Le phénomène "mot-clé": une limitation de la recherche d'informations

publié le **15/02/2012**, vu **3462 fois**, Auteur : Info Juridique

Le droit évolue constamment et les professionnels se doivent d'être au fait de l'actualité juridique. Cependant, du fait de la dématérialisation de l'information, de sa production et de sa diffusion à l'ère du numérique, la densité et l'abondance des informations sont telles qu'il est humainement impossible de veiller personnellement à l'évolution juridique ; c'est pourquoi il existe des bases de données juridiques en ligne.

Image not found or type unknown

Dès lors, et comme dans toute recherche d'information à l'ère du numérique, il est nécessaire de mettre en place une stratégie qui consiste à :

- cerner le besoin d'information ;
- formuler le besoin d'information ;
- repérer les sources pertinentes ;
- identifier les outils à exploiter en fonction de ces sources ;
- interroger les outils ;
- et évaluer la qualité des résultats.

Dans le cadre d'une recherche traditionnelle en bibliothèque, un catalogue répertorie et localise les documents afin d'orienter les usagers, que les documents recherchés se présentent sous format papier ou sous une forme numérisée.

Ce catalogue de bibliothèque donne accès aux notices bibliographiques des documents qu'il contient.

Une notice bibliographique est la description d'un document. Elle est organisée selon des champs : auteur, titre, éditeur, collection, ISBN, année de publication, type de publication, résumé, mots-sujet, etc…

Ainsi, pour permettre une recherche par le sujet, les documents doivent être indexés, c'est-à-dire associés à des mots-clés.

Traditionnellement, l'indexation des documents d'une bibliothèque est manuelle et s'appuie sur un thésaurus ; c'est-à-dire un ensemble de mots-clés ou mots-sujets prédéfinis que le bibliothécaire choisit d'associer au document pour décrire son contenu. Dès lors, le problème principal est que le vocabulaire d'un thésaurus est fermé, de sorte que tous les mots du dictionnaire ne s'y trouvent pas forcément.

Cependant, les catalogues intègrent désormais une indexation automatique, il est alors possible d'entrer une requête composée de mots-clés qui seront cherchés dans tous ces champs à la fois, ce qui lève partiellement la contrainte du vocabulaire fermé du thésaurus, et rapproche cette pratique de celle de l'interrogation des moteurs de recherche.

Le moteur de recherche est un outil qui permet de retrouver des pages web à partir d'une requête. Conséquemment, un logiciel réalise ainsi les tâches suivantes :

- moissonnage : les pages du web sont parcourues automatiquement par un robot ;

- indexation automatique : les pages sont analysées pour en extraire les mots significatifs ;

- interrogation en fonction d'une requête : les mots clés de la requête sont comparés aux mots extraits par l'indexation et une liste de pages web sélectionnées est affichée par ordre de pertinence.

Le moissonnage comporte cependant quelques limites car certaines pages du web ne peuvent pas être moissonnées par les moteurs de recherche : c'est le web invisible ou web profond. Notamment certaines pages ne sont accessibles qu'après authentification ; d'autres sont des pages dynamiques qui sont composées à partir de données stockées dans des bases de données.

L'administrateur d'un site web peut également interdire aux robots des moteurs d'indexer certaines pages.

Quant à l’indexation automatique, il est constaté que certains documents en ligne ne sont pas indexés car le moteur ne sait pas traiter leur format.

Copyright © 2025 Légavox.fr - Tous droits réservés

Par exemple, pour indexer les images, les moteurs utilisent les métadonnées associées, le texte entourant l'image ou les tags laissés par les internautes. Les textes très courts ou « mal écrits » (langage SMS) représentent alors un défi pour l'indexation automatique.

Avec les bases de données, on sort alors du livre numérique pour entrer plus spécifiquement dans l'information juridique. Il y a là une évolution considérable, très différente de la bibliothèque physique. Alors que celui qui effectue une recherche grâce à cet outil numérique pense qu'il en a la maîtrise à travers des mots-clé qu'il insère, il se trouve en réalité entièrement tributaire, pour le résultat de sa recherche, du moteur et des données qui y sont été intégrées.

Le en ligne, la base de données supprime assez largement le hasard, la lecture de parcours et la vue d'ensemble. On voit souvent imprimer quelques paragraphes isolés d'un ouvrage mis à jour, sans qu'il ait toutefois été vérifié qu'ailleurs dans ce même ouvrage, la contradiction ne serait pas portée (ou « portable ») à l'analyse ou l'opinion émise dans ces paragraphes.

Cela entraîne une dépendance et une certaine limitation de l'investigation intellectuelle et matérielle qui semble préoccupante.

Dès lors, si la bibliothèque numérique composée, d'une part, de livres structurés de manière identique aux ouvrages papiers et, d'autre part, de bases de données, il semble, d'un point de vue technologique, propre à susciter un émerveillement complet et sans réserve, qu'elle présente en revanche très probablement de graves insuffisances du point de vue intellectuel, scientifique et de recherche pratique.

L'avocat Ariel DAHAN, dans un message sur la liste Juriconnexion à propos de la formation des étudiants en droit, le disait dans des termes proches : « Mon expérience, qui ne vaut certainement que pour moi, me porte à penser que la recherche papier est infiniment plus riche que la recherche en ligne, en ce qui concerne la recherche en droit. [...] Je m'interroge sur la formation d'étudiants qui ne seraient formés qu'à la compulsion de bases de données. Leur habitude de travail serait faussée, ab initio, en raison de ce que la base de données ne donne que des informations pertinentes ou supposées l'être. Mais elle ne laisse pas l'étudiant parcourir l'ouvrage à la recherche ne serait-ce que de la bonne page ! Elle l'aiguille sur un résultat donné, qui a tendance à devenir « le » résultat universel ».

Ainsi, ce n'est pas tant l'ère du numérique que l'abondance des sources qui, de nos jours, doit être pris en compte dans les recherches.

Certes, l'ordre des réponses d'un moteur à une requête tient compte de la pertinence thématique, c'est-à-dire de l'importance des mots-clés de la requête dans la page web ; mais il tient aussi compte de la popularité de la page web, que les moteurs déduisent soit de l'abondance et de la qualité des liens qui mènent à la page, soit de la tendance des internautes à cliquer sur la page en réponse à la même requête.

Copyright © 2025 Légavox.fr - Tous droits réservés

Certains moteurs sont en effet explicitement collaboratifs : l'ordre des réponses est déterminé par les notes laissées par les internautes et l'indexation peut s'appuyer sur leurs tags.

Ministère de l'Enseignement Supérieur et de la Recherche [En ligne], Nathalie DENOS et Karine SILINI, « Organiser la recherche d'informations à l'ère du numérique » [PDF]

Réseaux du droit, Le journal du village de la justice, « Les bases de données en ligne : états des lieux » [PDF]

www.precisement.org, Un bloc pour l'information juridique [En ligne], Emmanuel BARTHE, « Débat papier contre numérique : quelques apports récents »

Copyright © 2025 Légavox.fr - Tous droits réservés