

Panteda

Production & ANalyse of TEmporal Data
Lig Leader , 11 Avril 2016

Constats

méconnaissance des algorithmes, des applications développés au sein du laboratoire pour traiter les séries temporelles



Alors que des outils existent au laboratoire

méconnaissance des données existantes qui pourraient être utilisées pour tester des algorithmes



Alors que des données existent

données temporelles produites au laboratoire qui sont perdues. Des algorithmes ne marchent plus au départ des projets



Alors que la répétabilité est de plus en plus souvent évoquée

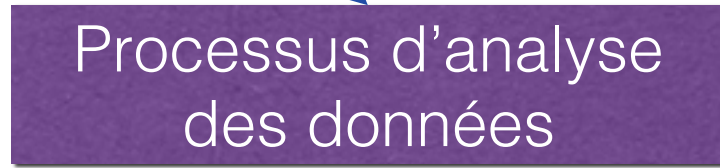
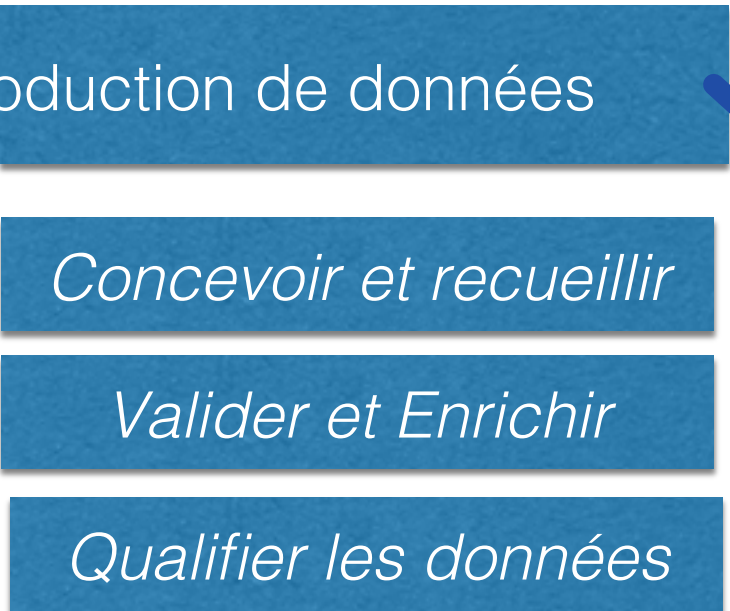
Problématiques au laboratoire

Problématiques de stockage des données et des
algorithmes d'analyse,

Problématiques de conception d'opérateurs d'analyse
et de visualisation des données,

Renforcer car = Des volumes de données différents
big data et des small data/big knowledge

Chaîne de traitements



données avec une dimension temporelle ou séquentielle

Environnement du LIG

	Production de données	Capitalisation des données	Production d'opérateurs d'analyses	Capitalisation des opérateurs	Construction des processus	Capitalisation des processus	Accompagnements des utilisateurs
weeps - LIG/SLIDE	x	x, big data	x				
cks -LIG/Metah		x, small data big knowledge	x	X, python	X, orange 2.7	x	
-/LIG- AMA et CS		X, big data	x	X	x	x	x
ation - LIG/IHM			x				
patterns -Convecs LIG			x				
ries temporelles MA/LIG			x				
ésentations ques LIG/Steamer			x				
l'activité par des s LIG/SIGMA			x				
omus/Fablab	x						
o/ data center ersyval		x tous les types de données					
er avec entratiens							

Objets de Pantéon



Séries temporelles

Capitaliser les données dans une structure qui gère le « temporel »

- Offrir les moyens de stocker des données avec les metadata pour les diffuser (A4H, Domus, plateformes des équipes) (=> Perscideo)
- Procurer des jeux de données temporelles pour tester algorithmes. (Ama, Convecs, Sigma, Slide, Steamer ...).

Avoir une plateforme disponible pour capitaliser les opérateurs de données des équipes et (tenter de) garantir la pérennité de ces opérateurs.

Avoir un outil pour construire des processus d'analyse, les sauvegarder et les partager.

Accompagner et guider les chercheurs non spécialistes dans l'analyse des données.

Zoom sur les Plateformes

Points forts

Points faibles

ets

- Big data, adaptée à une grande quantité de données temporelles et spatiales
- Orientée collecte et stockage de données
- Optimisée pour les requêtes en géomatique
- Couplable avec des autres données (en particulier géographiques)
- Capture de données de grands volumes en temps réel
- Des utilisateurs

- Pas d'opérateurs
- Droit d'utilisation des données limité par T

racks

- Capitalisation des données, des opérateurs et des chaînes de traitements
- Un outil de construction des processus
- Opérationnelle, interfaces conviviales
- Plateforme V1 bientôt achevée
- Une gestion de la propriété des données
- Des utilisateurs

- Traitements et stockages non optimisés,
- Lié à Python pour le développement des opérateurs
- Lié à Orange 2.7 pour les traitements,
- Pas de typage informatique
- Pas de typage statistique des données
- Données téléchargées sur le poste client
- Traitements limités au million de données

es

- Capitalisation des données, des opérateurs et des chaînes de traitements
- Une architecture technique orientée données temporelles

- Dépendance de l'architecture fournie par
- Mise en production septembre 2016
- Initié chez CS
- Compétences au LIG

Capitaliser

Conception et développement de la structure de stockage : Septembre 2016

Conductre une étude des besoins au niveau des équipes du laboratoire

Conductre une veille technologique sur les plateformes existantes et les outils disponibles

Étudier les problèmes et limites de l'architecture de Ikats

Identifier des données du laboratoire à intégrer dans la plateforme (A4H, Do

corpus, applications du labos,)

Spécifier et développer une plateforme de capitalisation des données temporelles

Interface d'importation et de consultation : Décembre 2016

Définir les besoins en termes de métadonnées pour décrire ces données et

opérateurs,

Développer une interface pour déposer les données,

Déposer les données dans la plateforme et intégrer les anciennes données.

Analyse => Avril 2017

Spécifier et Développer une interface pour déposer les opérateurs

Adapter l'opérateur de visualisation de R.Blanch pour la plateforme

Spécifier un outil de création des processus d'analyse

Ressources

Étudiants PIMLIG : E.Dublé, M.Ortega, N.Mandran

Rechercheurs: G.Bisson, P.Reignier, D.Bouhineau, F.Jambon, Bouanneau, ...

Ressources contractuelles : CDD PIA IKATS

Contribuer avec la société CS Toulouse (projet Ikats)