



Mutation des comportements pour maîtriser le numérique et l'IA

Hypertextes sémantiques, Unix, Xml, Dilib et Wicri



Mobilisation pour l'autonomie numérique des chercheurs, ingénieurs, praticiens

► Plan

- Quelques données sur l'information élaborée (France vs USA, etc)
- Valorisation des recherches en cours avec MediaWiki
- Exploration de corpus médicaux avec Dilib (IA, Unix, XML)
- Bibliothèques numériques avec approche encyclopédiques
 - Musique => Chanson de Roland
- Mobilisation générale avec :
 - l'Histoire Naturelle
 - et l'Académie lorraine des Sciences

Mobilisation pour l'autonomie européenne sur l'information scientifique et technique

Information élaborée (chiffres très partiels)

▶ USA : National Library of Medicine

- 1990 – 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 + ...)

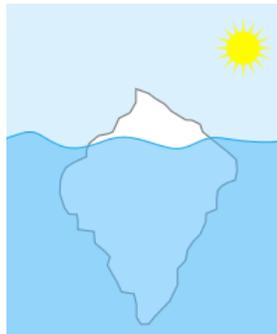
- 750 permanents à San Francisco, CA : 225 millions de dollars
- filiale commerciale => Google

▶ 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...



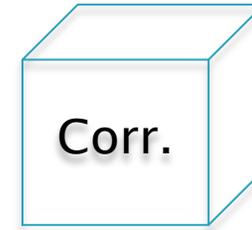
L'information scientifique en France

- ▶ 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)
- ▶ 1960 – le Trésor de la Langue Française, **Dynamique du Plan Calcul**
 - Dictionnaire réalisé avec des mécanismes d'Intelligence Artificielle 80 ingénieurs sur 40 ans
- ▶ 1971 – Pascal => Pascaline 500.000 références par an => Cyclades/Mistral
 - 1988 : CDST + CDSH => Inist + coopérations : de 500 à 1000 agents ou coopérants
 - 1993 -> 2015 démantèlement progressif (TLF+Pascal+Francis)
- ▶ 2024 – Personnels affectés à la production d'information scientifique pour l'IA : **ZERO**

Comment remonter la pente...

Nouveaux comportements individuels, collectifs...

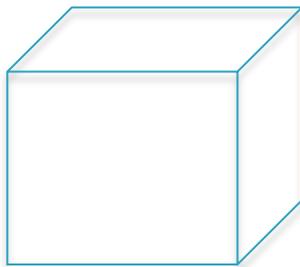
TLF et groupes binaires (pour la mémoire)



Texte corrigé

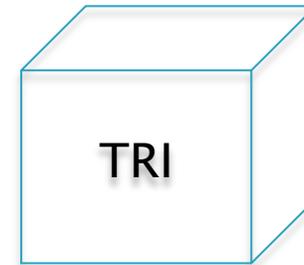


Lemmatisation



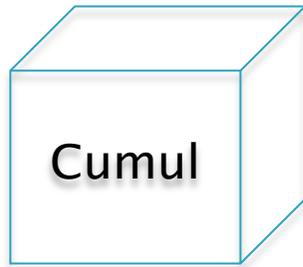
Livres

Mots, identification livre



Mots triés

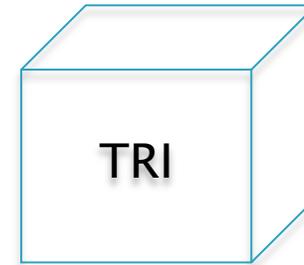
Groupes binaires (suite)



Mots triés
Abaca, T21
Abaca, T55
Abat, T3
....

Couples
Fréquence - mot

2 abaca (2,5)
5 abat (3 ...)
...



Couples triés

...
5 abat (3 ...)
...
2 abaca (2,5)

ETC durée 1 mois



Faire et écrire la science vivante en numérique

Valoriser le patrimoine écrit, scientifique ou culturel

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 assistable par des robots



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



Valoriser les recherches en cours



MediaWiki



- ▶ Publications, documents réédités en hypertexte
- ▶ Modélisation des comités de programme
- ▶ Inventaires avec données numériques consolidées

Lorraine

Contrat de Plan Etat Région

1 wiki par projet (environnement)

Maquette Grande Région

Lorraine, Luxembourg, Wallonie

Sarre, Rhénanie-Palatinat

Réseau de wikis multilingue (FR, EN DE)

TIC, SIC

Colloques du DCMI (Dublin Core)

⇒ Publications internationales

Colloques CIDE H²PTM

Revue : les Mots de l'Agronomie (INRAE)

Relations IHEST avec le Brésil

Couple bilingue (FR, PT)

Recherches en cours, bilan et moyens

- ▶ Ressources humaines (Projet Ticri => Wicri'2008)
 - 1 permanent CNRS, expert en numérique
 - 1 généraliste ETP sur contrat,
 - Informaticiens à temps partiel au démarrage (postdoc...)
- ▶ Peu de développements, forte expertise dans l'accompagnement
 - Exemples
 - Rodage Semantic MediaWiki ex nihilo
 - installation LaTeX compatible MediaWiki sur Ubuntu
- ▶ Quelques problèmes pour ce type d'application
 - Implication indispensable des unités de communication des établissements
 - Exemple : la page « Université de Lorraine »
 - Garantie de pérennité
 - Sinon l'information devient obsolète (et fausse...)

Explorations de corpus en santé

► Sources d'inspirations

- le TLF et les moteurs de recherche (1975)
- Unix (1985), dopé par Xml (1992), pour la manipulation des corpus
- Boîte à outils DILIB (petites sélections --- > serveurs d'exploration IA)



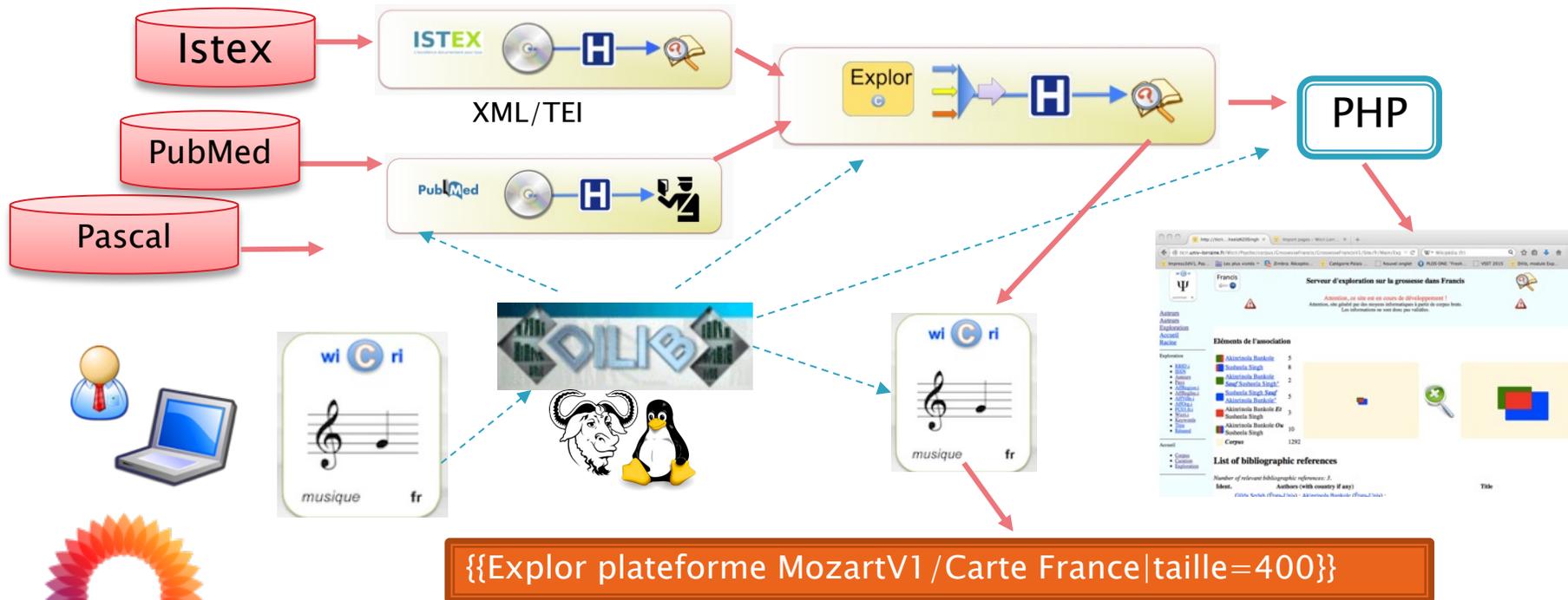
```
cat myIndex.xml \  
| SxmlSelect -g idx/f# -g idx/k# -p @g1 -p @g2 \  
| sort -n
```

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

◦ Problèmes initiaux

- Développements conséquents pour les interfaces avec les serveurs d'exploration
- Formation utilisateur imposant un démarrage « Unix – Langage C »

Solution : couplage Boite à outil XML et MediaWiki DILIB – ISTEK – 150 serveurs d'exploration



`{{Explor plateforme MozartV1 / Carte France|taille=400}}`



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 (5) 🔗	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. Québec (2) 🔗	5. marine (5) 🔗	5. 2324-6250 (3) 🔗
6. Perspective (2) 🔗	6. genome (5) 🔗	6. 1051-5400 (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
- Traduction française du MeSH

Besoins d'expertises, exemples

- ▶ Saturation du site par les robots
 - Les wikis représentent 200.000 pages
 - Les serveurs d'explorations représentent 600.000 documents générant des milliers de pages d'exploration
 - Les programmes PHP des serveurs génèrent des warnings quiaturent les fichiers log
- ▶ Conception, développement maintenance DILIB (ou MediaWiki)
- ▶ Robots Xml pour les évolutions des wikis

De la santé à la musique

- ▶ Problème du développeur :
 - Impossibilité d'évaluer un serveur Santé sans être médecin
 - => création d'un wiki dédié à la musique
 - avec des serveurs d'exploration évaluables (bruitage du serveur Mozart)



Démonstrateur Wicri/Musique

- Partitions musicales
- Articles de musicologie
- Interaction de dictionnaires
- Serveurs d'exploration

```
{  
  \time 2/4  
  \clef bass  
  c4 c g g a a g2  
}
```

Les commandes sont précédées d'un \
Les lettres représentent les notes
Les chiffres annoncent la durée



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

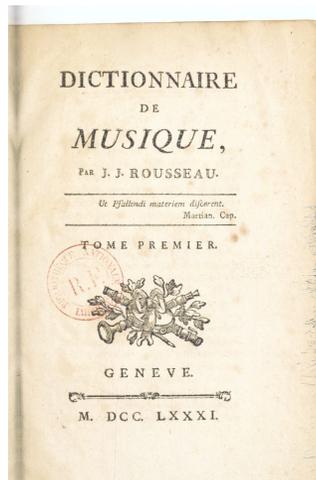
Sur ce wiki (par génération automatique)

- 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 :
```

Réédition numérique hypertexte Dictionnaire de musique (J.-J. Rousseau)

Romilly

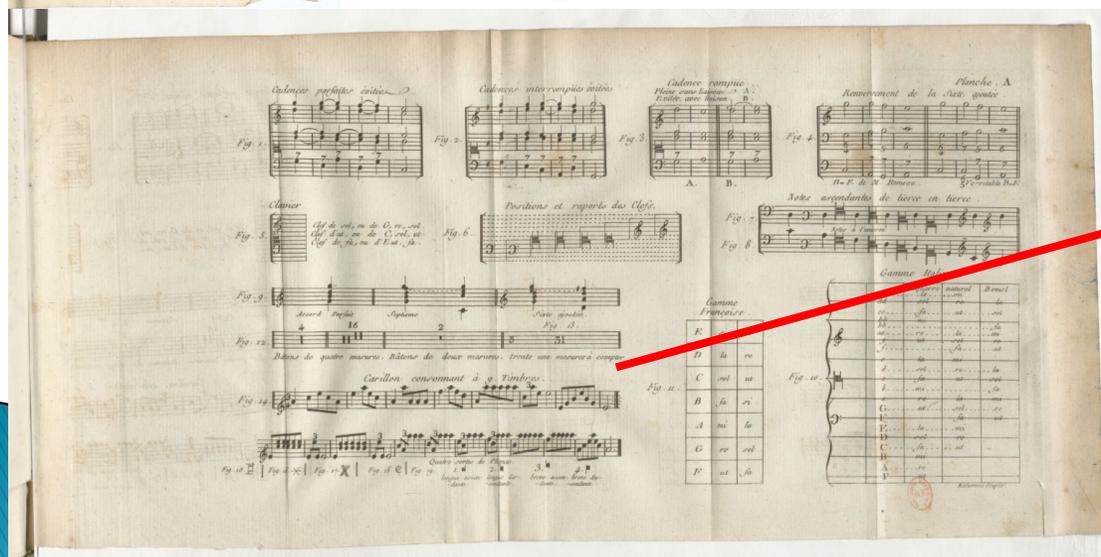


Dans le tome 1

pays des vaines spéculations.
CARRILLON. Sorte d'Air fait pour être exécuté par plusieurs Cloches accordées à différens Tons. Comme on fait plutôt le Carrillon pour les Cloches que les Cloches pour le Carrillon, l'on n'y fait entrer qu'autant de Sons divers qu'il y a de Cloches. Il faut observer de plus, que tous leurs Sons ayant quelque permanence, chacun de ceux qu'on frappe doit faire Harmonie avec celui

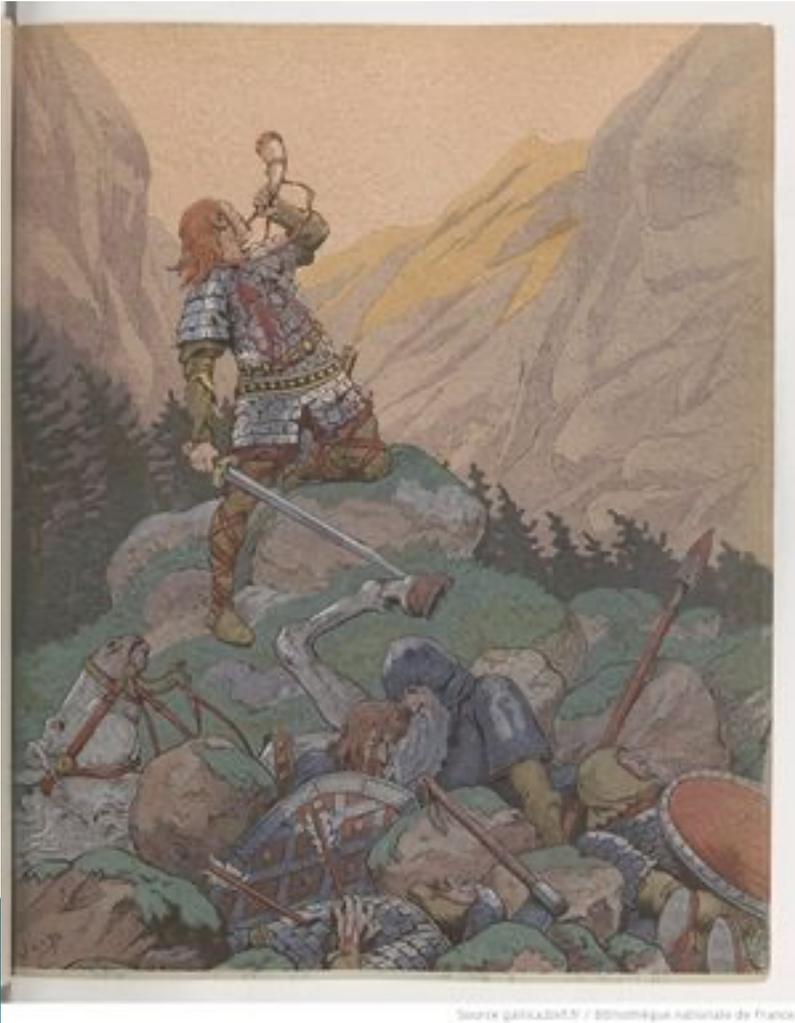
Lien

Dans le tome 2



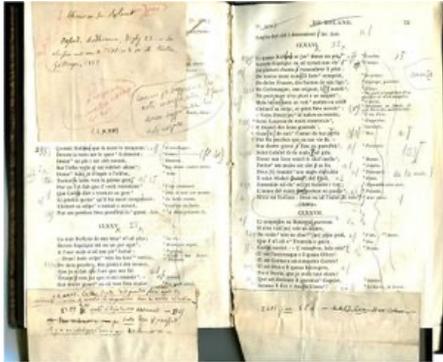
Le lecteur « non solfégiste » peut écouter la musique

La Chanson de Roland, une histoire, une épopée

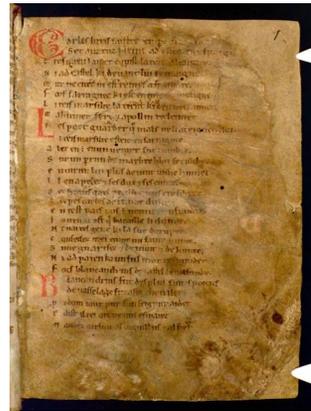


- ▶ Le 14 août 778, Charlemagne, de retour d'Espagne, perd l'arrière garde de son armée au col de Roncevaux
 - Ainsi est mort Roland
- ▶ La légende
 - la trahison de Ganelon
 - L'attitude héroïque de Roland
 - Qui attend de mourir pour sonner de l'olifant
 - Qui brise le rocher avec Durandal son épée
- ▶ La littérature
 - Manuscrits à partir du XIIe siècle
 - Transcriptions, traductions au XIX siècle
 - Des dizaines de milliers d'articles
 - De multiples valorisations (opéras, films...)

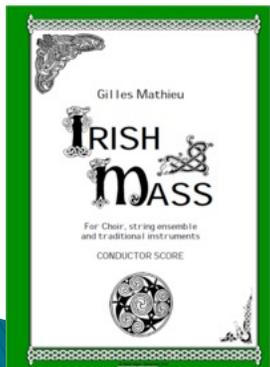
Un sujet de stage... 😊 imaginé tranquille !



Chanson de Roland
Edition critique de Francisque Michel
Annoté par Paul Meyer

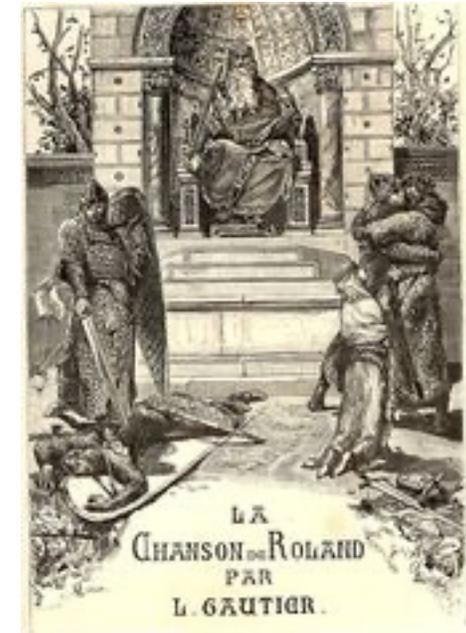


Manuscrit d'Oxford,
accessible sur Wikipédia



Gilles Mathieu
Un oratorio sur la Chanson de Roland

Un invité imprévu
Léon Gautier...



La Chanson de Roland un gigantesque hypertexte



Feuillets



Laises

F.

L.

V.

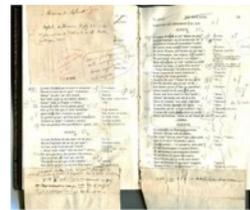
T.

Vers

Termes



Chapitres



L.

V.

Pages

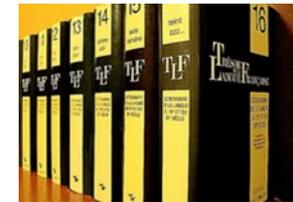
Notes



Mouvements

Partitions
musicales

V.



Mots

Encyclopédie
De la Chanson de
Roland

Résultats significatifs en 3 mois !

De la chanson à l'ALS en passant par Buffon

▶ Chanson de Roland

- Thème trop spécialisé pour des formations,
- Intérêt assez faible pour mobiliser des politiques,
- Plusieurs hommes années pour une référence internationale,
- Un projet potentiel, par exemple pour/avec l'Ecole des Chartes,

▶ L'Académie Lorraine des sciences

- Créée à Strasbourg (Muséum d'Histoire naturelle) en 1820
- Deux siècles de littérature scientifique très hétérogène
- Politiquement porteur dans le Grand Est

▶ Buffon et l'Histoire naturelle des animaux

- L'Histoire naturelle : Buffon et 25 siècles de littérature
- En relation avec la biodiversité, sujet porteur
- Une mine d'exemple pour les formations

Buffon, Histoire naturelle, Biodiversité



Buffon : une Histoire naturelle aux multiples aspects

- 44 volumes, multiples suites, multiples éditions
 - 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...)
 - Citations : de Newton à Archimède via Pline l'Ancien, Conrad Gessner...

▶ Un grand projet numérique potentiel

- De très nombreux ouvrages en français sur le site BHL (Washington)

▶ Un support de formation

- Formation par les ouvrages pour la jeunesse
 - Exemple instructif de stage écolier

▶ Un support d'expérimentation numérique

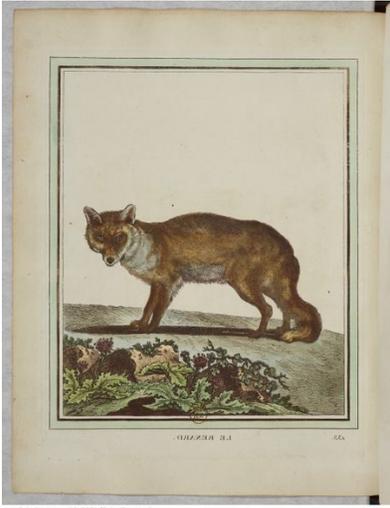
- Très grande hétérogénéité thématique et bibliographique

▶ Un support d'apprentissage découverte

- Exemple : La résolution des références bibliographiques



Histoire naturelle, exemple



* Le Renard; en Grec, *Ἀλώπηξ*; en Latin, *Vulpes*; en Italien, *Volpe*; en Espagnol, *Raposa*; en Allemand, *Fuchss*; en Anglois, *Fox*; en Suédois, *Raef*; en Polonois, *Liszka*.
Vulpes. Gesner. *Icon. animal. quadrup. pag. 88.*
Vulpes. Ray. *Synops. animal. quadrup. pag. 177.*

Simianipa caudam Cercopitheci habet, quod hæc icon non ostendit: ut inde cætera etiam non probe expressa aliqui suspicentur.

ΣΙΜΙΟΥΛΠΑ dici potest hoc animal peregrinum ficto nomine: Græcè *ἰνδικαλόωψ*, aut *ἰνδικαλοπίθηκος*: propter ambiguum inter hæc animalia corporis speciem.

GERM. Disf sömmd thier mag ein Fuchsfass genemnt werdē: daß es mit dem kopff gestaltet ist wie ein Fuchf/ hündē auf aber wie ein Aff mit einē langē schwanz wie ein Meerkatz: wiewol das in der beygesetzten figur / die wir vñ einer getruckten Mappen der wält/ nachgemacht/ mit angezeigt wirt. Arbores in Pariana regione ingentes sunt, (inquit Petrus Martyr, Oceanæ Decadis 1. lib. 9.) inter quas repertum est animal monströsum, uulpino rostro, Cercopitheci cauda, uesperulionis auribus, manibus humanis, pedibus Simiam æmulans: quod natos sibi catulos circumfert, quocunq̃ proficiscatur, utero exteriorē in modum magnæ crumenæ dependente. Id, licet mortuū, ipse uidi, conuolui, crumenamq̃ illam nouum uterum, nouum naturæ remediū, quō à uenatoribus aut alijs à cæteris uiolentis ac rapacibus animalibus natos liberet, illos secum



Monodelphis domestica



Opossum gris

Classification

Règne	Animalia
Embranchement	Chordata
Sous-embr.	Vertebrata
Classe	Mammalia
Sous-classe	Theria
Infra-classe	Metatheria
Ordre	Didelphimorphia
Famille	Didelphidae
Sous-famille	Didelphinae

Académie lorraine des Sciences



Académie lorraine des sciences

[Accès aux collections] Sciences naturelles de Strasbourg // Sciences de Nancy // Bulletin de l'ALS



▶ Repères historiques

- **Création Strasbourg** 1820 (Muséum Hist Nat) + Faculté de Médecine
 - 6 tomes (500 à 1000 pages) numérisées aux USA (site BHL)
- **Transfert à Nancy** en **1870** (en liaison avec faculté de Médecine)
 - Bulletin officiellement trimestriel arrêté en 1914 et 1939
- **Devient Académie Lorraine des Sciences** en 1961

▶ **Bibliothèque patrimoniale** de milliers d'articles

- Des centaines de PDF de 20 à 1000 pages, simplement juxtaposés

▶ Réédition hypertexte : grande variété de chemins éditoriaux

- Publications scientifiques de 10 à 30 pages,
- 100 pages pour thèse ou pour un texte de référence sur les bélemnites,
- 250 pages sur un répertoire archéologique galloromain en Moselle
- La vie de l'association (et des Universités associées) : Comptes rendus, listes des membres...

▶ Grande variété de chemins thématiques

- En relation avec le réseau multidisciplinaire des wikis thématiques
- Histoire naturelle, Santé, Histoire lorraine etc.

Pasteur

Poincaré

Bernheim

Coué

Un sujet idéal pour tester l'approche WICRI

ALS – Modèle collaboratif reproductible

Inspiré par la dynamique Wikipédia

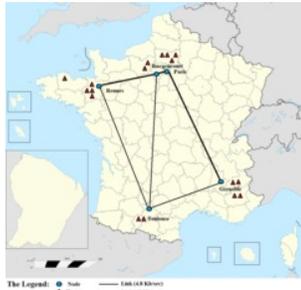
- ▶ **Wicri** : Projet mené par un seul ETP permanent
 - 150 wikis, 41.000/210.000 pages, 610.000 interventions...
 - 1 séance de deux heures pour corriger des texte
 - 1 semaine pour créer des pages sur un modèle défini avec assistance
 - 1 à 3 mois pour former un rédacteur administrateur avec assistance
- ▶ Peut alimenter des société savante thématiques du réseau Wicri
- ▶ **ALS** : 2 à 5 années de travail (ETP) avec déjà des résultats partiels
 - La société savante détermine des priorités
 - Les sociétaires peuvent participer puis assister
 - Par exemple des étudiants en Master ou en thèse
- ▶ **En Grand Est**, une vingtaine de sociétés savantes
 - Dispositif régional d'assistance et de production (5 personnes pour permanences),
 - pouvant aborder les recherches en cours
- ▶ **En France**, une dizaine de régions, et des sociétés thématiques
 - Dispositif national,
 - à partir de 5 personnes support Wikis sémantiques
 - 15 personnes, avec un atelier dédié aux logiciels libres utilisables dans le réseau (nouveau DILIB, Omeka ...)
 - 100 personnes pour un noyau de forte couverture multilingue
- ▶ **En Europe**, fédération multilingue des wikis thématiques

Oui nous pouvons !

- ▶ Retrouver la grandeur de notre système d'information scientifique et culturelle
 - ▶ Dans une coopération internationale

 - ▶ Du carburant pour l'Intelligence artificielle
 - ▶ En conservant notre tradition humaniste
- 

Hommages



1977
Pascal
Mistrail
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