

Synthèse bibliographique

Collecte de documents et mise en forme

Alexandre Pauchet

INSA Rouen - Département ASI

BO.B.RC.18, pauchet@insa-rouen.fr

Plan

- 1 Démarche méthodologique
- 2 Ressources d'indexation
- 3 Types de ressources
- 4 Identification des sources
- 5 Évaluation
- 6 Évaluation de l'information sur Internet
- 7 Synthèse bibliographique
- 8 Références

Démarche méthodologique

Précisez le sujet

- Spécifiez le sujet et le besoin
- Délimitez la discipline
- Identifiez les bons mots clefs

Recherchez des ressources

- Multipliez et adaptez les supports de recherche
- Classez, organisez et évaluez chaque ressource (indexation)
- Résumez chaque ressource

Synthétisez

- Introduisez clairement votre problématique
- Représentez vos classes de ressources suivant un(des) axe(s) spécifique(s)
- Présentez les limites actuelles ainsi que les problèmes non résolus
- Formatez proprement vos références bibliographiques (bibtex, zotero, ...)

Exemple de délimitation du sujet

Sujet : “*Évaluation de l’information sur Internet*”

- Discipline : informatique
- Besoin : présenter à l'utilisateur des pages Web au contenu **fiable** et adapté à son **besoin en information**
 - ⇒ restriction au contenu textuel des pages Web
 - ⇒ évaluation de l'adéquation des informations à un besoin
 - ⇒ évaluation de qualité intrinsèque des informations
- Sous-disciplines : text/web mining, recherche d'information
- Mots-clefs : RI, web intelligence, internet, www, *etc.*

Ressources d'indexation

- Bibliothèque (livre, revues, *etc.*)
- Annuaire (Yahoo!, Nomade, *etc.*)
- Moteurs de recherche génériques (Google, Bing, Yahoo!, *etc.*)
- Moteurs de recherche spécialisés :
 - Google scholar : <http://scholar.google.fr/>
 - Scirus : <http://www.scirus.com/>
 - Athenus : <http://www.athenus.com/>
 - DBLP : <http://www.dblp.org/search/>
- Archives et dépôts :
 - CiteSeerX : <http://citeseerx.ist.psu.edu/>
 - Sciencedirect : <http://www.sciencedirect.com/>
 - HAL : <http://hal.archives-ouvertes.fr/>

Types de ressources

- Format

- Papier : livre, revues, actes, *etc.*
- Électronique : idem (livre, revues, actes, *etc.*)
À consulter avec méfiance : blogs, forum, pages personnels

- Type

- Livre, chapitre d'ouvrage
- Revues, journal et articles de revue ou de journal (parfois uniquement en ligne)
- Actes de conférence, posters
- Manuscrit de thèse, de stage de master 2
- Rapports techniques ou internes

⇒ **Nécessité d'être critique sur la source !**

Identification des sources

Recherche "historique" et géographique

- Ressources "de référence"
 - Articles de type *Survey* (état de l'art)
 - États de l'art d'articles

- Identifier les communautés de référence
 - Chercheur(s)
 - Laboratoires
 - Groupes de travail, projets (ANR, européen), GRR, *etc.*
 - Entreprises
 - Conférences/Journaux

Recherche bibliographique

Évaluation de la source et de la ressource

- Qualité d'une ressource
 - Type de publication : processus d'évaluation, taux de sélection
 - Reconnaissance : nombre de citations
- Qualité du support de publication
 - Classements des conférences et journaux (<http://core.edu.au/>, <http://www.scimagojr.com/>) et référencement (DBLP : <http://www.dblp.org/db/>)
 - Facteur d'impact des journaux
 - Comité de programme, procédure de sélection et taux de sélection pour les conférences
- Confiance dans les auteurs
 - Reconnaissance des auteurs : h-index, publications (DBLP, GoogleScholar)
 - Évaluation des laboratoires de recherche : classement AERES

Évaluation de l'information sur Internet

- Source identifiée et compétente
 - Auteur(s) identifié(s)
 - Auteur(s) compétent(s) (sur le sujet)
 - Nature du site appropriée au besoin
 - Propriétaire et financeur du site connus et sans conflit d'intérêts
- Information vérifiée
 - Raisonnement cohérent et argumenté
 - Sources citées et vérifiables
 - Comité éditorial
 - Peu (pas) de fautes d'orthographe et de grammaire
- Information « informative »
 - Ton modéré, raisonnements objectifs
 - Intention d'informer (et non de convaincre)
 - Cible adaptée au discours
 - Publicités séparées du contenu

Évaluation de l'information sur Internet

- Information actualisée
 - Dates du cycle de vie du document sont connues
 - Site est maintenu et mis à jour
 - Liens hypertextes fonctionnels
- Information complète et pertinente
 - Tous les aspects du sujet sont couverts par le texte
 - Le contenu correspond au besoin
- Information issue d'une source populaire
 - Références entrantes (liens vers)
 - Nombre d'utilisateurs
 - Référencement (annuaires et moteurs de recherche)
 - Critiques positives

Évaluation de l'information sur Internet

- Ergonomie
 - Moteur de recherche interne, fil conducteur, plan
 - Standards d'accessibilité et d'ergonomie
 - Liens sortants clairement identifiés

- Présentation “professionnelle” et interactive
 - Charte graphique
 - Design et couleurs lisibles
 - Contact possible (téléphone, adresse et email)

- Accréditation externe

Exemple de ressources identifiées

Sujet : “Évaluation de l'information sur Internet”

- Identification de conférences (ex : SIGIR, WWW, WI, ...), de journaux (ex : WIAS)
- Identification de chercheurs de référence (ex : Baeza-Yates, Page, Brin, Gandon, etc.)
- Identification de laboratoires (ex : équipe WIMMICS à l'INRIA) et d'entreprises (ex : Google)
- Livre/survey de référence : M. Rami Ghorab, Dong Zhou, Alexander O'Connor, Vincent Wade : *Personalised Information Retrieval : survey and classification*. International Journal of User Modeling User-Adaptated Interaction (UMUAI), 23(4), pp. 381-443, 2013.

Synthèse bibliographique

- Sélection de documents : titre, auteur, résumé
- Classement grossier et résumé long de chaque document
- Choix de(s) axe(s) de présentation et reclassement
- Présentation critique de l'existant
- Tableau synthétique
- Conclusion et introduction à rédiger en dernier
- Outils de génération automatique de bibliographie (bibtex, zotero)

Exemple de plan

2	Évaluation de l'information	13
2.1	Procédures d'évaluation de l'information	15
2.1.1	Cotation de l'information issue de sources ouvertes	15
2.1.2	Évaluation de l'information sur Internet dans un contexte civil	23
2.1.3	Discussion	31
2.2	Partage de métadonnées pour l'aide à l'évaluation	33
2.2.1	Annotation des ressources sur Internet	33
2.2.2	Exploitation d'annotations issues de tiers de confiance	35
2.2.3	Exploitation d'annotations générées par les utilisateurs	37
2.2.4	Discussion	39
2.3	Systèmes automatisés d'évaluation de l'information	40
2.3.1	Évaluation de la popularité	40
2.3.2	Évaluation de la qualité des documents	44
2.3.3	Évaluation des informations dans le cadre du renseignement	48
2.3.4	Discussion	52
2.4	Proposition	53

Exemple de bibliographie

- [MPGC10] Yann MOMBRUN, Alexandre PAUCHET, Bruno GRILHÈRES et Stéphane CANU : Collecte, analyse et évaluation d'informations en sources ouvertes. *In COTA@IC'10 : Journées francophones d'Ingénierie des Connaissances, Cotation des informations : Théorie et Applications*, juin 2010.
- [MPS08] Christopher D. MANNING, Raghavan PRABHAKAR et Hinrich SCHÜTZE : *An introduction to information retrieval*. Cambridge University Press, juillet 2008.
- [MSSV09] Justin MA, Lawrence K. SAUL, Stefan SAVAGE et Geoffrey M. VOELKER : Identifying suspicious urls : an application of large-scale online learning. *In ICML'09 : Proceedings of the 26th Annual International Conference on Machine Learning*, page 86, juin 2009.
- [Nat06] NATIONAL LIBRARY OF MEDICINE : Evaluating Internet health information, mars 2006. Available from : <http://www.nlm.nih.gov/medlineplus/webeval/webeval.html>.

Conclusion

Références

Liens

- <http://aeris.11vm-serv.net/cours/chercher/methodes.html>
- <http://tecfa.unige.ch/perso/lombardf/calvin/methodologie/recherche/recherche.html>