


## Quand Wicri rencontre ISTEEX autour de la didactique...

*Jacques Ducloy  
Ingénieur conseil*



---

---

---

---

---

---

---

---

### Réseaux de wikis sémantiques et exploration de corpus, Application à H<sup>2</sup>PTM

- ▶ 1 - Introduction
- ▶ 2 - Wicri, un réseau de wikis sémantiques
- ▶ 3 - ISTEEX, des corpus pour la recherche et l'enseignement
- ▶ 4 - Exemple en didactique
- ▶ 5 - Conclusion : les nouveaux usages

IDEXI 2013 29/11/2013 2

---

---

---

---

---

---

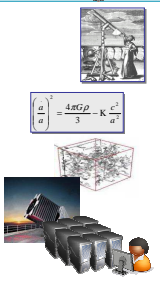
---

---

## Emergence d'un nouveau Paradigme ?

- ▶ Thousand years ago - **Experimental Science**
  - Description of natural phenomena
- ▶ Last few hundred years - **Theoretical Science**
  - Newton's Laws, Maxwell's Equations...
- ▶ Last few decades - **Computational Science**
  - Simulation of complex phenomena
- ▶ Today - **eScience or Data-centric Science**
  - Unify theory, experiment, and simulation
  - Using data exploration and data mining
    - Data captured by instruments
    - Data generated by simulations
    - Data generated by sensor networks

▶ Scientists over-welmed with data  
▶ Computer Science and IT companies have technologies that will help  
(With thanks to Jim Gray)



... Merci à Tony Hey (Microsoft Research)

IDEXI 2013 29/11/2013 3

---

---

---

---

---

---

---

---

## Nouveau paradigme, Web 2.0 la *success story* : Wikipédia

- ▶ 20 000 000 pages
  - la technologie est au point !!!
  - Modération a posteriori - vs - validation a priori.
- ▶ Un site massivement consulté par :
  - les chercheurs dans une démarche interdisciplinaire,
  - des industriels à la recherche de partenaires.
- ▶ Une collaboration de milliers de volontaires,
  - noyau de centaines d'experts eux aussi volontaires.

IDEXI 2013 29/11/2013

4

---

---

---

---

---

---

---

---

## Au fait, c'est quoi un wiki ???

- En 2000 :
- ▶ Logiciel facile à mettre en œuvre pour une rédaction collaborative
- 2 axes :
1. facilité de mise en œuvre dans les cas simples
  2. Aller le plus loin possible dans la construction de documents complexes.
- Pour la recherche, le transfert de technologie et de la culture scientifique :
- ▶ Système de construction collaborative de connaissance scientifique, culturelle et technique
  - ▶ Opéré par des groupes possédant une bonne maîtrise technologique
  - ▶ Modéré, et non contrôlé a priori

IDEXI 2013 29/11/2013

5

---

---

---

---

---

---

---

---

## Exemple de traitement scientifique : géolocalisation

On superpose des  
images sur un  
fond de carte

```
[[Début de carte]]
[[Image:474x488-Carte-Lorraine-A.png|250px]]
{{Géolocalisation
|Lorraine
|49.1203
|6.1778
|Metz
|Lac
|se}}
[[Fin de carte]]
```



[http://beta.cadecancy.lorcem.net/index.php/index.php?cat=5&file=se\\_g1C75&id=6&id=6](http://beta.cadecancy.lorcem.net/index.php/index.php?cat=5&file=se_g1C75&id=6&id=6)

IDEXI 2013 29/11/2013

6

---

---

---

---

---

---

---

---

## Le réseau WICRI > 100 wikis



- ▶ Wikis communs :
  - ▶ 7 wikis généraux ou service
  - ▶ 10 familles thématiques
  - ▶ 16 familles géographiques

Une équipe de soutien

30 wikis institutionnels

IDEXI 2013 29/11/2013 7

---

---

---

---

---

---

---

---

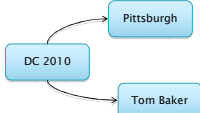
---

---

## Liens simples/ sémantiques

### Liens simples

DC 2010 takes place in [[Pittsburgh]]  
 ==Program Committee==  
 \* [[Thomas Baker]]



- ▶ A chaque lien on associe
  - + un attribut (property)
- ▶ Extension Semantic MediaWiki

DC 2010 takes place in [[Has location city:Pittsburgh]]  
 ==Program Committee==  
 \* [[Has PC member:Thomas Baker]]

IDEXI 2013 29/11/2013 8

---

---

---

---

---

---

---


---

---

---


## Liens sémantiques

**Pittsburgh** est située au confluent des rivières **Allegheny** et **Monongahela** qui forment l'**Ohio**, un affluent du **Mississippi**.



Pittsburgh est située au confluent des rivières [[sur le cours d'eau::Allegheny (rivière)|Allegheny]] et [[sur le cours d'eau::Monongahela]] qui forment l'[[sur le cours d'eau::Ohio (rivière)|Ohio]], un affluent du [[Mississippi (fleuve)|Mississippi]].

Faits relatifs à Pittsburgh ⓘ — Recherche de pages similaires avec + 🔍

Voir comme RDF 

Sur le cours d'eau Allegheny (rivière) + 🔍, Monongahela + 🔍 et Ohio (rivière) + 🔍

IDEXI 2013 29/11/2013 9

---

---

---

---

---

---

---

---

---

---

# L'Ohio

==Les affluents de l'Ohio==  
 ("liste calculée")

```

      {{#ask: [[est un affluent: {{PAGENAME}}]]
      |format=ul
      |sep=-
      |intro=Rivières citées sur Wicri Eau ... }}
      
```

==Les villes traversées par l'Ohio==  
 ("liste calculée")

```

      {{#ask: [[sur le cours
      d'eau: {{PAGENAME}}]]
      |format=ul
      |sep=-
      |intro=Villes citées sur Wicri Eau ... }}
      
```

IDEXI 2013 29/11/2013 10

---

---

---

---

---

---

---

---

---

---

---

---

## Nouveau paradigme, Web 3.0 les wikis sémantiques

- ▶ Wikipédia
  - ▶ = du texte + des liens simples
- ▶ Wikis sémantiques
  - ▶ = du texte + des liens qualifiés
  - ▶ Meilleure navigation
  - ▶ Raisonnement
  - ▶ Continuum entre le texte et la base de données
- ▶ Wicri : réseau de wikis sémantiques...
- ▶ La didactique peut s'exprimer dans différents contextes...

IDEXI 2013 29/11/2013 11

---

---

---

---

---

---

---

---

---

---

---

---

## Istex, contexte national

- ▶ 60 millions € ;
  - 60 millions de documents en texte intégral
- ▶ Décision nationale et exceptionnelle
  - pour un accès « sans contraintes » aux corpus par les chercheurs des établissements français
    - (Jean-Pierre Finance)
- ▶ Originalité du projet Istex - vs - Allemagne
  - Accès aux sources et pas seulement aux PDF
    - L'accès au document n'est pas l'objectif prioritaire
    - Priorité : nouveaux usages de corpus pour la recherche

IDEXI 2013 29/11/2013 12

---

---

---

---

---

---

---

---

---

---

---

---

### ISTEX : maîtriser l'infodiversité...

- ▶ Gisements gigantesques par leur volume et leur hétérogénéité
  - structurelle (formats), vocabulaires, thématiques...
- ▶ Très grande diversité des usages potentiels :  
Exemples :
  - Réponse dans les 3 heures/demande ministérielle (DRRT)
    - ou - construire un observatoire thématique, régional...
  - Open data - ou - tout confidentiel
- ▶ Les traitements systématiques ne sont plus possibles => il faut gérer le spécifique.
- ▶ Une requête en bibliothèque
  - Avant : 2 à 3 documents « a lire tranquillement »
  - Demain : 1000 docs en texte intégral « à explorer »

---

---

---

---

---

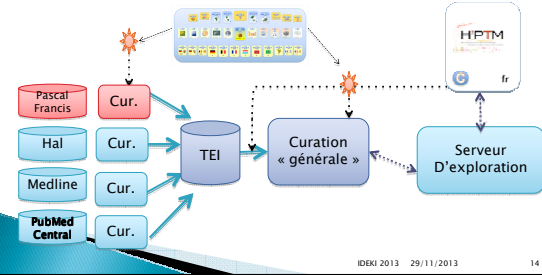
---

---

---

### Wicri - Istex

- ▶ couplage avec des plateformes de curation et d'exploration de corpus



---

---

---

---

---

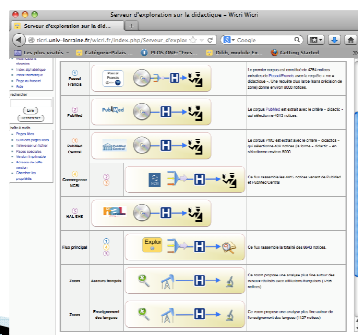
---

---

---

### Exemple sur la didactique

9 000 ref. biblio.  
Dont 400 oc Xml  
20 000 auteurs  
86 pays



---

---

---

---

---

---

---

---

## Curation de données, exemples

- ▶ Trouver et homogénéiser les entités géographiques
  - France, FRA, FR, Paris, INRIA
  - Royaume-Uni, UK, GBR, London...
  - Lorraine, Nancy, Metz, UHP
  - Etc etc

---

---

---

---

---

---

---

---

## Curation de données, exemples sur les auteurs

serveur d'exploration sur la didactique

Attention, ce site est en cours de développement !  
Attention, une partie des données n'est pas encore mise à jour.

Corpus

- Author's unit id: Levine D
- Total number of names: 4
- Total number of records with affiliations: 9

Publications

ID	Author	Count
D' Levine	NONE (181819)	1
DM Levine	NONE (181819)	1

Auteurs

Name	Levine	Count
David Levine	NONE (181819)	1

---

---

---

---

---

---

---

---

## Modes d'interaction

- ▶ Lecteur final : Wiki / rapport
- ▶ Lecteur averti : Wiki + exploration
- ▶ Rédacteur : exploration (+analyses) -> Wiki
- ▶ Rédacteur averti, documentaliste, etc
  - Procède aux extractions de corpus,
  - Effectue des paramétrages (génération automatique)
  - Lance les générations (mode itératif)
  - Complète des tables de curation
  - Peut utiliser des filtres Unix pour des raffinements non prévus ou non encore implantés.
- ▶ Chercheur :
  - Interopérabilité => peut intervenir n'importe où...

---

---

---

---

---

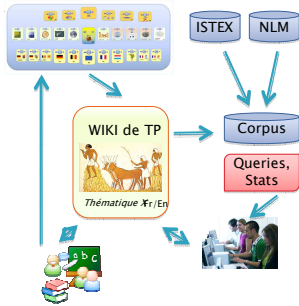
---

---

---

### Démarche Ticri pour la formation des ingénieurs : Veille + restitution collective dans un espace structuré

- Un wiki (ou plusieurs wikis) de référence dans le réseau Wicri traitent d'un sujet à approfondir,
- En parallèle on extrait des corpus d'informations (NIST, Istex et/ou autres bases telles que Medline (NLM) de taille significative (exemple 5000 à 20000 articles))
- On extrait du wiki de référence un sous-ensemble thématique à partir duquel on extrait un wiki de TP
- Les étudiants doivent y contribuer, de façon collective, en y écrivant de nouveaux articles en utilisant notamment les corpus bibliographiques via des outils d'analyse (par exemple un outil statistique autour de BaseX, outils de recherche avancé dans des corpus en Xml).
- En fin de travail, l'enseignant peut intégrer les informations pertinentes aux wikis de référence.




---

---

---

---

---

---

---

---

### Conclusion

- ▶ Faut-il repenser les programmes dès le lycée ???
  - Au même titre que : latin, nombres complexes, chargement d'une batterie électrique...
  - Ingénierie de l'information numérique à l'heure de l'internet pour mieux comprendre le monde informationnel omniprésent...
- ▶ Merci de votre attention...
- ▶ Pour retrouver le serveur, sous Google :
  - ▶ "Wicri" didactique

---

---

---

---

---

---

---

---