21e journée sciences et technique de la production du GDR MACS

Personnalisation de masse pour le patrimoine industriel numérique

Gestion de connaissances, indexation supervisée, auto-indexation

Matthieu QUANTIN

IS3P / IRCCyN / École Centrale de Nantes Centre François Viète / Université de Nantes



- Introduction
- 2 Constats et Objectif
- **6** L'existant
- 4 Proposition
- Résultats
- Perspectives et conclusio

2/31



Présentation du contexte de recherche

Interdisciplinarité SPI/SHS

Groupe de recherche EPOTEC ^a

- IRCCyN^b (École Centrale de Nantes) Génie industriel / Gestion des connaissances
- CFV ^c (Université de Nantes) Épistémologie / Histoire des techniques

Thèmes de recherche

Gestion des connaissances appliquée à la conservation et la valorisation du patrimoine technique et industriel

- Rétro-conception
- Gestion des connaissances historiques
- Patrimoine industriel
- a. Évolution des Procédés et Obiets Techniques
- b. Institut de Recherche en communication et cybernétique de Nantes
- c. Centre François Viète





Valorisation, médiation, analyse

- Cintreuse à membrure "Bennie"
- Machine à laver le sel de Batz-sur-Mer
- Le pont transbordeur de Nantes
- La poudrerie Royale de Saint Chamas
- Explornova
- Nantes1900
- → Nombreuses études historiques associées, formats sources conservés.

Gestion de connaissances

- modèle conceptuel : Digital Heritage Reference Model [?]
- Structure de donnée : Nantes1900 [?]
- Extraction de connaissances : Haruspex [?]

Quelques réalisations



FIGURE : Le pont transbordeur de Nantes en Réalité Augmenté in situ



Présentation d'un cas d'étude

150 ans d'histoire du port de Nantes

Un vivier d'objets techniques remarquables :

- le pont transbordeur
- les grands moulins
- les grues de chantiers navals
- la salle à traçer
- les halles Alstom
- la maguette au musée d'histoire

Tradition forte de réhabilitation/conservation

Mise en abyme : la maquette du port de Nantes en 1900 (P-A. Duchesne)





Présentation d'un cas d'étude

Focus sur la maquette du port de Nantes et les halles Alstom

Maquette du port de Nantes

- \sim 500 POIs
- > 1k documents icono et biblio
- Modèle 3D : 100 M de points
- Dispositif muséographique en RA

Halles Alstom

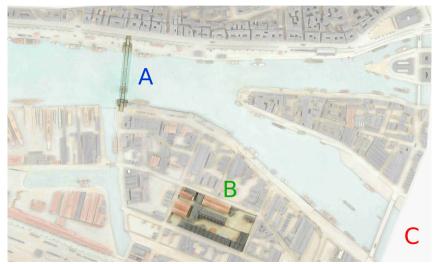
- Halls de construction mécanique
- En cours de réhabilitation
- Une monographie d'histoire des techniques (M2)
- ~ 250 références biblio et archives





Présentation d'un cas d'étude

Focus sur la maquette du port de Nantes et les halles Alstom





- Introduction
- 2 Constats et Objectifs
- **6** L'existant
- 4 Proposition
- 6 Résultats
- Perspectives et conclusio



Numérisation de masse

Production numérique

Dans le domaine de l'histoire des techniques

- production de récits historiques nativement numérique
 - → aucune capitalisation hors papier/pdf

Acquisition numérique

- scan et OCR de corpus de documents historiques
 → pas d'outillage pour analyser les textes bruts
- scan 3D (objets et bâtiments du patrimoine)





Numérisation de masse

Production numérique

Dans le domaine de l'histoire des techniques

- production de récits historiques nativement numérique
 - → aucune capitalisation hors papier/pdf
- production de maquettes 3D
 - → mise en relation manuelle des récits

Acquisition numérique

- scan et OCR de corpus de documents historiques
 - ightarrow pas d'outillage pour analyser les textes bruts
- scan 3D (objets et bâtiments du patrimoine)





Enjeux

Construire une chaîne de traitement personnalisée pour chaque cas d'étude

- à l'indexation
- à l'interfaçage

Donner des accès à la masse de donnée obtenu

- niveau d'exploration ajustable
- permettre l'enrichissement d'un réseau

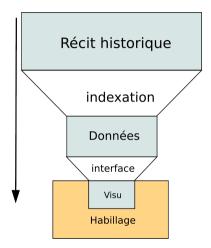


FIGURE : Entonnoir de la connaissance en histoire



Enjeux

Construire une chaîne de traitement personnalisée pour chaque cas d'étude

- à l'indexation
- à l'interfaçage

Donner des accès à la masse de donnée obtenu

- niveau d'exploration ajustable
- permettre l'enrichissement d'un réseau

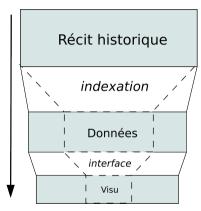


FIGURE : Tendance désirée sur l'entonnoir



Construire une chaîne de traitement personnalisée pour chaque cas d'étude

- à l'indexation
- à l'interfaçage

Donner des accès à la masse de donnée obtenu

- niveau d'exploration ajustable
- permettre l'enrichissement d'un réseau

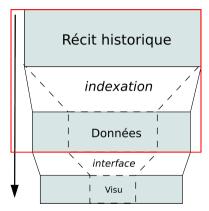


FIGURE : Tendance désirée sur l'entonnoir



Outil

Proposer un mécanisme d'indexation supervisée de récits historiques

Objectifs

- faciliter l'intégration de nouveaux contenus
 - plusieurs interprétations = confrontation de récits
 - plus d'objets
 - plus de détails
- proposer des accès multiples
 - utilisateurs / acteurs
 - lectures transverses
 - niveaux de détail



Outil

Proposer un mécanisme d'indexation supervisée de récits historiques

Objectifs

Matthieu Quantin

- faciliter l'intégration de nouveaux contenus
 - plusieurs interprétations = confrontation de récits
 - plus d'obiets
 - plus de détails
- proposer des accès multiples
 - utilisateurs / acteurs
 - lectures transverses
 - niveaux de détail

Hypothèse de travail

Exploiter un outil de TAL : ANA [?]

- indépendant de tout dictionnaires
- agrégation itérative de termes (expansion, expression, noyaux)
- 80% de résultats pertinents sur notre corpus
- pertinence significative : « maconneries de moellons hourdés de chaux »

À combiner avec une chaîne de traitement pour l'indexation supervisée



Problématique

Problématique : Comment indexer les connaissances historiques pour en optimiser l'exploitabilité?

Coté données historiques

- Considérer la spécificité de chaque récit
- Construire des réseaux d'informations
- Hiérarchiser selon une échelle de spécificité

Coté valorisation

- Offrir une interface d'exploration visuelle
- Adapter cette interface d'exploration à l'utilisateur
- → Seulement le coté "données historiques" ici.

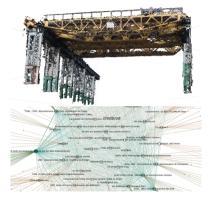


FIGURE: Scan 3D d'un pont roulant et représentation des connaissances associées



- 1 Introduction
- Constats et Objectif
- 3 L'existant
- 4 Proposition
- 6 Résultats
- Perspectives et conclusion

13/31



Les données

Format des données

- texte brut (corpus)
- XML schema TEI (corpus ou monographie)
- .tex .doc (monographie)

Représentation minimale

3 descripteurs pour analyse :

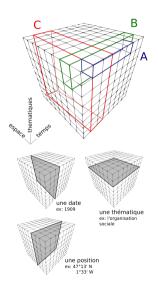
- temps
- espace

Matthieu Quantin

sémantique

Composantes temporelles et spatiales "optionnelles" ex : Une étude sur les "conditions de travail au début du XX^e siècle" \rightarrow pas de composante spatiale. → pas de donnée sémantiquement non analysable

Nous prenons cela comme point de départ





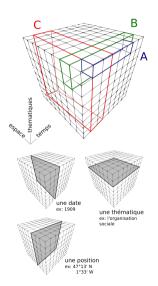
Contraintes

Caractéristiques des données

- récit non constant dans le degré de précision
- aucun vocabulaire contrôlé ne régit le récit
- aucun marqueur sémantique ou presque (TEI)

Contraintes humaines

- s'adapter aux pratiques des historiens (et pas l'inverse)
- simple et configurable
- produire des niveaux d'accès au réseau d'information

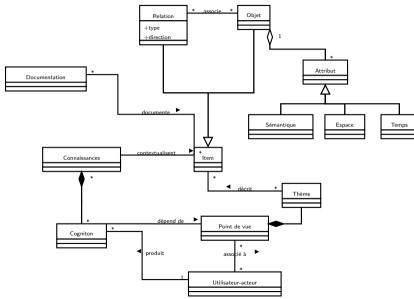




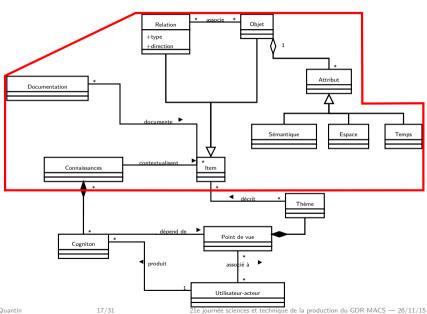
- Introduction
- Constats et Objectif
- **B** L'existant
- 4 Proposition
- 6 Résultats
- Perspectives et conclusion



Le modèle









Le modèle

La chaîne de traitement

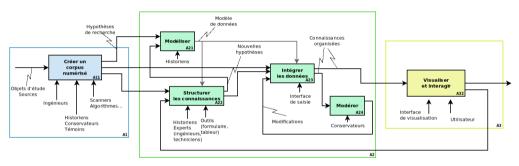
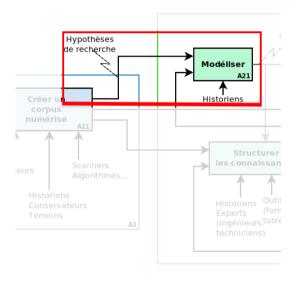


FIGURE : chaîne de traitement des données historiques



La chaine de traitement



19/31

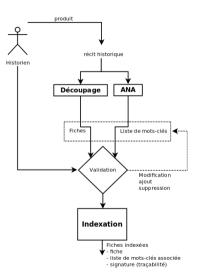
Modéliser

- Modèle = structure du récit ou du corpus
- Hypothèses de découpe configurables
- Conservation de la validité historique du texte

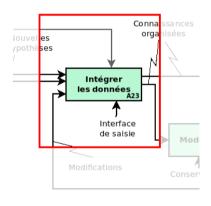


Structurer les connaissances Outils Historiens (formulaire, Experts (ingénieurs tableur) techniciens)

La chaine de traitement







La chaine de traitement

Extraction

- termes métiers, jargon...
- du générique à la composition spécifiques
- nom propres, dates

Intégrer

- un lien = une co-occurrence de mot-clef entre fiches
- lien pondéré : TF-IDF du mot-clef
- démonstrateur avec Neo4.1

TF-IDF : métrique de distance sémantique



- Introduction
- Constats et Objectif
- **6** L'existan
- 4 Propositio
- **5** Résultats
- 6 Perspectives et conclusion

22/31



Réseau sémantique

Mise en relation

- interne = dans un récit linéaire
- externe = plusieurs récits

Exploitation

- métrique de sémantique
- niveaux d'exploration non discrétises
- suggestion de liens = sérendipité
- → contextualisation des éléments les uns par les autres

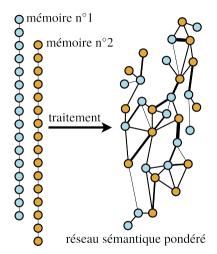


FIGURE : Réticulation des informations historiques



Résultats



FIGURE : Interface graphique du logiciel développé



- Introduction
- Constats et Objectif
- **6** L'existan
- 4 Proposition
- Résultate
- 6 Perspectives et conclusion



Perspectives

Mise en relation

- liens récits / corpus
- liens récits / données physiques (3D, SIG)
- application à des cas industriels

Pre et post-process

- aide à la création de corpus
- application au projet "Salons Mauduit" pour janvier 2016
- standardisation de certains résultats (requêtes ext.)

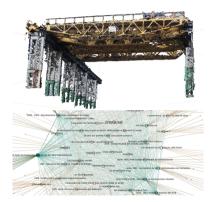


FIGURE : Scan 3D d'un pont roulant et représentation des connaissances associées



Conclusion

Résultats

- Capitalisation de récits en histoire des techniques
- Réseau sémantique vs. linéarité et juxtaposition
- Approche centrée sur les pratiques de l'historien
- Utilisateur au cœur de la méthode (indexation supervisée)
- Outil complémentaire aux pratiques usuelles
- Détection de "zones" faiblement documentées
- Bonne traçabilité de l'information (sources)

Usagers

- satisfaction du producteur sur les 1^{er} tests
- corrélation des analyses humaines et machine sur corpus
- possibilité de moduler le degré d'exploration pour l'usager



Conclusion

Personnalisation de masse

A deux niveaux :

- processus : extraction de vocabulaire personnalisé à chaque récit
- accès : exploration modulaire de l'environnement historique d'un objet.

L'interface

Un premier prototype conceptuel d'interface a été développé :

pict/screenshot.png



References I



Merci de votre attention

matthieu.quantin@irccyn.ec-nantes.fr
http://epotec.fr



Introduction