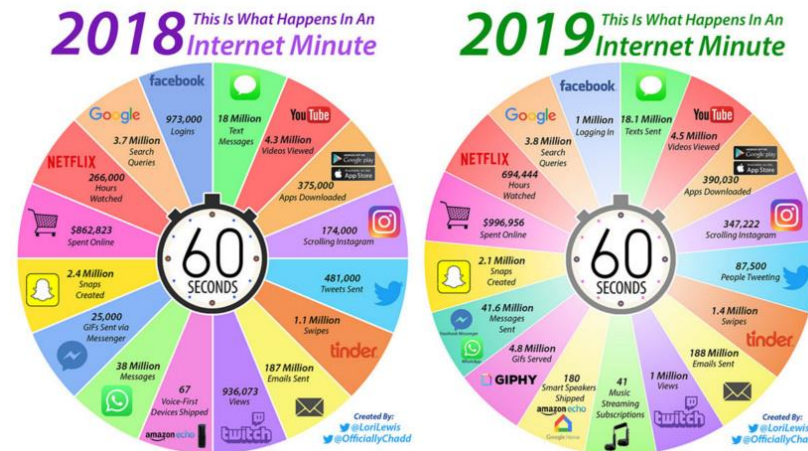


Archivage du Web et des médias sociaux

Alain Dubois

Quelques statistiques

- 4.65 milliards de pages sur Internet pour 1 milliard de sites (uniquement le côté « visible » du Web par les moteurs de recherche) (*Science et vie*, mars 2016)
- Internet Live Stats (<http://www.internetlivestats.com>)



- Quelle politique d'archivage électronique mettre en œuvre?

Différentes approches

- Approche intégrale
 - Il s'agit de collecter l'entier du web, sans distinction, ni critère de sélection (cf. projet Internet Archive (<https://archive.org/>))
- Approche exhaustive
 - Il s'agit de viser la complétude dans un périmètre circonscrit (domaine, espace national ou type de sites) (cf. bibliothèques nationales)
- Approche sélective
 - Il s'agit de sélectionner des sites web en fonction de critères thématiques, qualitatifs...
- Approche thématique
 - Il s'agit d'une approche complémentaire de l'approche sélective, qui consiste à sélectionner une collection de sites web en fonction d'un événement particulier

Différentes stratégies de collecte

- Stratégie automatisée
 - mise en place d'un logiciel-robot (moissonneur et collecteur du web). Cette stratégie est utilisée dans le cadre d'une collecte d'un domaine complet
- Stratégie semi-automatisée
 - mise en place d'un logiciel-robot à laquelle on ajoute des critères de sélection plus précis. Cette stratégie est utilisée dans le cadre d'une approche sélective du web
- Approche manuelle
 - sélection manuelle de sites pertinents. Cette approche est utilisée dans le cadre d'une sélection thématique

Exemple – projet Archives Web Suisse

- Critères de collecte – sélection de sites
 - publiés et rendus accessibles par l'intermédiaire des réseaux de communication publics (World Wide Web)
 - pas protégés par des mots de passe et/ou non payants
 - ayant un rapport avec les cantons ou la Suisse ou contenant des thèmes significatifs ou importants d'un point de vue historique, politique, économique, social, scientifique, culturel ou religieux sur les cantons et la Suisse
 - rédigés par des auteurs suisses ou créés et conçus par des auteurs suisses ou en relation avec la Suisse
 - s'adressant à un large public
- site Web peut être hébergé sur un serveur situé en Suisse ou à l'étranger

Exemple – projet Archives Web Suisse

- Domaines de collecte
 - Sites web sur les cantons
 - sites Web représentant les cantons en tant que collectivité territoriale
 - sites Web représentant la Confédération
 - sites Web sur des événements et des collections spéciales (élections fédérales, Euro 2008)
- élaboration d'une grille d'évaluation
- principe du *fair use* au niveau de la collecte

Archives des réseaux sociaux

- Comment aborder cette question?

Merci!



Questions?



“Study School for Archival Educators” copyright International Council on Archives, 2019, is licensed under Creative Commons License Attribution-Noncommercial 4.0.