

---

# La constitution d'un corpus spécialisé à partir des ressources ISTE<sup>X</sup>

Sabine Barreaux\*<sup>1</sup> and Camille De Salabert\*<sup>1</sup>

<sup>1</sup>Institut de l'information scientifique et technique (Inist) – Centre National de la Recherche Scientifique  
- CNRS – France

## Résumé

### Présentation de la conférence

La plateforme ISTE<sup>X</sup> 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 ISTE<sup>X</sup> afin de simplifier et de développer son exploitation, notamment par le téléchargement de corpus volumineux (ISTE<sup>X</sup>-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 ISTE<sup>X</sup> afin d'obtenir à un corpus de qualité qui pourra ensuite être utilisé par des outils de fouille de texte.

### Programme de l'intervention

Constitution d'un corpus spécialisé à partir des ressources ISTE<sup>X</sup>

- Présentation du réservoir ISTE<sup>X</sup>
- Construction d'une requête avec ISTE<sup>X</sup>-DEMO

Valorisation d'un corpus spécialisé à l'aide des services ISTE<sup>X</sup>

- Téléchargement du corpus avec ISTE<sup>X</sup>-DL
- Exploration du corpus avec l'outil LODEX pour analyser son contenu et en éliminer les éventuelles anomalies

### En savoir plus

Tutoriel vidéo API ISTE<sup>X</sup>  
Tutoriel vidéo LODEX

---

\*Intervenant