



Analyser et écrire la science multidisciplinaire,

dans un réseau d'hypertextes sémantiques avec Dilib et Wicri





1977
Pascal
Mistrall
Iris 80
Cyclades



1991 Pascal en SGML/XML

- ▶ *« L'ingénieur, le physicien se trouvent souvent devant les problèmes que les mathématiciens classiques n'ont pas pu résoudre. [...] Il est alors indispensable que l'ingénieur, le physicien et tous ceux qui s'occupent de mathématiques appliquées, soient capables de se dégager du complexe inhibitif de rigueur que leur a imposé leur éducation, et qu'ils osent se lancer à l'aventure : la vérification expérimentale sera là pour leur crier casse-cou le cas échéant. »*

C'est la même chose avec la science basée sur l'exploration des connaissances

L'information scientifique en France

- ▶ 1665 – Journal des Sçavants
- ▶ 1746 – Encyclopédie Diderot et d'Alembert
 - 150 collaborateurs, 35 volumes
- ▶ 1750 – Histoire naturelle Buffon
 - 20 principaux collaborateurs; 44 volumes (dont post-mortem)
 - Œuvre composite (inventaire, articles de synthèse, comparaisons anatomiques...)
- ▶ 1960 – le Trésor de la Langue Française,
 - Dictionnaire réalisé avec des mécanismes d'Intelligence Artificielle (groupes binaires)
 - 80 ingénieurs sur 40 ans
- ▶ 1971 – Pascal => Pascaline 500.000 références par an => Cyclades/Mistral
 - 1988 : CDST + CDSH => Inist + coopérations : plus de 500 agents
 - 1993 -> 2015 démantèlement progressif (TLF+Pascal+Francis)
- ▶ 2024 – Personnels affectés à la production d'information scientifique pour l'IA :
 - **ZERO**

Dans le Monde (chiffres très très partiels)

▶ USA : National Library of Medicine

- 1990 – même profil que l'INIST : 500.000 analyses /an (400 personnes)
- 2024 – 750 personnes , 4000 contacts, 1.000.000 analyses par an !!!
 - Chaque notice = indexation assistée + 3 interventions humaines

▶ USA : Wikimedia Foundation (Wikipédia + ...)

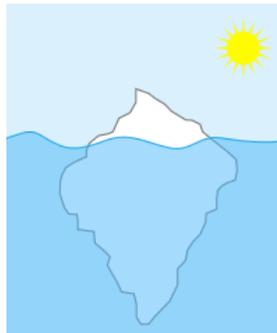
- 750 permanents à San Francisco, CA : 225 millions de dollars
- filiale commerciale => Google

▶ USA :

- OCLC (WorldCat) : 1.200 p.
- ACS (American Chemical Society) : 2.000 p. // ACM ? // IEEE 1.000
- Clarivate (WoS) 11.600 p.
- Total entre 15.000 p. et 25.000 p.

▶ Allemagne : FIZ : 330 personnes + ???

▶ Russie : VINITI : 800 personnes, 12.000.000 analyses par/an...



Démonstrateur Wicri : Ecrire la science vivante en numérique

Chaque discipline peut développer son approche



Démonstrateur : 150 wikis

Un wiki par grand sujet

Activité et édition scientifique

- Contributions transparentes (pas d'anonymat)
- Pilotage (ligne éditoriale) par comités scientifiques

Contenus diversifiés

- Articles de recherche en hypertexte
- Ouvrages anciens réédités en données de la recherche.
- Extraits de dictionnaires, encyclopédie réédités
- Articles encyclopédiques
- Ontologies liées aux données
- Références bibliographiques sélectionnées (Pascal..)

Un réseau

- Mutualisation des modèles, modules, catégorie
- Notion de wiki de référence
- Maintenance assistée par des robots



Moteur de Wikipédia...
Commons 150.000.00 pages
Modèles, modules...
Catégories
API ouverte...



Démonstration Wicri/Musique

- Partitions musicales
- Interaction de dictionnaires

Structurer la connaissance avec des liens sémantiques dans les modèles...



sur un livret de
[[A pour auteur de livret::Lorenzo da Ponte]].



Semantic MediaWiki :
Constructions sémantiques
dans une approche RDF
Organisation SGBD
incrémentale

A pour auteur de livret



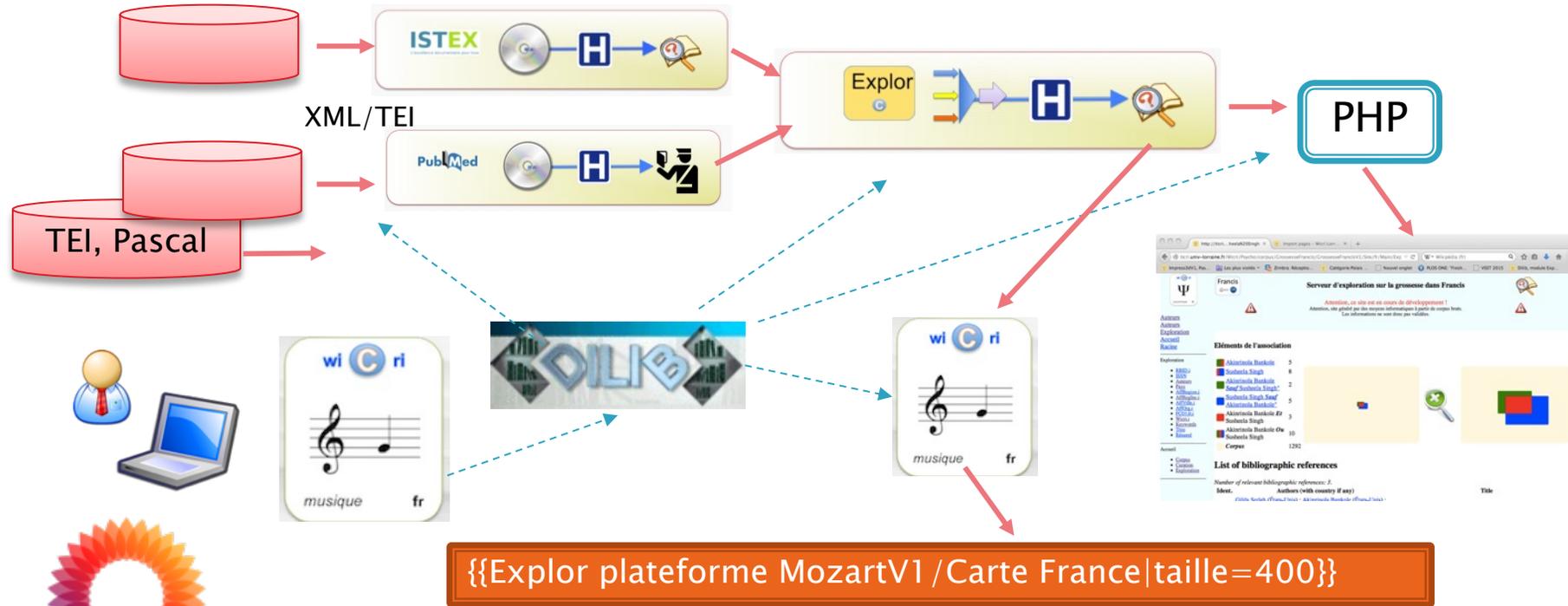
Liste des opéras ayant un livret de Lorenzo da Ponte

Sur ce wiki (par génération automatique)

- Axur, re d'Ormus (Antonio Salieri)
- Così fan tutte (Wolfgang Amadeus Mozart)

```
{{#ask: [[a pour auteur de livret::{{PAGENAME}}]]  
| format=ul | ?A pour compositeur=compositeur :
```

DILIB – ISTEK – 150 serveurs d'exploration



```
<idx>
  <k>Salieri</k>
  <f>2</f>
  <l>
    <e>001230</e>
    <e>002345</e>
  </l>
</idx>
```



| Pays | Région | Villes |
|---------------------------------------|--|--------------------------------------|
| 1. France (67) 🔗 | 1. Californie (11) 🔗 | 1. Paris (9) 🔗 |
| 2. États-Unis (31) 🔗 | 2. Île-de-France (9) 🔗 | 2. Marseille (5) 🔗 |
| 3. Royaume-Uni (14) 🔗 | 3. Occitanie (région administrative) (7) 🔗 | 3. Montpellier (4) 🔗 |
| 4. Allemagne (14) 🔗 | 4. Massachusetts (6) 🔗 | 4. Londres (4) 🔗 |
| 5. Canada (11) 🔗 | 5. Angleterre (6) 🔗 | 5. Grenoble (4) 🔗 |
| 6. Italie (10) 🔗 | 6. État de New York (5) 🔗 | 6. Berlin (4) 🔗 |
| 7. Espagne (8) 🔗 | 7. Maryland (5) 🔗 | 7. Toulouse (3) 🔗 |
| 8. Suisse (6) 🔗 | 8. Caroline du Nord (5) 🔗 | 8. Prague (3) 🔗 |
| 9. Australie (6) 🔗 | 9. Arizona (5) 🔗 | 9. Montréal (3) 🔗 |
| 10. Pays-Bas (6) 🔗 | 10. Washington (État) (4) 🔗 | 10. Zurich (2) 🔗 |

| Mots-clés anglais | Mots des titres | ISSN/revue |
|---|--------------------------------------|--|
| 1. Astrophysics (3) 🔗 | 1. data (10) 🔗 | 1. SPIE proceedings series (6) 🔗 |
| 2. State of the art (2) 🔗 | 2. analysis (7) 🔗 | 2. 1932-6203 (5) 🔗 |
| 3. Software package (2) 🔗 | 3. software (6) 🔗 | 3. Lecture Notes in Computer Science (4) 🔗 |
| 4. Real time (2) 🔗 | 4. microbial (6) 🔗 | 4. Eos Trans. AGU (3) 🔗 |
| 5. Quebec (2) 🔗 | 5. marine (5) 🔗 | 5. 2324-6250 (3) 🔗 |
| 6. Perspective (2) 🔗 | 6. genome (5) 🔗 | 6. 1051-5420 (3) 🔗 |
| 7. Open source software (2) 🔗 | 7. distributed (5) 🔗 | 7. 0096-3041 (3) 🔗 |
| 8. Measurement sensor (2) 🔗 | 8. genomic (4) 🔗 | 8. 0027-8424 (3) 🔗 |
| 9. Library network (2) 🔗 | 9. control (4) 🔗 | 9. 2047-217X (2) 🔗 |
| 10. Information policy (2) 🔗 | 10. web (3) 🔗 | 10. 1545-7885 (2) 🔗 |



Démonstration COVID

- 25 serveurs d'exploration
- Procédure rapide sur PubMed

La Chanson de Roland un gigantesque hypertexte

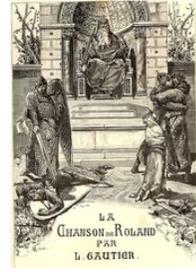


Feuillets



Laises

F.



Chapitres



Pages

L.

Notes

V.

Vers

L.

V.

Termes

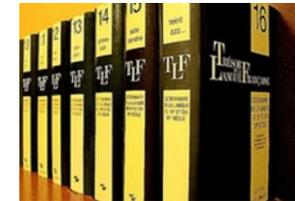
T.



Mouvements

Partitions
musicales

V.



Mots

Encyclopédie
De la Chanson de
Roland

Conclusion inspirée par Buffon...



Buffon : une Histoire naturelle aux multiples aspects

- Données de la recherche : Inventaire du Cabinet du Roi
- Articles anatomiques de Daubenton
- Articles descriptifs et scientifiques => pour le Roi (et ses sujets)
- Articles de synthèse (épistémologie...)
- Un homme des lumières : de Newton à Archimède via Plin l'Ancien

Wicri : 150 wikis gérés par un seul ETP permanent

- 41.000/210.000 pages, 610.000 interventions...

Objectif stratégique : refaire de la France un acteur mondial :

- 500 ingénieurs, profil siècle des lumières (excellence thématique et numérique IA)
- X fois 1000 wikis ???

Retombées multiples dans les SHS ou sciences guidées par l'observation

- Exemple : les Histoires naturelles en réseau hypertexte
- Voir les démonstrations

Merci aux équipes de l'INIST pour leur soutien logistique

Merci pour votre écoute, vos questions, vos visites...

