
Informatique, norme et temps

Isabelle Boydens

Présentation par Marie-Anne Chabin
Réunion PIN
15 janvier 2004

Les bases de données empiriques

- Collection finie et structurée de données codifiées, textuelles ou multimédia, destinées à représenter certains aspects du réel observable, appelés « domaine d'application »
- Le réel empirique est continu et hétérogène : on ne trouve jamais dans les concepts généraux que ce qu'on y met
- Absence de référentiels

Qualité des données

- Les domaines de la Data quality research
 - data cleansing : mise en cohérence automatique,
 - data tagging : documentation du schéma d'une base de données à l'aide d'indicateurs de qualité (complexe, peu efficace),
 - data tracking et reengineering : méthode statistique pour détecter les erreurs logiques, conduisant à une rationalisation des processus

Qualité des données (suite)

- Les constats de départ :
 - efforts de conception concentrés sur l'intégration de nouvelles technologies au détriment des difficultés conceptuelles
 - flux d'information qui alimentent les bases de données ne suffisamment pris en considération (usine fantôme)
 - on ne fait pas assez attention aux usages de l'information
 - contexte de l'information insuffisamment pris en compte
 - la fusion de bases de données hétérogènes à d'autres fins que le but initial détourne les usages
- Seule la « field theory of information » prend vraiment en compte le contexte

La bases de données de la Sécu belge

- base LATG :
 - >300 données potentielles chaque trimestre pour
 - 200.000 employeurs et 3.000.000 de travailleurs
 - perception annuelle d'environ 38 milliards d'euros
 - base de 60 Go en 1998
 - 100.000 consultations ou modifications par jour

Statistiques versus administration

- Les statistiques incluent explicitement une tolérance à l'erreur... mais la représentation de la loi au sein d'une base de données requiert des arbitrages humains qui peuvent s'appuyer sur :
 - l'heuristique (collecte et confrontation de toutes les sources écrites en l'absence de référentiel) et reconstitution de la généalogie des données (*stemma codicum*)

Arbitrages humains

- et sur :
 - l'herméneutique, science de l'interprétation

 - un système d'information administratif a une incidence sur le processus d'interprétation et d'application de la loi ainsi que sur le réel normé ; le système est directement constitutif de l'objet qu'il représente

Approche orientée-objet

- Un diagramme de flux d'information est orienté sur les tâches (functional orientation), lesquelles sont décrites en termes de transformation des données (donc contraint par l'existant)
- L'analyse orientée-objet, en encapsulant à la fois états et comportements... offre une heuristique créative...

Transformations dans le temps

- les versions (de schéma ou de processus) se traduisent par une succession de transformations ponctuelles
- Mais le changement est inscrit au coeur même de la nature de certains processus
- Au concept discret de version se substitue celui de processus « fluctuants » (*forma fluens* => *fluxus formae* de Guillaume d'Ockham)

Construction progressive des concepts administratifs

- La conformité des valeurs à la sémantique d'une base de données ne constitue pas une condition suffisante afin de garantir la validité des données qui s'étend au-delà de la partition dichotomique entre valeurs formellement correctes et incorrectes
- Il faut analyser le processus de construction du couple « intention / extension » au sein d'une base de données administrative (signification / l'ensemble des objets auxquels les caractéristiques du concept peuvent s'appliquer)

Bases de données temporelles

- Ajout d'attributs et/ou entités permettant la prise en compte de l'évolution dans le temps de l'extension de la base de données (la base de données « non-temporelle » permet seulement conserver le dernier état de la base)
- + statut hétérogène de l'information saisie, adéquation relative des valeurs aux spécifications formelles et nature des opérations appliquées à la base de données

Versions de schéma

- Versions de schéma linéaires
- Versions de schéma rétroactives (mesure légale à effet rétroactif)
- Versions extensionnelles : prise en compte de nouveaux cas de figure => modifier le domaine de définition et étendre la modification à tous les cas concernés
 - concept de boucle étrange (Ost et Kerchove, Hofstadter) : « quelque chose d'intérieur sort et agit sur le système »

Notion de contraintes d'interprétation

- Méthode d'automatisation de certaines procédures d'interprétation humaine car dans tout système d'information empirique, l'émergence d'informations « exceptionnelles » fait partie intégrante du contrat initial
- Identification de cas de figure jugés problématiques *a priori* et dont l'interprétation repose sur un jugement humain

Expérimentation

- Base LATG : chaque trimestre, 1200 millions de valeurs – environ 600 000 anomalies par trimestre (0,05 à 0,09 % des valeurs saisies), 300 personnes à temps plein,
- La prolifération de la réglementation fait augmenter les contraintes d'interprétation (de 1000 à 15000 par trimestre de 1994 à 1997) – 25 % des valeurs vérifiées étaient finalement valides
- Pour 1997, la méthode permet une diminution de 50% des anomalies au trimestre suivant

Les contraintes d'interprétation (suite)

- Méthode : analyse diachronique du nombre d'anomalies formelles et de leur traitement ; interprétation progressive de la base de données en extension, transformation de la base de données en intention
- Accompagner et rationaliser la gestion quotidienne en enrichissant le schéma.
« L'enrichissement du schéma de la base de données s'apparente à la conception d'un appareil critique automatisé »

Questions d'après lecture

- Faut-il pérenniser ce type de base de données administratives, et comment ?
- Les contraintes d'interprétation peuvent-elles ou doivent-elles être utilisées ou prises en considération pour l'archivage ?
- La notion du temps, vue par I. Boydens pour son influence sur la construction de la base et de son interprétation, pourrait être approfondie dans une perspective Archivage