
La constitution d'un corpus spécialisé à partir des ressources ISTEEX

Sabine Barreaux*¹ and Camille De Salabert*²

¹Institut de l'information scientifique et technique (Inist) – CNRS : UPS76 – France

²Institut de l'information scientifique et technique (Inist) – CNRS : UPS76 – France

Résumé

Présentation de l'atelier

La plateforme ISTEEX constitue un réservoir riche de plus de 23 millions de publications scientifiques accessibles à la communauté française de l'enseignement supérieur et de la recherche.

Par sa mise à disposition des documents en texte intégral, cette vaste ressource textuelle représente un matériau de choix pour réaliser des travaux de fouille de textes à des fins de recherche scientifique.

En outre, elle est associée à toute une gamme de services conçus autour de l'API ISTEEX afin de simplifier et de développer son exploitation, notamment par le téléchargement de corpus volumineux (ISTEX-DL) et par l'exploration des corpus extraits permettant leur exposition sur le web (LODEX).

Cet atelier se propose de présenter la méthode de constitution d'un corpus spécialisé sur un sujet universel en utilisant les différents services associés à ISTEEX afin d'obtenir à un corpus de qualité qui pourra ensuite être utilisé par des outils de fouille de texte.

Programme détaillé

Constitution d'un corpus spécialisé à partir des ressources ISTEEX

- Présentation du réservoir ISTEEX
- Construction d'une requête avec ISTEEX-DEMO

Valorisation d'un corpus spécialisé à l'aide des services ISTEEX

- Téléchargement du corpus avec ISTEEX-DL
- Exploration du corpus avec l'outil LODEX pour analyser son contenu et en éliminer les éventuelles anomalies

En savoir plus

Tutoriel vidéo API ISTEEX

Tutoriel vidéo LODEX

*Intervenant