

GdR MADICS
Journées Science des Données 22 - 23 Juin 2017
Lieu : Ecole de Management de Marseille

Compte rendu sur l'organisation de l'atelier QUALIMADOS Qualité des masses de données scientifiques

Objectif

L'objectif de l'atelier était de créer un espace d'échanges entre chercheurs, experts et professionnels sur les nouveaux enjeux et défis liés à la qualité des données scientifiques dans l'ère du Big Data.

L'atelier a offert également une bonne occasion de discuter et de réfléchir sur la création d'une action liée à "la qualité des masses de données" dans le GdR MADICS

1) Organismes

Allel Hadjali (LIAS/ENSMA, Poitiers, allel.hadjali@ensma.fr)
Laure Berti-Equille (IRD, Montpellier, laure.berth@ird.fr)
Angela Bonifati (LIRIS/CNRS, Lyon, angela.bonifati@univ-lyon1.fr)

2) Lieu et date

L'atelier est organisé conjointement avec les journées Science des Données du GdR Madics:

- Lieu : Ecole de Management de Marseille
- Date : 23 juin 2017

3) Participants

L'atelier a réuni une vingtaine (20) de personnes : des enseignants-chercheurs/chercheurs, des professionnels et des doctorants.

4) Invités

L'atelier a invité deux orateurs pour leur expertise dans le domaine de la qualité des données:

- Cédric Pruski, ITIS, Luxembourg
- Bastien Rance, HEGP, Paris

La prise en charge des frais (liés à leur déplacement et hébergement) a été assurée par le GdR Madics.

5) Programme

Le programme proposé est composé de 6 présentations abordant différents aspects liés à la qualité des données en général, et plus spécifiquement, des données scientifiques.

Certaines d'entre elles ont fourni un tour d'horizon sur les défis actuels posés par la qualité des données scientifiques et d'autres ont présenté certaines solutions, traitant la qualité, mises en place dans des applications réelles.

Une discussion riche et constructive a eu lieu à la fin de la journée où un bilan a été dressé et un débat sur l'intérêt, de créer une action au sein de Madics sur la qualité de données, a également été mené.

La liste des présentations orales est :

- 10h15 – 10h30 : Présentation de l'atelier et de ses objectifs (Allel Hadjali, LIAS/ENSMA, Poitiers)
- 10h30 – 11h00 : Data Quality: where are we on the journey from theory to practice? (Angela Bonifati, LIRIS, Lyon)
- 11h – 11h30 : Tour d'horizon des données scientifiques et des problématiques particulières liées à leur qualité (Laure Berti-Equille, IRD, Montpellier)
- 11h30 – 12h15 : Gestion des annotations sémantiques en santé - Le projet ELISA (Cédric Pruski, ITIS, Luxembourg)
- 14h – 14h45 : Prise en compte des données manquantes dans les modèles de mélanges : Application aux séries temporelles d'images multispectrales (Serge Iovleff, Université de Lille)
- 14h45 – 15h30 : Qualité dans l'entrepôt de données cliniques de l'HEGP (Bastien Rance, HEGP, Paris)
- 15h30 – 16h : Discussions et clôture de l'atelier

6) Bilan

La diversité des présentations a soulevé de nombreuses questions sur le sujet tant au niveau théorique que pratique. Ce qui montre l'intérêt du sujet au sein de la communauté scientifique.

L'idée de poursuivre notre travail vers la création d'une action sur la qualité des données, au sein de Madics, a reçu une adhésion totale de la part des participants. Les fondations et les objectifs de cette action seront présentés en détail dans un prochain courrier au responsable de Madics.

Remerciements.

Nous tenons à adresser nos vifs remerciements aux responsables de Madics et à l'ensemble des personnes œuvrant dans ce GdR pour leur soutien financier et aussi pour toutes facilités mises à notre service pour la réussite de cet atelier.