

Domaine D4

Organiser la recherche d'informations à l'ère du numérique

Version 1.0 ++

Nathalie DENOS

Maître de conférences en informatique
UPMF - Grenoble 2

Karine SILINI

Maître de conférences en informatique
ULCO

Licence Creative Commons :
Patrimoine - Pas d'Utilisation Commerciale
Partage des Conditions Initiales à l'Identique

D4.1 : Rechercher de l'information avec une démarche adaptée

La recherche d'information

La **recherche d'information** nécessite la mise en place d'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 ;
- **évaluer** la qualité des résultats.

Exemple :

On désire étudier l'évolution des supports de stockage numérique en vue de rédiger un article de vulgarisation de 5 à 10 pages.

Pour formuler ce besoin sur un domaine que l'on connaît mal, on commence la recherche par des sources de type lexiques, dictionnaires, encyclopédies : Wikipédia, Encyclopedia Universalis en ligne, FranceTerme (terminologie officielle en France). On utilise les moteurs de recherche dédiés à chacune de ces sources, et on obtient des résultats de bonne qualité (articles de fond, définitions) que l'on filtre selon leur capacité à répondre aux questions posées et à couvrir l'évolution des supports de stockage numérique au fil du temps. Cette première phase permet de cerner ce sujet qu'on connaît mal, en repérant de nouveaux mots-clé (loi de Moore, technique holographique, mémoire flash, etc.) et des aspects de la question à traiter auxquels on n'aurait pas pensé (ordres de grandeur, fonctionnalités, caractéristiques de durée de vie ou de coût, solutions de stockage interne, externe, partagée, distribuée, par un prestataire, etc.).

On se tourne alors vers de nouveaux types de sources compte tenu du public visé (vulgarisation) et du domaine (numérique) : les manuels fondamentaux et les ressources pédagogiques (commentçamarche, dépôts de ressources pédagogiques pour le C2i niveau 1). On y trouve des fiches pratiques pédagogiques, des échanges dans des forums, et on mobilise sa capacité à évaluer les résultats trouvés : sont-ils utiles ? sont-ils fiables ?

En rédigeant l'article, on sera certainement amené à prolonger la recherche pour documenter des points précis ; on pourra aussi se tourner vers des sources d'actualités pour donner du relief à l'introduction ou à la conclusion, ou repérer les dernières évolutions des supports de stockage (technique holographique).

Le web et les sources d'information

Il ne faut pas croire qu'un moteur de recherche généraliste interroge une unique source et donne accès à toutes les informations du web :

- les sources d'information sont multiples et variées : elles n'ont pas toutes la même valeur ;

Pour la requête « compétences numériques », un moteur de recherche sur le web peut retourner côte à côte un article du Monde, une définition de Wikipédia, un billet du blog d'un internaute influent sur ce sujet, un rapport de l'UNESCO ou une question d'un étudiant dans un forum.

- certaines informations ne sont pas accessibles par un moteur généraliste : la majeure partie des ressources en ligne font partie du **web invisible**.

Le « web invisible » ou « web profond » désigne la partie du web que les moteurs de recherche classiques ne peuvent pas indexer. Il recouvre notamment les banques de données et les ressources dont l'accès requiert une authentification.

En savoir plus...

Cerise Conseils aux Étudiants pour une Recherche d'Information Spécialisée Efficace (URFIST)

Les sources d'information

Parmi les sources d'information sur le web, on distingue :

- **les sources d'information traditionnelles** (agences de presse, grandes organisations d'intérêt privé ou public, associations d'experts, publications scientifiques, etc.) qui diffusent des informations validées selon un processus reconnu ;

L'agence de presse AFP, l'organisation internationale OCDE, la Bibliothèque Nationale de France, les éditeurs de publications scientifiques sont des sources d'information traditionnelles.

- **les sources émergent de structures collectives plus ou moins formelles**, dont les règles de publication et de validation de l'information sont plus ou moins strictes.

L'encyclopédie en ligne Wikipédia est une source qui émerge d'une communauté d'auteurs ; la qualité de l'information qu'elle diffuse est liée aux règles et processus d'édition qui lui sont propres.

- **les sources informelles** (pages web personnelles, blogs d'internautes, etc.) publiées sans aucun contrôle.

Le blog d'un prix Nobel d'économie relève des sources informelles. Les informations qui y sont publiées peuvent être d'une grande valeur, sans toutefois bénéficier d'une validation par d'autres spécialistes du domaine.

Les différents types de source

Les sources d'information peuvent être **généralistes** ou **spécialisées**. On peut les organiser ainsi :

- manuels fondamentaux (écrits par les auteurs fondateurs d'un domaine) et ressources pédagogiques ;
- encyclopédies, dictionnaires et lexiques (édités par les éditeurs importants d'un domaine) ;
- revues scientifiques (souvent éditées par des sociétés savantes, et dont les articles sont validés par des experts) et littérature grise (rapports d'études ou de recherches, actes de congrès, thèses, brevets, etc.) ;
- vulgarisation scientifique (souvent sous la forme de revues ou d'ouvrages destinés au grand public) ;
- actualité (presse quotidienne tournée vers le grand public ou un lectorat plus spécialisé selon le cas) ;
- rapports et bilans (produits par des organisations et institutions reconnues au niveau national ou international) ;
- sources officielles (produites par les gouvernements ou organismes gouvernementaux) ;
- banque de données et de statistiques (produites par des groupements de chercheurs ou des organismes officiels) ;
- blogs et podcasts (édités par des internautes relevant ou non d'institutions bien identifiées).

Lorsqu'on se documente sur un domaine que l'on connaît mal, il faut commencer par repérer les sources pertinentes qui permettront de cerner les concepts et la terminologie associés : dictionnaires, lexiques, encyclopédies, manuels, etc.

La qualité des sources

Pour évaluer la qualité d'une source, on pourra s'intéresser à sa **fiabilité** (propension à diffuser des informations exactes, liée au processus de validation de l'information) et à son **autorité** (réputation dont elle jouit dans son domaine auprès des experts).

Les catalogues de bibliothèque

Les bibliothèques conservent des documents. Pour les répertorier et les localiser, ils sont décrits par leur notice bibliographique.

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.

Un **catalogue de bibliothèque** donne accès aux notices bibliographiques des documents qu'elle contient.

Il peut arriver que le document lui-même soit accessible sous une forme numérisée à partir du catalogue.

L'indexation du contenu

Pour permettre une recherche par le sujet, les documents doivent être **indexés**, c'est-à-dire associés à des **mots-clé**.

Traditionnellement, l'indexation des documents d'une bibliothèque est manuelle et s'appuie sur un **thésaurus** ; c'est un ensemble de mots-clé ou mots-sujet prédéfinis que le bibliothécaire choisit d'associer au document pour décrire son contenu. Le vocabulaire d'un thésaurus est fermé : tous les mots du dictionnaire ne s'y trouvent pas !

Lorsqu'un usager formule une requête avec des mots-clé dans le champ correspondant à l'indexation manuelle du catalogue, il peut ne rien trouver simplement parce qu'il n'a pas utilisé les mots du thésaurus.

Les catalogues intègrent désormais une **indexation automatique** des divers champs des notices, notamment du champ « Titre » et du champ « Résumé ». Il est ainsi 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 thésaurus reste un outil intéressant pour le repérage des termes importants pour une recherche.

Le catalogue du **SUDOC** (Système Universitaire de Documentation) est le catalogue collectif français réalisé par les bibliothèques et centres de documentation de l'enseignement supérieur et de la recherche.

En savoir plus...

[SUDOC](#) catalogue du Système Universitaire de Documentation

Les portails documentaires

Un **portail documentaire de bibliothèque** centralise l'accès à un ensemble de services liés à la recherche d'information.

On trouve notamment :

- l'accès au catalogue de cette bibliothèque et d'autres bibliothèques ;
- la recherche fédérée, qui consiste à interroger simultanément plusieurs outils (catalogues de bibliothèque, sites web, moteurs de recherche, etc.) ;
- l'accès à des documents en texte intégral ;
- l'accès à des encyclopédies, des dictionnaires, la presse en ligne, etc. ;
- des services personnalisés comme l'accès à son dossier de lecteur, à des informations ciblées, à des outils de veille.

Les annuaires de recherche

Un **annuaire de recherche** est un outil de recherche qui répertorie des sites web sélectionnés pour leur qualité par des spécialistes du domaine. Les sites sont organisés dans une hiérarchie de **catégories**.

Le processus d'alimentation d'un annuaire reposant sur une **sélection humaine**, les ressources qu'on y trouve sont théoriquement de bonne qualité. Cependant, elles sont relativement peu nombreuses, et n'intègrent pas toujours les dernières nouveautés apparues sur le web.

On distingue :

- les **annuaires généralistes** : ils sont difficiles à tenir à jour ;

Exemples d'annuaires généralistes mondiaux :

[Open Directory Project – DMOZ](#) (mise à jour par les internautes)

[ipl2 - information you can trust](#) (sélection du web par des professionnels de l'information)

Ces annuaires généralistes sont peu utilisés car peu connus.

- les **annuaires spécialisés** : très utilisés, ils font l'objet d'une activité de mise à jour régulière.

Exemples d'annuaires spécialisés :

[Signets de la BNF](#) (sélection des bibliothécaires de la Bibliothèque Nationale de France)

[Signets de CERIMES](#) (sélection des bibliothèques universitaires)

[WWW Virtual Library](#) (le premier de l'histoire du web)

[Infomine](#) (ressources académiques)

On peut utiliser un annuaire selon **deux modalités** :

- en naviguant de catégorie en catégorie ;
- en faisant une recherche par mots-clé.

L'indexation est limitée aux champs que l'annuaire renseigne pour chaque site web : elle ne s'applique pas au contenu intégral des sites web.

Les moteurs de recherche

Un **moteur de recherche** est un outil permettant de retrouver des pages web à partir d'une requête.

C'est un logiciel qui réalise 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.

Google, Yahoo Search ou Bing sont des moteurs de recherche généralistes.

Il existe des **moteurs de recherche spécialisés** dans certains types de ressources (images, vidéos, publications académiques, blogs, flux RSS, personnes, etc.) ou dans certains domaines (technologie, musique, mathématiques, santé, chimie, etc.).

Sur certains sites web, on trouve un moteur dédié aux pages du site lui-même.

Les limites du moissonnage

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.

Il existe des forums de discussion privés (soumis à authentification) dont les posts ne sont pas répertoriés par le moteur de recherche.

Si vous consultez le site web de la SNCF pour faire une réservation en ligne, la page web est dynamique : ses informations changent en fonction du remplissage du train.

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

Les limites de l'indexation automatique

Certains documents en ligne ne sont pas indexés car le moteur ne sait pas traiter leur format.

Les documents au format PDF sont désormais indexés par la plupart des moteurs.

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 un défi pour l'indexation automatique.

Un **méta-moteur** est un outil permettant de soumettre une même requête à plusieurs moteurs de recherche simultanément.

Exemples de méta-moteurs : Dogpile, Yippy

Les requêtes

Lorsqu'on interroge un catalogue de bibliothèque ou un moteur de recherche, on utilise une requête pour préciser les critères de la recherche.

Catalogue de bibliothèque

Lorsqu'on fait une recherche dans un catalogue de bibliothèque, la **requête** exprime des **critères** qui sont structurés selon les champs des notices bibliographiques. On peut choisir les opérateurs pour combiner ces critères :

- l'**opérateur ET** (celui proposé par défaut) rend tous les critères obligatoires ;

Si on utilise à mauvais escient cet opérateur, cela peut éliminer des réponses pertinentes des résultats : c'est le phénomène de « silence »

- avec l'**opérateur OU** il suffit que l'un des critères soit vérifié ;

Si on change un ET en un OU dans une requête, cela augmente le nombre de résultats, au risque de voir apparaître des documents non pertinents : c'est le phénomène de « bruit ».

- l'**opérateur SAUF** permet de signaler un mot-clé non souhaité.

Dans certains cas, on utilise la version anglaise des opérateurs : AND, OR et NOT.

Au-delà de ces opérateurs booléens, il existe l'opérateur de **troncature *** qui permet de ne donner que le début d'un mot-clé pour inclure virtuellement toutes les terminaisons possibles.

*Le critère **certificat*** est vérifié si on a **certificat, certificats, certification, etc.***

Moteur de recherche

Les moteurs de recherche tendent à limiter l'effort de formulation de l'utilisateur en décidant automatiquement de la façon de combiner et ajuster les mots-clés saisis. Cependant, il est possible de réaliser une recherche plus fine en utilisant les opérateurs booléens pour combiner les critères. Les moteurs proposent généralement des notations plus légères :

- le signe + devant un mot-clé signifie qu'il est obligatoire (proche du ET) ;
- le signe – signifie qu'il ne doit pas apparaître (proche du SAUF) ;

Au-delà de ces opérateurs booléens, on utilise souvent les **guillemets** pour rechercher une expression exacte (l'expression entière doit être présente, et sous cette forme exactement).

Les moteurs offrent souvent un mode de recherche avancée où on peut préciser par exemple la langue des réponses, les sites auxquels doivent se limiter les réponses, etc.

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-clé de la requête dans la page web. Il tient aussi compte de la **popularité** de la page web, que les moteurs déduisent :

- de l'abondance et de la qualité des liens qui mènent à la page ;
- de la tendance des internautes à cliquer sur la page en réponse à la même requête.

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

La compétence en questions

Q 1 : Que contiennent les catalogues de bibliothèque ?

- 1 - Des adresses de bibliothèques.
- 2 - Des notices bibliographiques.
- 3 - Des livres à acheter.
- 4 - Des articles de journaux.

Q 2 : Qui parcourt et indexe les ressources du web dans un moteur de recherche généraliste ?

- 1 - Des internautes.
- 2 - Des experts.
- 3 - Des logiciels robots.
- 4 - Des bibliothécaires.

Q 3 : Quelle est la requête la plus précise pour rechercher les ressources contenant l'expression poisson rouge au singulier ou au pluriel ?

- 1 - poisson rouge
- 2 - poisson OR rouge
- 3 - "poisson rouge" OR "poissons rouges"
- 4 - poisson* rouge*

La compétence en situations

Trouver les bonnes sources...

Repérez les sources utiles pour élaborer votre projet professionnel.

Dressez-en une liste en commentant leur utilité et leur qualité.

Comment vole un avion ?

Dans le catalogue du SUDOC, formulez une requête qui permet de trouver des documents pour préparer un exposé sur les avions.

Annuaire ou moteur ?

Reprenez cette recherche en utilisant d'abord un annuaire de recherche puis un moteur de recherche.

Quelle approche a été la plus rapide pour repérer des informations intéressantes ?

Avez-vous trouvé les mêmes réponses avec l'annuaire et sur la première page de réponses du moteur ?

D4.2 : Évaluer les résultats d'une recherche

Les critères d'évaluation

Une **ressource** est l'entité concrète qui véhicule de l'information.

Dans le cadre d'une recherche d'information sur le web, les ressources contribuant à répondre au besoin d'information peuvent prendre des formes diverses : page web, notice bibliographique, document au format PDF, ouvrage, image, etc.

Évaluer une ressource consiste à étudier :

- sa **pertinence** : l'information qu'elle véhicule correspond-t-elle au besoin d'information ?
- sa **qualité** : l'information qu'elle contient est-elle fiable et est-elle bien présentée ?

La **pertinence** d'une ressource est l'adéquation de son contenu informationnel avec le **besoin d'information**.

Pour cela on peut répondre aux questions suivantes :

- le contenu relève-t-il bien du **domaine** sur lequel porte la recherche ?
- le contenu répond-il aux **questions posées** initialement ?
- le contenu apporte-t-il des **informations utiles** à la recherche au-delà des questions posées initialement ?

La **qualité** d'une ressource repose sur :

- la **fianabilité** de l'information, c'est-à-dire la confiance qu'on peut accorder à sa **véridicité** ;
- la qualité de la **présentation** de l'information, qui recouvre la qualité de la rédaction (structuration de l'information, expression écrite, illustrations), ainsi que sa lisibilité et son accessibilité.

La **fianabilité** d'une information est liée au contexte dans lequel elle a été communiquée (processus de validation) et à l'intention sous-jacente (commerciale, académique, officielle, etc.).

On peut distinguer « l'information connaissance » (information scientifique et technique, garantie par la validation de chercheurs et spécialistes) et « l'information opinion » (information journalistique, publicitaire, issue des pseudo-sciences, qui est le produit du sens commun, de l'évidence, des préjugés, des stéréotypes).

L'évaluation de la fiabilité

En pratique, pour évaluer la **fiabilité d'une information** trouvée sur le web, il faut :

- identifier sa **source** et en évaluer la **fiabilité** ;
- identifier l'**auteur**, repérer son lien avec la source (affiliation, hébergement, etc.) et évaluer sa **notoriété** ;
- évaluer la **fraîcheur** de l'information ;
- évaluer l'**objectif** de l'information (informer, expliquer, convaincre, etc.) ;
- identifier le **public visé**.

Identifier la source

On peut souvent identifier une source en consultant la **page d'accueil** du site et en décryptant l'**URL** de la ressource.

Dans une URL, on peut discerner :

- le domaine (fr, com, org, etc.) : il caractérise l'orientation générale du site (français, commercial, organisation, etc.) ;
- le nom de domaine (education.fr, over-blog.com, openoffice.org, etc.) : il permet souvent de repérer la **nature** (universitaire, institutionnel, commercial, associatif, etc.) et le **statut** du site (officiel, privé, personnel, etc.) ;
- le sous-domaine éventuel (indiqué en préfixe).

Par exemple : dans <http://c2i.education.fr/>, le nom de domaine est education.fr et le sous-domaine est c2i.

Tout le monde peut acheter un nom de domaine. Il existe des sites spécialisés de **WHOIS** sur le web qui permettent de consulter les informations relatives au propriétaire d'un nom de domaine.

Par exemple : en consultant le [Whois de l'Afnic](#), on voit que le titulaire du nom de domaine education.fr est « MIN EDUCATION NAT ENSEIGN SUP RECHERCHE ».

Évaluer la notoriété de l'auteur

Pour évaluer la notoriété d'un auteur, on peut se fier aux indices fournis par lui-même ou d'autres en lien avec :

- son **activité** professionnelle ou associative (métier, expérience, réalisations, statut, niveau d'études) ;
- son **affiliation** (entreprise, organisation, gouvernement, université, etc.) ;
- sa **reconnaissance** par ses pairs.

Évaluer la fraîcheur de l'information

Pour évaluer la fraîcheur de l'information, il faut repérer la **date de création** de la ressource et éventuellement la date de sa **dernière mise à jour**. La **fréquence de mise à jour** de la source est aussi un repère très utile.

Évaluer l'objectif de l'information

Le **type de la source** (politique, promotionnel, journalistique, éducatif, institutionnel, scientifique et technique, vulgarisateur, personnel, culturel, etc.) révèle souvent l'objectif sous-jacent (influence de l'opinion, consommation, information, communication ou mise à disposition d'un savoir au public, expression d'une opinion, enrichissement intellectuel, etc.).

Identifier le public visé

Le **public visé** (grand public, spécialistes) contribue à éclairer l'utilité d'une ressource pour le besoin d'information.

Les règles de publication

Pour évaluer la fiabilité d'une source, on pourra s'intéresser à ses règles de publication.

Processus de validation inexistant

Le contenu des **sites web personnels, commerciaux ou d'opinion** ne font généralement l'objet d'**aucune validation**. Cela ne va pas à l'encontre de leur popularité, qui peut être excellente alors même que les informations peuvent s'avérer fausses.

Processus de validation traditionnel

Les **publications scientifiques** produites dans un contexte académique font souvent l'objet d'un processus de **relecture par les pairs**, c'est-à-dire par des spécialistes du domaine eux-mêmes reconnus par leurs pairs. C'est une excellente garantie de fiabilité. Lorsque ce n'est pas le cas, on parle de littérature grise.

De même, les **sites institutionnels** publient des informations qui sont validées au préalable au sein de l'institution, du fait du rôle performatif des informations diffusées (textes de loi, recommandations, informations officielles, etc.).

Une ressource qui cite ses sources est toujours plus fiable !

Processus de validation original

Certains sites contiennent des ressources qui font l'objet d'un **processus de validation original**.

C'est le cas du projet d'encyclopédie collective Wikipedia, que chacun peut modifier et améliorer. Les principes fondateurs du projet sont publiés et décrivent les conditions de son élaboration notamment en termes de :

- nature des contenus : c'est une encyclopédie généraliste et spécialisée, qui ne doit pas contenir de productions originales, de propagande, de débat d'opinion ;
- qualité des contenus : la recherche de l'exactitude est une règle qui s'impose aux contributeurs, qui doivent citer leurs sources et s'astreindre à une neutralité de point de vue.

Il va sans dire que certaines pages de cette encyclopédie peuvent contenir des erreurs, ou adopter des points de vue qui ne sont pas neutres. C'est un collectif qui organise la maintenance de cette encyclopédie et en fait respecter les principes fondateurs, dans la mesure de ses moyens.

On repère parfois des bandeaux en tête d'un article de Wikipédia : ils aident le lecteur à évaluer la qualité de l'information contenue dans l'article. Ils signalent par exemple lorsqu'un article est une ébauche, ne cite pas suffisamment ses sources, ou contient des informations à vérifier.

La compétence en questions

Q 1 : Quels facteurs entrent en compte dans l'ordre des réponses d'un moteur de recherche sur le web ?

- 1 - La réputation de l'auteur.
- 2 - La popularité de la page web.
- 3 - Le nombre de mots de la page web.
- 4 - La correspondance avec les mots de la requête.

Q 2 : Dans l'URL <http://catalogue.bnf.fr/>, quel est le nom de domaine ?

- 1 - catalogue
- 2 - bnf
- 3 - fr
- 4 - catalogue.bnf.fr
- 5 - bnf.fr

Q 3 : Que peut-on dire de Wikipédia ?

- 1 - Toute information est contrôlée par un expert avant d'être publiée.
- 2 - Tout internaute peut compléter un article.
- 3 - Tout internaute peut voir les modifications récentes d'un article.
- 4 - Tout article peut être modifié à n'importe quel moment.

La compétence en situations

Commentez avant de jeter...

Dans le cadre d'une recherche d'information que vous devez mener, vous sélectionnez des ressources qui vous serviront de référence.

Pour une fois, gardez la trace des ressources que vous éliminez et explicitez pour chacune les raisons de votre choix.

Fiable ou pas fiable ?

A partir d'une ressource du web, pouvez-vous identifier l'auteur et ses compétences ?

Quelles informations vous donne l'URL ?

Connaissez-vous la dernière date de mise à jour de la ressource ?

Controverse et qualité

Trouvez un sujet d'actualité qui fait polémique.

Recherchez une page de Wikipedia qui traite de ce sujet.

Cette page contient-elle un bandeau d'information ou d'avertissement ?

En quoi les principes fondateurs de cette encyclopédie collective sont-ils respectés ou enfreints ?

D4.3 : Référencer une ressource numérique en ligne

La disponibilité d'une ressource

Le web contient de nombreuses ressources auxquelles on peut être amené à faire référence pour citer ses sources. Cependant, bon nombre de ces ressources sont **éphémères**, ou **instables** dans la mesure où leur contenu change au fil du temps.

En voici quelques exemples typiques.

Un article de Wikipedia a été modifié entre le moment où on l'a consulté pour y faire référence dans un rapport, et le moment où le lecteur du rapport le consulte.

La page d'accueil d'un quotidien présente sa « une », qui change tous les jours, voire plusieurs fois par jour.

Lorsqu'on fait référence à des ressources, il faut être conscient de ces difficultés et prendre quelques précautions.

Trouver la bonne URL

Une page web affiche parfois plusieurs ressources, qui sont réparties sur l'ensemble de la page. C'est le cas notamment des pages d'accueil des sites d'actualités. Pour identifier une de ces ressources, il faut la sélectionner : l'URL correspondante est alors spécifique à la ressource, et présente généralement une meilleure pérennité.

Exemple : La page d'accueil du quotidien Le Monde affiche aujourd'hui plusieurs articles, dont un qui est intitulé « Une pluie exceptionnelle d'étoiles filantes attendue samedi ». L'URL <http://www.lemonde.fr/> permet d'accéder à cette page aujourd'hui, mais ce ne sera plus le cas demain. En revanche, en cliquant sur le titre de l'article, la nouvelle URL intègre la date et le titre de l'article, ce qui laisse espérer une plus grande pérennité : http://www.lemonde.fr/planete/article/2011/10/08/une-pluie-d-etoiles-filantes-exceptionnelle-attendue-samedi-soir_1584572_3244.html

Indiquer la date de consultation

Lorsqu'on construit une argumentation en se référant au contenu d'une ressource en ligne, il est embarrassant que ce contenu change entre le moment où on l'utilise et le moment où d'autres sont susceptibles de vérifier cette référence.

Pour se prémunir contre toute critique, il est recommandé de noter la date à laquelle on a consulté la ressource.

Exemple : L'article « Une pluie exceptionnelle d'étoiles filantes attendue samedi » publié le 8 octobre 2011 à 16h51 sur le site du Monde a peut-être été mis à jour le lendemain suite aux événements survenus dans la nuit. En indiquant qu'on a consulté l'article le 8 octobre 2011, on permet au lecteur de se rendre compte qu'il ne consulte peut-être pas la même version.

Faire une sauvegarde de la ressource

Pour retrouver à coup sûr le contenu d'une ressource dans l'état où on l'a consultée, on peut en faire une sauvegarde.

Certains moteurs de recherche permettent d'accéder pendant quelques temps à une ressource disparue par l'intermédiaire du lien « En cache ».

La récupération

Sauvegarder une ressource en ligne consiste à l'enregistrer sur son poste de travail.

Enregistrer une page web

Une page web est composée d'un code source en langage HTML, qui ne comporte que du texte et des liens vers d'autres ressources. Ces ressources (images, vidéos, feuille de style, etc.) sont également localisée sur le serveur web mais dans des fichiers à part.

Les navigateurs permettent d'enregistrer une page web pour la conserver sur son ordinateur. Deux options sont offertes :

- **page web complète** : dans ce cas l'intégralité de la page est enregistrée, y compris les ressources qui la composent (images, vidéos, etc.) qui sont placées dans un dossier portant le même nom que le fichier au format HTML ;
- **HTML seulement** : seul le code source de la page est enregistré ; cela permet de récupérer le texte de la page mais pas les ressources qui la composent.

Lorsqu'une page web est composée de façon dynamique en puisant certaines données dans une base de données, celles-ci peuvent être perdues à l'enregistrement.

Enregistrer un document accessible en ligne

On trouve souvent en ligne des documents qui peuvent être téléchargés pour être consultés en dehors du navigateur. Lorsqu'on clique sur le lien correspondant, le navigateur propose généralement de l'enregistrer ou de l'ouvrir.

On peut aussi directement demander à **enregistrer la cible du lien**.

Les navigateurs ont souvent des extensions qui permettent d'afficher des documents de divers types ; cela permet de les ouvrir sans les enregistrer explicitement.

Récupérer une partie d'un document

On peut être amené à récupérer une partie d'une page web pour l'intégrer comme une citation dans un document.

S'il s'agit d'une portion de texte, il peut être utile de l'intégrer par copier/coller comme du **texte non formaté** : cela évite que les styles de texte de la page web s'intègrent mal à la charte graphique de son propre document.

S'il s'agit d'une image, le navigateur permet d'**enregistrer l'image** sur laquelle on clique : cela permet de réutiliser ce fichier image dans n'importe quel autre document.

S'il s'agit d'un son ou d'une vidéo, il n'y a en général pas de fonctionnalité standard intégrée dans le navigateur pour le télécharger, mais on peut ajouter des **extensions** au navigateur qui le permettent.

Consulter un site web hors connexion

La plupart des navigateurs permettent de naviguer hors connexion sur les pages déjà consultées dans un passé récent. Il suffit d'enclencher la fonctionnalité « **Travailler hors connexion** » pour que le navigateur affiche la page web qu'il a stockée dans le **cache du navigateur** lors de la dernière consultation.

Cette approche est limitée, puisqu'on ne peut accéder qu'aux pages web déjà visitées.

Si on souhaite pouvoir accéder hors connexion à un site web complet, il faut utiliser un aspirateur de sites web.

La référence à une ressource en ligne

Quand on utilise des ressources, que ce soit pour en citer des extraits ou pour y faire référence, il faut dresser la liste des références bibliographiques en respectant certaines règles.

Dans le cas des ressources numériques en ligne, il faut ajouter des renseignements spécifiques.

Norme

Il existe une norme qui définit précisément la composition et la typographie d'une référence bibliographique à un document électronique (référence ISO 690-2 ou Z 44-005-2).

Les règles sont les mêmes que pour les documents traditionnels, mais il faut ajouter :

- le type de support (en ligne, CD-ROM, enr. sonore, etc.) ;
- l'adresse de la ressource (URL) ;
- la date de consultation de la ressource.

Modèle de base pour un site web

Auteur. Date. *Titre* [Support]. Adresse de la ressource. (consulté le ...)

- Auteur : l'organisme ou la personne physique dans le cas d'une page personnelle.
- Date : celle de la dernière mise à jour si elle est connue.
- Titre : celui de la page d'accueil du site.
- Support : il s'écrit entre crochets ; on trouve : [En ligne] [CD-ROM] [Enr. sonore], etc.
- Adresse de la ressource : c'est l'URL.
- Date de consultation par l'utilisateur : elle s'écrit (consulté le jour mois année).

Exemple : Ministère de l'enseignement supérieur et de la recherche. *Accueil portail des C2i* [En ligne]. <http://www.c2i.education.fr/> (consulté le 10/10/2011).



Modèle de base pour une ressource Internet, y compris une page web

Auteur. Date. «Titre de la ressource». In : Ressource plus large, [Support] Adresse de la ressource (consulté le ...)

- Ressource plus large : s'il y a lieu, ajouter la ressource plus large à laquelle le document cité est rattaché; procéder de la façon suivante: In Auteur/titre ou titre du site ou du document qui contient la ressource.

Exemple : Caron, Rosaire. «Comment citer un document électronique?». In : Université Laval - Bibliothèque. *Site de la Bibliothèque de l'Université Laval*. [En ligne]. <http://www.bibl.ulaval.ca/doelec/citedoce.html> (consulté le 10 octobre 2011)

En savoir plus...

[Citer une référence électronique](#) un résumé des pratiques avec des exemples

[Comment citer un document électronique ?](#) une description très complète

La compétence en questions

Q 1 : Un document cite un article du Monde intitulé "Ce week-end, canicule dans le Pas-de-Calais" en le référant par l'URL <http://www.lemonde.fr/>. A quoi peut-on s'attendre dans les jours qui suivent ?

- 1 - L'URL n'existera plus.
- 2 - L'URL ne présentera plus le même article.
- 3 - L'URL redirigera vers la page d'accueil des archives du Monde.
- 4 - L'URL redirigera vers un article sur la météo du jour dans le Pas-de-Calais.

Q 2 : Dans une page web, il y a un lien vers un fichier au format PDF. Comment faire pour télécharger ce fichier sur son poste ?

- 1 - On enregistre l'URL.
- 2 - On enregistre la page html seule.
- 3 - On enregistre la cible du lien.
- 4 - On enregistre la page web complète.

Q 3 : Quelles sont les informations à indiquer quand on fait référence à une page web dans un rapport ?

- 1 - L'URL.
- 2 - La date de consultation.
- 3 - L'auteur.
- 4 - Le propriétaire du nom de domaine.
- 5 - Le mot de passe pour y accéder.

La compétence en situations

Pages d'accueil changeantes...

Allez sur le site de l'UNESCO.

Selon vous, à quelle fréquence la page d'accueil change-t-elle ?

Repérez une actualité de ce site et écrivez sa référence bibliographique.

Quels éléments retrouve-t-on dans l'URL ?

Que laissent-ils penser de la pérennité de cette URL ?

Enregistrer une page web avec une vidéo

Allez sur le site de partage de vidéos de votre choix.

Trouvez une vidéo qui vous intéresse.

Enregistrez la page web qui affiche cette vidéo.

Que constatez-vous lorsque vous consultez la page que vous avez enregistrée ?

Citer un article d'encyclopédie en ligne

Repérez un article de l'encyclopédie en ligne de votre choix.

Écrivez sa référence bibliographique.

D4.4 : Organiser une veille informationnelle

La veille informationnelle

La veille informationnelle consiste à se tenir au courant de l'actualité d'un domaine ou d'un sujet.

Pour faciliter ce travail, on peut :

- agréger les flux d'actualités diffusés par différents sites ;

Cela permet d'afficher les derniers articles dans un même espace pour éviter de visiter les différents sites un par un.

- s'abonner à des lettres d'information ou *newsletters* ;

Ces lettres peuvent être hebdomadaires ou mensuelles, et sont généralement éditées par des « veilleurs » spécialisés dans un domaine.

- suivre un ou plusieurs veilleurs sur un système de microblogage ;

Sur un site de microblogage, on peut choisir de suivre un petit nombre de personnes parmi les bons « veilleurs » du domaine qui nous intéresse : on peut ainsi recevoir dans les meilleurs délais une sélection de l'actualité sans aucun effort.

- paramétrer des notifications pour être prévenu par courriel de nouveaux événements.

On peut se faire notifier de l'arrivée de nouveaux messages dans un forum pour éviter de le visiter tous les jours inutilement.

On trouve souvent une fonction de notification ou d'alerte dans les outils de partage de documents, ou de calendrier partagé : on est alerté lorsqu'une modification a été apportée.



L'agrégateur de flux

Un **flux d'informations** sur le web est une succession d'articles ou de contenus multimédia (podcasts) diffusés à une même URL sur le web.

On parle aussi de **fil d'actualité** ou de fil d'information.

Si on fait afficher un flux directement dans un navigateur, on voit la liste des titres des articles ou contenus par ordre chronologique inverse, sans aucune mise en forme.

La **syndication de contenus** est le procédé qui consiste à récupérer les informations d'un site web pour les intégrer dans un autre.

Le terme **flux RSS** (Real Simple Syndication) désigne la technologie la plus répandue de flux web.

Cette technologie s'appuie sur un format de diffusion basé sur le langage XML.

Le flux Atom est une autre technologie possible.

Un **agrégateur de flux** est un logiciel qui permet d'afficher dans un même espace toutes les actualités en provenance de flux d'informations du web.

Un agrégateur de flux peut se présenter comme :

- un service en ligne ;

Ce type de service permet souvent de disposer sur une page web des « blocs » dans lesquels arrivent les différents flux choisis.

A l'URL d'un flux, on trouve souvent des fonctionnalités d'ajout en un clic à un des grands services d'agrégation en ligne du marché (Netvibes, iGoogle, etc.).

- un logiciel à installer en local.

Ces logiciels présentent les flux comme une liste de courriels, où on peut afficher tous les flux mélangés par ordre chronologique inverse, ou au contraire consulter chaque flux séparément.

La plupart des logiciels de messagerie comportent un agrégateur de flux. C'est le cas de Mozilla Thunderbird par exemple.



Agréger un flux d'information

Pour ajouter un flux dans un agrégateur, il faut indiquer l'URL du flux.

Pour trouver l'URL d'un flux, le plus simple est de :

- se rendre sur le site qui diffuse le flux d'actualité en question ;
- repérer la petite icône orange typique des flux RSS et cliquer dessus.

Il existe de nombreux annuaires de flux sur le web, qui permettent de trouver de nouveaux flux par domaine.

Exemple : l'URL du flux de la Une du quotidien « Le Monde » : <http://rss.lemonde.fr/c/205/f/3050/index.rss>

La lettre d'information

Une **lettre d'information** électronique, ou *newsletter*, diffuse régulièrement par courriel des informations sur un sujet donné.

S'abonner à une lettre d'information revient le plus souvent à s'abonner à une **liste de diffusion** fonctionnant selon des règles particulières :

- seuls les propriétaires de la liste peuvent envoyer des messages ;
- les abonnés ne peuvent que recevoir les messages ainsi diffusés.

Les différents types de lettres d'information

Les lettres d'information **thématiques** sont produites par une **communauté rassemblée autour d'un thème**. C'est le cas de communautés scientifiques notamment, qui échangent beaucoup par cet intermédiaire.

Dans ce contexte, les lettres d'information sont souvent une compilation des messages transmis par les membres de la communauté au modérateur de la liste.

Les lettres d'information **institutionnelles** sont diffusées par une institution ou un de ses **services**.

Par exemple le service d'une université chargé de l'orientation et de l'insertion professionnelle peut rédiger et diffuser une lettre mensuelle contenant des actualités et des conseils.

S'abonner et de désabonner

Les lettres d'information étant des cas particuliers de listes de diffusion, le système de gestion des abonnements est le même : il est automatisé par un **robot logiciel** auquel on passe commande (abonnement, désabonnement) par courriel.

Certains sites web permettent de s'abonner facilement à leur lettre d'information.

Le site de la CNIL propose sur sa page d'accueil un lien pour s'abonner à sa lettre d'information.

Souvent, les lettres d'informations contiennent un lien permettant de se désinscrire en un clic.

Le microblogage

Les plateformes de **microblogage** permettent aux usagers de publier des messages très courts ou *tweets* destinés à informer les usagers qui les « suivent » de leur propre actualité, ou de celle du domaine qui les intéresse.

Le microblogage est un outil de veille thématique très utile.

Comment se fait la veille par les tweets ?

Les « veilleurs » professionnels d'un domaine sont à l'écoute des flux d'information des sites d'actualité de leur domaine. Ils sélectionnent les informations intéressantes et rédigent un tweet qui décrit brièvement l'information et intègre un lien vers l'article complet. Ainsi les usagers qui les suivent reçoivent une sélection de l'actualité du domaine.

La taille des tweets étant limitée à 140 caractères, il existe des services de « raccourcissement d'URL » qui transforment une URL ordinaire, bien souvent trop longue pour un tweet, en une URL courte.

Ces veilleurs suivent aussi les tweets des autres veilleurs de leur domaine, ce qui les amène souvent à « retweeter » un tweet reçu afin d'en informer leur propre réseau de « suiveurs ».

Comment trouver un veilleur à suivre ?

L'auteur d'un tweet est repérable par son identifiant qui est souvent un pseudo.

Les plateformes de microblogage permettent :

- de rechercher un usager sur la base de son pseudo, voire des éléments de son profil ;

Il faut connaître le pseudo d'un usager pour le trouver facilement. Cependant les veilleurs professionnels utilisent souvent leurs véritables nom et prénom.

- de rechercher des tweets par mots-clé : cela peut aider à repérer les usagers qui émettent des tweets intéressants.

On peut décider de suivre directement l'auteur d'un message qu'on nous a retweeté.

Le nomadisme

Pour organiser sa veille, l'utilisateur agrège des flux d'informations, s'abonne à des lettres d'informations, suit des « tweeters avertis », mémorise ses sites favoris, etc.

Pour rendre cette **veille nomade**, il suffit de s'inscrire sur certains services en ligne qui donnent accès à ses abonnements et préférences à partir de tout navigateur connecté à Internet.

Parmi ces services, on distingue :

- les tableaux de bord personnalisés qui intègrent des « fenêtres gadgets » ou des widgets (window + gadget) de veille spécifiques (actualités, résultats sportifs, météo, les films de la semaine, courriel, réseau social, etc.). Certaines de ces fenêtres permettent d'agréger ses flux d'information ou ses signets préférés ;

Netvibes ou iGoogle offrent de tels services de tableau de bord personnalisé.

- les ENT qui donnent accès à un certain nombre de services en ligne : stockage de signets, annonces, messagerie, etc. ;
- les sites de microblogage.

Pour passer le temps dans les transports en commun, Gérard consulte les derniers tweets arrivés sur son téléphone portable.

Les signets en ligne

Le système des favoris intégré au navigateur n'est pas adapté aux usages nomades actuels.

Il existe des sites spécialisés dans le stockage des signets en ligne offrant parfois la possibilité de les partager.

Ces plateformes proposent souvent des fonctionnalités collaboratives : taguer les signets, faire une recherche dans les signets des autres en tenant compte de la popularité (combien d'utilisateurs ont conservé ce signet), suivre les signets ajoutés par un utilisateur.

Zotero est une plateforme qui permet d'alimenter très facilement une bibliographie en ligne, et même de la partager avec d'autres.

La compétence en questions

Q 1 : Quels sont les outils permettant d'organiser une veille informationnelle ?

- 1 - Les agrégateurs de flux.
- 2 - Les services d'alertes par courriel.
- 3 - Les vérificateurs d'orthographe.
- 4 - Les outils de mise à jour automatique de logiciels.
- 5 - Les abonnements à des listes de diffusion.

Q 2 : Si un usager est abonné à la lettre d'information d'un site commercial, que peut-on affirmer ?

- 1 - Il connaît les adresses électroniques des autres abonnés.
- 2 - Il peut se désabonner en envoyant un courriel spécifique.
- 3 - Il est abonné à un flux RSS.
- 4 - Il reçoit l'information dans sa boîte à lettres électronique.

Q 3 : Quels termes évoquent un flux RSS ?

- 1 - La syndication de contenus.
- 2 - Le publipostage.
- 3 - Le fil d'actualités.
- 4 - L'agrégateur.
- 5 - Le débit binaire.

La compétence en situations

Agrégez vos flux !

Dans votre ENT, dans votre logiciel de messagerie, ou encore dans un service de tableau de bord en ligne, abonnez-vous à un flux traitant des TICE.

A quelle fréquence ces flux sont-ils actualisés ?

Suivre les discussions d'un forum

Trouvez un forum dont le sujet vous intéresse.

Quelles sont les différents types de notifications proposés ?

Veille par les tweets

Consultez la « Une » du Monde en ligne et repérez un thème d'actualité en lien avec vos préoccupations.

Allez sur Twitter et repérez les tweets qui traitent de ce sujet.

Repérez les comptes qu'il vous semble intéressant de suivre.

Les réponses aux questions

Les réponses aux questions

D4.1 : Rechercher de l'information avec une démarche adaptée

Q 1 : Que contiennent les catalogues de bibliothèque ?

- 1 - Des adresses de bibliothèques.
- 2 - [Des notices bibliographiques.](#)
- 3 - Des livres à acheter.
- 4 - Des articles de journaux.

Q 2 : Qui parcourt et indexe les ressources du web dans un moteur de recherche généraliste ?

- 1 - Des internautes.
- 2 - Des experts.
- 3 - [Des logiciels robots.](#)
- 4 - Des bibliothécaires.

Q 3 : Quelle est la requête la plus précise pour rechercher les ressources contenant l'expression poisson rouge au singulier ou au pluriel ?

- 1 - poisson rouge
- 2 - poisson OR rouge
- 3 - ["poisson rouge" OR "poissons rouges"](#)
- 4 - poisson* rouge*

D4.2 : Évaluer les résultats d'une recherche

Q 1 : Quels facteurs entrent en compte dans l'ordre des réponses d'un moteur de recherche sur le web ?

- 1 - La réputation de l'auteur.
- 2 - La popularité de la page web.
- 3 - Le nombre de mots de la page web.
- 4 - La correspondance avec les mots de la requête.

Q 2 : Dans l'URL <http://catalogue.bnf.fr/>, quel est le nom de domaine ?

- 1 - catalogue
- 2 - bnf
- 3 - fr
- 4 - catalogue.bnf.fr
- 5 - bnf.fr

Q 3 : Que peut-on dire de Wikipédia ?

- 1 - Toute information est contrôlée par un expert avant d'être publiée.
- 2 - Tout internaute peut compléter un article.
- 3 - Tout internaute peut voir les modifications récentes d'un article.
- 4 - Tout article peut être modifié à n'importe quel moment.

D4.3 : Référencer une ressource numérique en ligne

Q 1 : Un document cite un article du Monde intitulé "Ce week-end, canicule dans le Pas-de-Calais" en le référant par l'URL <http://www.lemonde.fr/>. A quoi peut-on s'attendre dans les jours qui suivent ?

- 1 - L'URL n'existera plus.
- 2 - L'URL ne présentera plus le même article.
- 3 - L'URL redirigera vers la page d'accueil des archives du Monde.
- 4 - L'URL redirigera vers un article sur la météo du jour dans le Pas-de-Calais.

Q 2 : Dans une page web, il y a un lien vers un fichier au format PDF. Comment faire pour télécharger ce fichier sur son poste ?

- 1 - On enregistre l'URL.
- 2 - On enregistre la page html seule.
- 3 - On enregistre la cible du lien.
- 4 - On enregistre la page web complète.

Q 3 : Quelles sont les informations à indiquer quand on fait référence à une page web dans un rapport ?

- 1 - L'URL.
- 2 - La date de consultation.
- 3 - L'auteur.
- 4 - Le propriétaire du nom de domaine.
- 5 - Le mot de passe pour y accéder.

D4.4 : Organiser une veille informationnelle

Q 1 : Quels sont les outils permettant d'organiser une veille informationnelle ?

- 1 - Les agrégateurs de flux.
- 2 - Les services d'alertes par courriel.
- 3 - Les vérificateurs d'orthographe.
- 4 - Les outils de mise à jour automatique de logiciels.
- 5 - Les abonnements à des listes de diffusion.

Q 2 : Si un usager est abonné à la lettre d'information d'un site commercial, que peut-on affirmer ?

- 1 - Il connaît les adresses électroniques des autres abonnés.
- 2 - Il peut se désabonner en envoyant un courriel spécifique.
- 3 - Il est abonné à un flux RSS.
- 4 - Il reçoit l'information dans sa boîte à lettres électronique.

Q 3 : Quels termes évoquent un flux RSS ?

- 1 - La syndication de contenus.
- 2 - Le publipostage.
- 3 - Le fil d'actualités.
- 4 - L'agrégateur.
- 5 - Le débit binaire.