

GRILLE D'ÉVALUATION DES CONTENUS DES DOCUMENTS WEB

**Yamina Zellouf , Jean-Jacques Girardot , Philippe Jaillon ,
Roland Jégou**

*Ecole Nationale Supérieure des Mines de Saint-Etienne
Département Réseaux Informations Multimédia
158, Cours Fauriel. 42023 Saint-Etienne*

{zellouf, girardot, jaillon, jegou}@emse.fr

Résumé :

Dans la nouvelle société de l'information, le développement de l'Internet expose les individus à des abus et à des risques de toutes sortes. En particulier, une nouvelle forme de pollution s'installe chaque jour un peu plus sur le World Wide Web, elle est due à l'excès d'information, à la prolifération de données qui n'ont pas été vérifiées entraînant de multiples effets pervers ou indésirables. Il est souvent difficile pour l'utilisateur, qu'il soit professionnel ou grand public, de faire la différence entre les informations qui sont utilisables, fiables et celles qui ne le sont pas. On entrevoit les conséquences néfastes de cet état de fait si certaines règles ne sont pas établies pour limiter les effets pervers et les abus. Pour résoudre ce problème il est nécessaire de savoir comment évaluer, critiquer ou vérifier l'information, de déterminer à quel moment une information doit être ignorée, rejetée et comment elle doit être lue, classée et référencée.

Le but de notre projet est de proposer une architecture globale pour l'évaluation, la validation et la certification de l'information sur le Web qui s'appuie sur des outils et des mécanismes déjà existants [ZELL99]. Il doit permettre d'offrir aux utilisateurs de consulter des documents certifiés par des organismes indépendants. Le contenu et la fiabilité de ces documents sont basés sur l'avis d'experts choisis pour leurs domaines de compétence. Les experts qui évaluent les documents les annotent grâce à des mécanismes d'étiquetage tels que ceux décrits dans [LIPP98] [LASS] et ils sont engagés par l'avis qu'ils ont donné. Les étiquettes associées aux documents expertisés sont conservées dans une base de données. Les experts étant engagés par leurs avis, il est important de mettre en œuvre des mécanismes d'authentification, de protection et d'actualisation de ces informations.

Dans cet article nous aborderons les problèmes liés à la définition exacte de la notion qualité d'information sur l'Internet et nous présenterons les critères que nous avons sélectionnés pour évaluer la qualité des sites Web.

Introduction

L'expression "écologie" revient à l'Italien Domenico Scavetta [SUTT98] qui la définit comme l'étude du milieu de vie et des rapports entre les êtres vivants et leur environnement. C'est aussi l'étude des grands déséquilibres qui menacent la vie par disparition d'une de ces composante utile.

Cette définition est aussi appropriée à la révolution qui est en cours avec la nouvelle société de l'information et le développement des réseaux de communication tel que l'Internet. En effet, l'Internet a pris de l'essor et influence nos méthodes de travail, nos modes de communication, nos activités et même certains de nos loisirs. Cette technologie nouvelle favorise l'accès à de nouvelles informations par la facilité et la liberté d'accès qu'elle propose. Cependant cette liberté de l'information sans contrainte est porteuse de risques que l'on peut nommer "infopollution".

L'infopollution apparaît lorsqu'il y a trop d'informations, provoquant une sorte de saturation où il devient difficile de différencier les bonnes sources des mauvaises. Ces sources d'informations sont souvent hétérogènes et peuvent varier d'un site commercial, à un site universitaire, gouvernemental voire personnel. La qualité des informations y est très variable : certaines sources sont fiables et validées, recommandées par des sources de grande renommée, d'autres sont fausses, intentionnellement ou non, risquant de désorienter dangereusement les utilisateurs.

Le but de cet article est d'aborder le problème de la qualité de l'information sur l'Internet. Des critères d'évaluation sont proposés et peuvent être employés pour évaluer cette qualité.

Pollution sur Internet

La pollution informationnelle sur le Web peut prendre des formes variées et des dimensions importantes, et en conséquence elle peut entraver l'accès aux informations réellement utiles. Parmi les causes de cette pollution nous pouvons citer :

- la **surabondance** qui affecte sensiblement l'efficacité : il faut savoir comment trouver l'information cherchée parmi toutes celles qui nous sont offertes sur l'Internet.
- la **désinformation**, un deuxième risque qui affecte le savoir (contenus de mauvaises qualités, rumeurs, images électroniquement trafiquées, etc.).
- la **médiocrité** des textes accessibles. Il faut savoir de quel repère devra disposer l'utilisateur pour accorder sa confiance à un contenu précis. Des services devront être mis en place afin de permettre aux utilisateurs de distinguer le simple brouillon d'un texte élaboré, le propos d'un inconnu de celui d'un prix Nobel, le texte orienté à des fins commerciales du texte objectif d'une tierce partie. [SUTT98]
- la **prolifération** de sites indésirables avec des textes à caractère raciste, pornographique, pédophile etc.) dont l'information est à rejeter.

Afin d'éviter ce nouveau phénomène de pollution, divers organismes tel que RSAC (Recreational Software Advisory Council) ont proposé des solutions telles que la mise en place d'outils de sélection et de filtrage [LIPP98] pour éviter tous les déchets informationnels qui se

trouvent sur l'Internet. Il est clair qu'un compromis est à établir, car d'un côté procéder à la mise en œuvre de systèmes de filtrage des contenus sur Internet pourrait entraîner la censure, mais aussi une certaine forme de monoculture et d'un autre côté, l'absence de dispositifs de sélection peut aussi générer des comportements de fuite, d'abstinence ou de rejet des informations présentée aux utilisateurs.

Evaluation des Informations

Internet est un réseau de réseaux qui regorge d'informations. Tandis qu'il présente une richesse d'informations disponibles de tout type et de tout genre, il n'y a souvent aucune garantie sur sa qualité. Certaines sources offrent des informations de haute qualité, d'autres laissent à désirer. En effet, ces informations peuvent provenir d'un expert reconnu avec une excellente documentation et des références riches, d'autres peuvent représenter une rumeur ou même une propagande commerciale. Il faut rajouter à cela l'absence de politique éditoriale qui pourrait gérer ces informations. Il incombe souvent à l'utilisateur de vérifier et analyser lui-même les sources d'informations. Mais comment un simple utilisateur pourrait-il évaluer critiquer ou même vérifier l'information qu'il trouve sur l'Internet ?

Il devient donc nécessaire d'établir des critères d'évaluation par lesquels la qualité des sites Web peut être évaluée [RETT96].

Rappelons que si l'Internet offre des services tels que le courrier électronique, les listes de diffusion, les news, notre travail s'intéresse plus particulièrement aux documents Web. Un document Web représente un ensemble des pages ayant des liens entre elles et qui sont diffusées par une même source.

Plusieurs experts se sont intéressés à cette question, et plusieurs grilles d'évaluation ont été élaborées. Ces dernières sont soit générales et donc destinées à des évaluations standards valables pour tout type de site, soit spécifiques et donc applicables à des sites traitant de domaines précis tels que la pédagogie, la santé... Dans ce dernier cas, les critères d'évaluation sont souvent plus détaillés.

Il faut noter aussi l'existence de nombreuses méthodes qui définissent des critères d'évaluation qui sont sans rapport avec la qualité des contenus du documents Web et que l'on n'abordera pas dans cet article [ADAM96] [COLL96] [GRAS98] [COOK96].

Approche

L'approche employée pour développer les critères présentés dans cet article s'est inspirée au départ essentiellement des critères d'évaluation utilisés dans les centres de documentation et plus particulièrement par les bibliothécaires pour l'évaluation des livres et des articles. On peut imaginer facilement que ces techniques sont utilisables dans un domaine tel que l'Internet. Nous avons alors élaboré une grille d'analyse standard.

Une première étape consistait en l'application des critères contenus dans cette grille aux documents Web dont la structure est souvent très différente de celle que l'on peut rencontrer dans un livre. Ces documents Web ont été choisis sur le site *Agora21* qui est spécialisé dans les domaines de l'environnement et du développement durable [AGOR98].

Une première remarque des évaluateurs concernait la difficulté de répondre à certains critères présents sur la grille standard. Ceci était dû essentiellement à un manque d'informations contenues dans le site.

Une deuxième remarque était que la grille ne pouvait pas informer entièrement l'utilisateur sur tous les aspects présents dans le document. Il a été nécessaire de l'étendre à d'autres points, malheureusement très spécifiques au domaine choisi qui était celui de l'environnement.

L'idée a été alors d'essayer d'établir une grille qui soit basée sur des critères les plus communs possibles à tous les types des sites Web et donc à toutes les expertises possibles.

Description de la grille

La grille contient une première partie générale d'identification et de présentation du site en question. Une deuxième partie est celle qui se rapporte au contenu du document. Elle est représentée sous forme de plusieurs sections.

Difficultés rencontrées

Il est évident que la définition des critères n'est pas suffisante. Il est nécessaire de savoir attribuer le poids respectif à chacun des critères afin d'obtenir un score global pour chaque document. Les sites concernés ayant une vocation d'information, la pondération doit être faite de manière à privilégier l'aspect "qualité du contenu".

Aussi, il faut noter un facteur très délicat qui est la divergence et la variabilité des opinions dans l'évaluation des contenus exprimées par les différents experts. Ces derniers peuvent avoir des opinions très divergentes et fortement soutenues, ce biais étant dû principalement à des expériences personnelles et professionnelles propres à chacun d'entre eux.

Critères d'évaluation

Les critères énoncés ci-dessous s'appliquent en général à tous types d'informations. Cependant, il est à noter que les informations se trouvant sur l'Internet doivent être évaluées avec discernement à partir du moment où n'importe qui peut y publier n'importe quoi, et que ces informations ne passent plus par des filtres traditionnels que sont les éditeurs et les maisons d'éditions.

Ces critères sont classés en 7 catégories :

- validité,
- réputation de la source,
- indépendance de la source,
- exactitude et précision,
- unicité,
- composition et organisation,
- intégrité.

Validité

Il s'agit de déterminer si la source d'information est trop générale, trop technique, ou trop spécialisée. Il faut vérifier aussi la pertinence et l'utilité du contenu, car la pertinence démontre comment le contenu réel d'un site correspond à l'information qu'il est supposé fournir. La nature du document doit être clairement explicité sur la première page du document. Par exemple, supposons qu'un utilisateur désire arrêter de fumer. S'il trouve un site consacré à l'arrêt du tabagisme, mais qui discute seulement des raisons d'arrêter sans fournir d'outils pour le faire réellement, ce site ne lui sera pas d'une grande utilité [CENT97].

Un autre point à soulever est celui de la citation des références. Les références doivent être montrées clairement. Les documents reliés (liens hypertextes) doivent cibler un ensemble de lecteurs avec des caractéristiques les plus similaires. Par exemple, un site pour enfants ne doit pas présenter des liens sur d'autres documents destinés aux adultes.

A noter qu'un site est qualifié de crédible lorsque ses références le sont aussi. Par contre des liens à un site dont la validité du contenu est faible entraîne la baisse de la crédibilité du site d'origine. Beaucoup de sites Web publient d'ailleurs des liens hypertextes dans le seul but d'améliorer leur validité et leur crédibilité.

L'information est-elle valide ?
L'information apparaît-elle très demandée ?
Quelles sont les sources de données qui ont été utilisées ?
Est-ce que l'information est détaillée en profondeur (sujet et thème), est-elle abondante, pertinente ?
Est-ce que l'information répond bien au but énoncé ?
Est-ce que l'information présentée est directement dérivée d'un autre format ? (Ex: version électronique d'une copie d'un livre/journal).
Est-ce que l'information prétend ne pas avoir de tendance (alors qu'en réalité c'est le cas ?)
Est-ce que les références sont précisées ?
Existe-t-il une bibliographie ?
Est-ce que la portée de l'information est énoncée ?
L'information est-elle complète ?
Y a-t-il présentation de la ressource sous un autre format ?
Quelle est la motivation de l'auteur ou du fournisseur de l'information lorsqu'il a publié l'information?
Est-ce que l'information possède un lien vers d'autres sources ?
L'information est-elle payante ?

Réputation de la source

Les problèmes qui surviennent lorsque l'on veut distinguer les informations crédibles des informations moins crédibles ou erronées sont en général liés aux problèmes de l'évaluation exacte de la source. En effet, la source de l'information est généralement le premier critère pour

sa crédibilité et sa qualité. Comme c'est le cas lorsque le contenu présenté dans un livre ou un article scientifique n'est pas l'information originale, sa source doit être clairement indiquée. De même, le nom d'un répondant doit être visible afin de permettre à l'Internaute d'envoyer des critiques, de dénoncer des omissions ou des biais dans la source de l'information. Les questions suivantes s'intéressent donc à la source de l'information.

Est-ce que la source est attribuable à un auteur ou à une organisation bien connue ?
L'origine du document est-elle mentionnée ?
Est-ce qu'il s'agit d'une source souvent citée ?
Est-ce que le nom du responsable du site est visible ?
Est-ce que l'information est attribuée nominativement à un auteur ou un éditeur ?
Avez-vous déjà entendu parler de cet auteur ?
S'agit-il d'une URL d'université ?
Est-ce que l'auteur est cité souvent par des sources respectables ?
Connaissez-vous l'expérience qu'a l'auteur dans ce domaine ?
Quel poste occupe-t-il ?
Est-ce que le nom de l'auteur est au complet ?
Est-ce que d'autres informations sur l'auteur sont précisées ?
Est-ce que l'information a été filtrée ?
Le site est-il sponsorisé par une compagnie, une organisation ou bien un expert dans le domaine ?
Le site a-t-il été recommandé ?
A travers quelle source avez-vous atterri sur cette page ?
Une autorité connue fait-elle référence à ce document ?
Est-ce que les adresses email et les coordonnées de l'éditeur, de l'auteur, et des sources sont fournies ?
Y a-t-il des adresses email pour contacter l'éditeur/le médiateur/ou les sponsors ?

Remarques:

Il faut noter qu'il est parfois difficile de juger de la crédibilité d'une source même si elle révèle correctement ses qualifications et ses références.

L'expert doit faire attention aux ressemblances des noms des sites

L'anonymat est un problème fréquemment rencontré sur Internet

Il est souvent difficile de vérifier la qualité des sites qui sont référencés par d'autres. L'expert peut être confronté à une situation dans laquelle un site respectable peut contenir un lien sur un site qui l'est beaucoup moins.

Indépendance de la source

Il faut vérifier que l'information est bien indépendante et surtout que les idées de l'auteur du site ne sont pas biaisées par des idées politiques ou autre. En effet un site peut être créé dans un but promotionnel, informationnel, commercial, politique, voire même pour toutes ces raisons. Il est important de déterminer les intentions des auteurs pour discerner les biais possibles. En général un auteur n'est jamais totalement neutre et il vaut mieux connaître son point de vue.

Enfin, s'il existe une source de financement public ou privé, il doit y être fait référence dans le document.

L'information est-elle autonome ?
La ressource contient-elle plus d'une adresse de contact ?
L'information est-elle biaisée ?
Est-ce simplement de l'information publicitaire ?
Est-ce que l'information est payante ?
S'agit-il d'une liste de liens ? Si oui ces liens sont-ils annotés ?

Exactitude et précision

L'exactitude de l'information est le critère le plus évident pour la qualité de l'information. Les informations citées par le site doivent pouvoir être vérifiées par l'expert et l'utilisateur en général. Par exemple, la preuve scientifique destinée à soutenir une théorie doit être présentée clairement, le cadre d'étude doit être décrit d'une manière compréhensible par tous. Aussi, un site bien documenté et la présence de multiples références indiquent généralement que le créateur du site maîtrise son sujet. Une autre méthode pour vérifier l'exactitude de l'information est de la retrouver dans plusieurs sources différentes. Pour que l'information contenue dans le document soit fiable il faut tout d'abord qu'elle soit complète. Elle devra donc inclure aussi des sources de bonne qualité.

Parmi les questions posées, nous citons :

L'information est-elle exacte ?
Est-il possible de vérifier son exactitude ?
Est-ce que la page cite des références pour confirmer l'exactitude de l'information ?
Est-ce que l'information est détaillée ?
Jusqu'à quel niveau ?
Quel est le degré de superficialité ou de complétude de l'information ?
Existe-t-il une partie incomplète ?
Le titre est-il informatif ? Y a-t-il un résumé ?
Quelle est l'étendue de l'information ?
Est-ce que tous les aspects du sujet sont abordés ?
Y a-t-il des trous logiques dans l'information ?
Est-ce que l'index et le sommaire sont clairs ?
Y a-t-il des mots clés qui indiquent le contenu de l'information ?

Unicité

L'information peut être unique comme des résultats scientifiques publiés dans un article, ou ne pas l'être comme par exemple s'il s'agit d'un fait rapporté par plusieurs journaux.

L'information est-elle unique ?

L'information est-elle importante ?
Y a-t-il un travail original sur le site ?
L'information est-t-elle en relation avec d'autres travaux ?

Composition et organisation

La qualité de l'organisation de l'information est primordiale car elle influence notre capacité de compréhension. Un site où l'information est bien organisée sera plus facile à consulter. Après l'organisation il y a la qualité du texte ; il faut savoir si l'auteur présente clairement ses idées. Y a-t-il une répétition dans son argumentation ? La qualité de la langue est également un indicateur pour évaluer une source d'information. Le contenu doit être exact, actuel et plausible, mais aussi correctement formulé.

Est-ce que l'information est stable ?
Est-ce que le texte vérifie les règles élémentaires de la grammaire, de l'orthographe ?
Est-ce que le texte contient du jargon ?
Est-ce que l'information est ambiguë ?
Est-ce que l'information est clairement organisée ?
A-t-elle une bonne structure ?
L'information est-elle arrangée d'une manière logique ?
Est-ce qu'elle est divisée en parties logiques ?
L'information est-elle organisée en fonction des besoins de l'utilisateur ?
Le contenu est-il clairement décrit ?
Est-ce que les parties en tête sont claires et descriptives ?

Intégrité

Un nombre croissant de sites est créé chaque jour, mais nombre d'entre eux sont rapidement laissés à l'abandon. L'information qui y est présentée risque d'être périmée et les liens hypertextes font souvent référence à des sources qui ont disparues. Il est donc certaines fois important que la date de création de l'information et celle de sa dernière mise à jour soient mentionnées. Ces dates peuvent être de bonnes indications sur la pertinence et la fiabilité de l'information.

L'information est-elle actualisée et mise à jour ?
Si le site contient des données sensibles, comment sont actualisées ces dernières ?
Est-ce que la date de création du site est mentionnée ?
L'information est-elle durable en nature ? Quelle est la relation avec la fréquence de mise à jour ?
S'il s'agit d'une information statique, sera-t-elle durable à ses utilisateurs ?
L'information a-t-elle une période limitée d'utilisation ?
Y a-t-il une maintenance adéquate du contenu de l'information ?
L'information a-t-elle été mise à jour récemment ?
Y a-t-il une indication sur la fréquence des mises à jour de l'information ?

Conclusion

Il est très important de connaître la validité et la fiabilité d'une information avant de l'utiliser, de la citer ou de la diffuser. L'absence de contrainte éditoriale sur le réseau Internet et la possibilité pour chacun d'y publier des informations dont il est le seul garant peut conduire à des risques quant à leur valeur. Il est donc judicieux de mettre en place un service qui jouerait le rôle d'un médiateur d'informations et orienterait les utilisateurs vers des sources jugées pertinentes. Sa finalité serait de faire évaluer les ressources disponibles sur Internet par des experts. Ceci résume le but de notre projet en cours de réalisation.

Dans cet article nous avons tenté d'énumérer les points essentiels dans la détermination de la qualité d'un document. Il faut noter que la première difficulté rencontrée est celle d'établir des critères généraux qui soient applicables à tous les types de sites, et donc de parvenir à une seule grille utilisable dans tous les domaines, car certains critères peuvent être présents, voire essentiels dans un domaine et complètement absents dans un autre. La deuxième difficulté réside dans le fait que le choix des notes attribuées à chacun des critères peut varier selon l'usage que l'on veut faire du et donc de la requête que l'utilisateur aura formulé.

Bibliographie

- [ADAM96] ADAM site, « Art, Design, Architecture and Media Information Gateway », <http://adam.ac.uk/adam/>. Novembre 1996.
- [AGOR98] Agora21, « Environment site ». Ecole Nationale Supérieure des mines de Saint-Etienne. <http://www.agora21.org/>. 1998-2000.
- [CENT97] Association des Centraliens, « Critères de qualité de l'information de santé sur Internet . Démarche de Netscoring ».. Ecole Centrale de Paris. Novembre 1997.
- [COLL96] Le collège Virtuel, « Critères d'Evaluation des sources d'informations d'Internet ». Collège Bois de Boulogne 1996. <http://www.virtuel.collegebdeb.qc.ca/critere.html>
- [COOK96] Cook, A., McNab, A., « The good the bad and the ugly » : Internet Review sites ; <http://omni.ac.uk/agec/iolim96.html>
- [GRAS98] Grassian, E., «Thinking Critically about World Wide Web Ressources ». UCLA College Library Instruction.1998.
- [LIPP98] Lipp, P., Chu, Y-H., Deautels, P., «PICS Signed Label (Dsig) 1.0 Spécification », W3C Recommandation. May 1998.
- [LASS] Lassila, O., Swick, R., «Ressource Description Framework ». Model and syntax Specification W3C.<http://www.w3.org/TR/WD-rdf-syntax-971002.html>.
- [RETT96] Rettig, J., «Beyond cool : Analog Models for Reviewing Digital Ressources » – Paris. France. Septembre 1996.
- [SUTT98] Sutter, E., «Pour une écologie de l'information ». Documentaliste-Sciences de l'Information, Volume 35, Numero2, Mars-Avril 1998, p83, 86.
- [ZELL99] Zellouf, Y., Jaillon, Ph., Girardot, J.J., «Providing Rated Documents on the Net ». WebNet99, Honolulu, USA, Octobre 1999.