

ISTEX

L'excellence documentaire pour tous

Ecosystème d'ISTEX : Chantiers d'usage et services à valeur ajoutée

Jean-Marie Pierrel, Université de Lorraine

L'écosystème d'ISTEX

- En liaison étroite avec l'équipe de développement de la plateforme, mise en place de projets d'**exploitation du plein texte pour**
 1. **Démontrer l'intérêt d'ISTEX dans une optique de Text Mining**
 1. Projets de services à valeur ajoutée
 2. Chantiers d'usage
 2. **Créer une dynamique de recherche développement** autour de la plateforme ISTEX qui puisse servir de déclencheur à des activités plus larges d'appropriation par les chercheurs des contenus d'ISTEX pour développer des recherches en fouille de textes

Trois grands types de projets

- **Enrichissement des métadonnées** grâce à des techniques de fouille de textes
- Mise en place de **projets de services à valeur ajoutée**
- Mise en place de projets **chantiers d'usage**

=> Interactions avec l'équipe de développement de la plateforme au cours de 3 séminaires techniques (2015, 2016, 2017)

Enrichissements des métadonnées fondés sur des techniques de TDM

- **Détection d'entités nommées (LDI Tours)**
 - Noms de personnes, lieux, organisations, financeurs, hébergeur, date, adresse web, etc.
- **Détection en plein texte de termes et de leurs variantes (LINA Nantes)**
- **Intégration d'outils et enrichissement complémentaires (INIST)**
 - **XMLisation** de documents
 - **Balisage** des principaux champs des références bibliographiques

Des projets de services à valeur ajoutée

- **CILLEX : classification et outils de métrologie** greffés sur le moteur de recherche ISTEEX (CLLE/IRIT Toulouse)
- **LorExplor : Bibliothèque open source de composants Xml** d'exploitation de sous-corpus ISTEEX (Université de Lorraine)
- **ISTEX_R : Caractérisation de l'évolution des recherches** et des connaissances dans le temps grâce à la construction de cartes diachroniques (LORIA, ATILF, INIST)

Des expérimentations grâce à des chantiers d'usage

- Sélection de **8 équipes-projets** « Chantier d'usage » en décembre 2016 pour démontrer les possibilités de fouille de textes sur les corpus ISTEEX.
- Les **corpus ISTEEX vus comme des données de base de recherches** de type fouille de textes ou «text mining».

Les projets Chantiers d'usage en cours

- **ALPAGE** : annotation de corpus ISTEEX et codage en TEI
- **ReITEX** : Usage du corpus ISTEEX pour l'exploitation de méthodes d'extraction de connaissances à partir de textes
- **3ST** : Surligneur Sémantique de Textes Scientifiques

Les projets Chantiers d'usage en cours

- **FULLLAB** : Comparaison fulltext et abstract en sciences environnementales
- **NEOTEX** : Exploitation de documents textuels d'un domaine par un néophyte
- **Terre-ISTEX** : Identification et analyse des TERRains d'Études dans les corpus ISTEX

Les projets Chantiers d'usage en cours

- **Biosystémique** : Recherche de résultats expérimentaux dans les publications scientifiques concernant la biologie systémique
- **OTELO** : chantier thématique sur les Terres rares, de l'exploration à l'impact environnemental

Pour la suite : choix de présentation de quelques résultats obtenus dans ce cadre

- **CILLEX : Classification et identification de l'information pertinente dans ISTEEX**

Yannick Chudy – IRIT / Univ.Toulouse et Bruno Gaume – CLLE / Univ.Toulouse

- **Terre-ISTEX : Identification et analyse des terrains d'étude dans les corpus ISTEEX**

Éric Kergosien – GERiICO / Université Lille 3

Pour la suite : choix de présentation de quelques résultats obtenus dans ce cadre

- **OTELo / Terres-rares : un outil cartographique d'exploration et exploitation des corpus ISTEEX**

Corinne Leyval et Pierre-Yves Arnould – Université de Lorraine

- **LorExplor : Une bibliothèque open source de composants XML d'exploitation du corpus**

Jacques Ducloy – LORIA

Des exploitations du plein texte dans d'autres projets partenaires

- Au sein de l'**Institut des Systèmes Complexes** du CNRS
 - Exploitation via **Gargantext**
 - Projet **EasISTEX**
- **ISTEX, source de corpus** dans le cadre de la plateforme européenne **OpenMinTeD**
-



Merci de votre attention.