

Analyse et modélisation des textes

TEI@Oxford

Magritte nous rappelle



Numériser et encoder...

La numérisation nous propose une *image*, une représentation d'un objet déjà existant

L'encodage nous permet de représenter l'image des *idées* résultant de ces représentations

Des idées sur quoi?

Tout! mais on peut distinguer plusieurs axes...

- infos 'intrinsèques' à l'objet : les formes, couleurs, etc. d'une image; les sons, rythmes, etc. d'une musique; les structures linguistiques (mot, phrase, paragraphe) ou formels (chapitre, titre, nom de lieu) d'un texte...
- infos 'extrinsèques' ou 'meta' sur l'objet : son type, ses origines, ses buts, ses usages ...
- infos 'interprétatives': la portée d'un texte ou d'un dessin, le programme d'une musique ou d'un rite...

distinctions floues mais pervasives...

Modélisation et structuration

Il existe (quelques) méthodes classiques d'analyses de données.
L'important c'est de comprendre:

- toute méthode ne serait qu'une modélisation
- la modelisation devrait faire ressortir la structuration essentielle d'un objet complexe

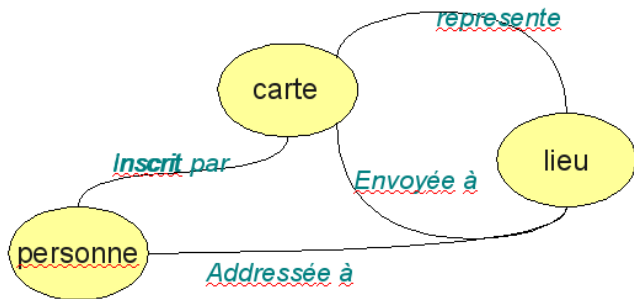
Après une modélisation, on peut donner une implémentation informatisée ; sans modèle, une implémentation risque d'être incompréhensible et aléatoire, inutile...

Analyse des données classique

On identifie...

- les 'objets d'intérêt'
- leurs attributs/propriétés
- les relations entre objets
- les procédures/traitements essentiels envisagés

Par exemple...



Analyse des documents

- quels sont les unités que nous souhaitons traiter?
- comment sont-ils structurés (quels composants, quels attributs)?
- est-ce que les occurrences de ces objets sont clairement identifiables dans un flux de texte?

Essayons cela avec ces trois documents:

Document A

78 - LES SABLES D'OLONNE (Vendée)
Le Port



Bientôt sur le chemin du retour
avons eu du beau temps tout le
mois de juillet. Vous souhaitez
à tous la même chose.

J'espère que vous êtes tous en bonne
santé - à chacun son tour, le
bon temps et le repos.

avec d'affectionneuses pensées et
bons baisers à tous.

Marcelle

ARTAUD Frères - Éditeurs
Avenue de la Cloze - NANTES

M^{me} Veyssade Lucienne

31, rue du Président

Kennedy -

ALFORTVILLE

~~Val de Marne~~ Val de Marne)

Document B

O/1/284/ 525

PC45014047

5 CM

Mai 1763 ordre du Roy

Le pavillon du Roy comme
al'ordinaire, mais l'on ne
donnera de bois d'arrivée à
personne



Le voyage étant comme
celui de Versailles, l'on n'est
obligé à rien, mais on peut
leur donner quelque secours en
bois et linge, Mr Boucheman
et le Moine qui suivent

ou par route

Le S. Luthier demande si l'on éclairera
et chauffera le Château de Marly le
voyage prochain, comme aux autres voyages
ordinaires.

Si le Concierge, les Garçons du Gardemeuble,
Le S. Boucheman, et le Moine, Les
Garçons du Château, l'Orloger, le Garçon
des Châux, et les Porteurs seront chauffés
et éclairés.

Si on éclairera les Salles des Gardes
de Votre Majesté, des Gardes de
la Porte, et des Com-Suisses.

Document C

F. LE LIONNAIS
23, route de la Reine

001

92 - BOULOGNE SUR SEINE

le 3 Mars 1969

Tél. 605 90 13

Chers Brigadiers,

Le sort en reste jeté. Raymond QUENEAU et René THOM viennent de me confirmer - avec des accents qui ne laissent aucun doute sur leur sincérité - que notre prochaine réunion aura bien lieu :

Lundi 17 Mars à 17 Heures, chez Georges PEREC (4° Etage)
92, rue du Bac PARIS 7° - Téléphone : 222 95 68

On est prié de ne pas arriver en retard, car certains participants m'ont exprimé leur désir de se libérer avant 18 H 45.

Il avait récemment couru le bruit que le signataire de cette lettre était coutumier du parjure et qu'on ne pouvait lui faire confiance pour la livraison du "Deuxième Manifeste" destiné aux Editions Cape. C'était là pure calomnie. Voici le texte ou du moins un projet que vous êtes invités à analyser, critiquer et compléter. Voyez-vous d'autres aspects de l'idéal OuLiPien qui mériteraient d'être mentionnés ? Ou croyez-vous que l'aspiration sémantique devrait être développée ?

Je vous laisse sur ces méditations et vous adresse mes pensées amicales.

P.S. Claude BERGE : Et bien entendu, je t'attends chez moi Jeudi 13 Mars à

Exercice 2 : analyse d'un document

- Dans votre groupe, sélectionner d'abord un rapporteur
- Regarder bien votre document. Supposer que vous en avez quelques centaines d'autres pareilles.
- Comment est votre document organisé?
- Quels sont ses composants essentiels, commun à tous ces documents ?
- Quels sont ses composants intéressants, possible dans tous ces documents ?
- Sauriez vous faire en sorte qu'une autre personne reconnaitrait les mêmes objets ?
- Faire une liste de tous les objets et propriétés essentiels de votre document
- Présenter cette liste ... et la justifier !