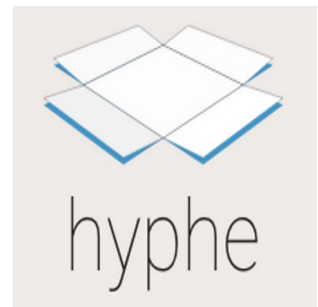


[LE LOGICIEL LIBRE DU MOIS]

Hyphe

[Hyphe](#) est un logiciel libre de crawler web permettant aux chercheurs de créer des corpus constitués de pages web et de liens entre elles sur un sujet spécifique (par exemple, l'huile de palme ou le coronavirus).



Ces pages web sont sélectionnées par les chercheurs et peuvent être regroupées en « webentités », qui peuvent être des pages individuelles, des sites web, des sous-domaines ou même une combinaison de ceux-ci. Ces entités représentent différents acteurs de la problématique à traiter (par exemple, une personne, une organisation, etc.).

En les crawlant, Hyphe permet de construire itérativement et de visualiser un réseau des relations entre ces acteurs à travers les hyperliens reliant ces webentités.

De nouvelles webentités sont automatiquement suggérées en explorant les hyperliens de chaque entité présente dans le corpus. Les chercheurs peuvent ensuite les examiner dans un processus itératif et qualitatif.

Hyphe permet aux chercheurs de choisir et de catégoriser manuellement les acteurs qu'ils souhaitent ajouter à leur corpus. À ce titre, il doit être considéré comme un outil qualitatif.

La cartographie des controverses peut largement bénéficier d'études sous l'angle du web. Analyser les sites internet des acteurs d'une controverse et établir la carte des liens entre eux peut constituer un objet d'étude des plus riches, mais qui peut s'avérer quelque peu complexe à réaliser pour des chercheurs en sciences humaines et sociales. Publié sous la forme d'un logiciel libre publié sur GitHub, Hyphe a été conçu pour offrir aux chercheurs et étudiants un outil de création et nettoyage de corpus web reposant sur un crawler orienté pour la recherche. Les utilisateurs sont accompagnés par une méthodologie pour construire leur corpus web de manière à la fois granulaire et flexible avec des principes de curation simples. Plutôt que de traiter de « sites web »,

Hyphe manipule des « Web Entités », qui peuvent être définies tout aussi bien comme de simples pages, un sous-domaine, une combinaison de sites, etc. Les pages résidant sous ces Web Entités sont ensuite crawlées, afin de collecter les liens sortants et les contenus textuels. Les entités les plus citées peuvent ensuite être prospectées afin d'enrichir le corpus avant de le visualiser sous la forme d'un réseau et de l'exporter pour nettoyage dans Gephi ou pour publication. Nous publions aujourd'hui une nouvelle version de Hyphe désormais capable de gérer plusieurs corpus simultanément et reposant sur une toute nouvelle interface web pour les utilisateurs implémentée en HTML5. Découvrez Hyphe sur son site dédié, essayez notre démo ou installez-le sur un serveur de votre labo !