

Environnement numérique d'information et de communication

M1104 D.U.T. GEA en alternance 2015-2016



domenico.ruoppolo@iutv.univ-paris13.fr

Domaine D4

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

Sous-domaine D4.1

Rechercher de l'information
avec une démarche adaptée

Cherchez *quoi*?

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.

Cherchez *quoi*?

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.

Cherchez *quoi*?

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

- les sources d'information **traditionnelles** qui diffusent des informations validées selon un processus reconnu;
- les sources **émergeant de structures collectives** plus ou moins formelles, dont les règles de validation de l'information sont plus ou moins strictes;
- les sources **informelles** (pages web personnelles, blogs d'internautes, forums, ...) publiées sans aucun contrôle.

Un autre critère de classification des **sources d'information** est:

- les sources d'information **généralistes**;
- les sources d'information **spécialisées**.

Googlez !

Un **moteur de recherche** est un *logiciel* – auquel on accède dans une page web – permettant de retrouver des pages web à partir d'une *requête*. Il réalise les tâches suivantes.

Googlez !

Un **moteur de recherche** est un *logiciel* – auquel on accède dans une page web – permettant de retrouver des pages web à partir d'une *requête*. Il réalise les tâches suivantes.

Avant de toute interaction avec l'utilisateur :

- **moissonnage** : parcour automatiquement par un *robot*;
- **indexation automatique** : extractions mots significatifs;

Googlez !

Un **moteur de recherche** est un *logiciel* – auquel on accède dans une page web – permettant de retrouver des pages web à partir d'une *requête*. Il réalise les tâches suivantes.

Avant de toute interaction avec l'utilisateur:

- **moissonnage** : parcour automatiquement par un *robot*;
- **indexation automatique** : extractions mots significatifs;

Après toute interaction avec l'utilisateur:

- **interrogation** en fonction d'une requête saisie par l'utilisateur;
- **affichage** des résultats dans un certain ordre.

Les limites du moissonage

Certaines pages du web ne peuvent pas être moissonnées: elles composent le **web invisible** ou **web profond** (**deepnet**).

- pages qui ne sont accessibles qu'après *authentification*;
- pages *dynamiques* qui sont composées à partir de données stockées dans des bases de données (exemple: page SNCF);
- l'administrateur d'un site web peut interdire l'indexation de certaines pages: cela arrive souvent pour des *forum*.

Remark: à ne pas confondre avec le **web ombre** (**darknet**)!!!

Les limites de l'indexation automatique

- Certains documents en ligne ne sont pas indexés car le moteur *ne sait pas traiter leur format* (gros problème dans le passé).
- pages *dynamiques* qui sont composées à partir de données stockées dans des bases de données (exemple: page SNCF).
- les textes très courts ou mal écrits sont difficiles à indexer.

Les limites de l'indexation automatique

- Certains documents en ligne ne sont pas indexés car le moteur *ne sait pas traiter leur format* (gros problème dans le passé).
- pages *dynamiques* qui sont composées à partir de données stockées dans des bases de données (exemple: page SNCF).
- les textes très courts ou mal écrits sont difficiles à indexer.

Remark: pour *indexer les images*, les moteurs utilisent les métadonnées associées, le texte entourant l'image ou les tags laissés par les internautes.

Quelques astuces pour les requêtes

La requête est l'ensemble des *mots-clés* de recherche saisis. Les moteurs de recherche décident automatiquement de la façon de combiner ces mots.

Pour forcer cette façon:

- le signe + devant un mot-clé signifie qu'il est *obligatoire*;
- le signe - devant un mot-clé s'il *ne doit pas apparaître*;
- les guillemets " " pour rechercher une expression *exacte*;
- l'astérisque * avec la signification de *n'importe quoi...*

Des choses similaires pour des recherches dans les catalogue de bibliothèque : ET , OU , SAUF ...

Sous-domaine D4.2

Évaluer les résultats d'une
recherche

Les ressources

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

évaluer une ressource consiste à étudier:

- sa **pertinence**, l'adéquation à notre besoin:
 - ▶ 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 d'autres informations utiles à la recherche?
- sa **qualité** :
 - ▶ **fiabilité**, confiance qu'on peut accorder à sa véridicité;
 - ▶ la **qualité de la présentation**.

La fiabilité des ressources

Pour évaluer la fiabilité d'une information trouvée sur le web :

- identifier sa source et en évaluer la fiabilité:
 - ▶ page avec processus de validation **inexistant**
 - ▶ page avec processus de validation **traditionnel**
 - ▶ page avec processus de validation **original**
- identifier l'auteur et évaluer sa notoriété:
 - ▶ activité professionnelle ou formation
 - ▶ son affiliation (entreprise, organisation, etc.)
 - ▶ sa reconnaissance
- évaluer la fraîcheur de l'information ;
- évaluer l'objectif de l'information (informer, expliquer, convaincre, etc.) ;
- identifier le public visé.

La fiabilité des ressources

Pour évaluer la fiabilité d'une information trouvée sur le web :

- identifier sa source et en évaluer la fiabilité:
 - ▶ page avec processus de validation inexistant
 - ▶ page avec processus de validation traditionnel
 - ▶ page avec processus de validation original
- identifier l'auteur et évaluer sa notoriété:
 - ▶ activité professionnelle ou formation
 - ▶ son affiliation (entreprise, organisation, etc.)
 - ▶ sa reconnaissance
- évaluer la fraîcheur de l'information ;
- évaluer l'objectif de l'information (informer, expliquer, convaincre, etc.) ;
- identifier le public visé.

La fiabilité des ressources

Pour évaluer la fiabilité d'une information trouvée sur le web :

- identifier sa source et en évaluer la fiabilité:
 - ▶ page avec processus de validation inexistant
 - ▶ page avec processus de validation traditionnel
 - ▶ page avec processus de validation original
- identifier l'auteur et évaluer sa notoriété:
 - ▶ activité professionnelle ou formation
 - ▶ son affiliation (entreprise, organisation, etc.)
 - ▶ sa reconnaissance
- évaluer la **fraîcheur** de l'information ;
- évaluer l'objectif de l'information (informer, expliquer, convaincre, etc.) ;
- identifier le public visé.

La fiabilité des ressources

Pour évaluer la fiabilité d'une information trouvée sur le web :

- identifier sa source et en évaluer la fiabilité:
 - ▶ page avec processus de validation inexistant
 - ▶ page avec processus de validation traditionnel
 - ▶ page avec processus de validation original
- identifier l'auteur et évaluer sa notoriété:
 - ▶ activité professionnelle ou formation
 - ▶ son affiliation (entreprise, organisation, etc.)
 - ▶ sa reconnaissance
- évaluer la fraîcheur de l'information ;
- évaluer l'**objectif** de l'information (informer, expliquer, convaincre, etc.) ;
- identifier le public visé.

La fiabilité des ressources

Pour évaluer la fiabilité d'une information trouvée sur le web :

- identifier sa source et en évaluer la fiabilité:
 - ▶ page avec processus de validation inexistant
 - ▶ page avec processus de validation traditionnel
 - ▶ page avec processus de validation original
- identifier l'auteur et évaluer sa notoriété:
 - ▶ activité professionnelle ou formation
 - ▶ son affiliation (entreprise, organisation, etc.)
 - ▶ sa reconnaissance
- évaluer la fraîcheur de l'information ;
- évaluer l'objectif de l'information (informer, expliquer, convaincre, etc.) ;
- identifier le **public visé**.

Sous-domaine D4.3

Référencer une ressource
numérique en ligne