

Groupe de contact FNRS

« Analyse critique et amélioration de la qualité de l'information numérique »

16 avril 2024

“Approches quantitatives de textes historiques : quelques (non-) problèmes et comment les aborder ?”

par

Simon Hengchen

Introduction

Isabelle Boydens

Groupe de contact FNRS

« Analyse critique et amélioration de la qualité de l'information numérique »

- **Qualité des données :**
 - « *fitness for use* » : adéquation des données à leurs objectifs (sociaux, industriels, juridiques, scientifiques, ...), sous *contrainte de budget*
 - pas de « *qualité totale* »
 - approche continue et pluridisciplinaire
 - Données empiriques et évolutives

- **Groupe de contact FNRS créé en 1994 (30 ans en 2024)**
 - <https://www.frs-fnrs.be/fr/financements/mobilite-pays-specifiques/groupe-de-contact/91-fr/nos-financements/mobilite-fnrs/groupe-de-contact/103-sciences-appliquees>
 - <https://www.frs-fnrs.be/fr/financements/mobilite-pays-specifiques/groupe-de-contact/91-fr/nos-financements/mobilite-fnrs/groupe-de-contact/104-sciences-humaines-et-politiques>

- **Centre de compétence en qualité des données, Smals**
 - <https://www.smals.be/fr/content/data-quality>
 - <https://www.smals.be/nl/content/data-quality>

- **Cours de Master en STIC à l' Université libre de Bruxelles :**
 - « **Qualité de l'information et des documents numériques** »
 - <https://www.ulb.be/fr/programme/stic-b510>
 - <https://isabelle-boydens.web.ulb.be/>

Enjeux stratégiques lorsque l'information empirique et évolutive est un instrument d'action sur le réel (question du temps !)

Quelques exemples :

- 1442, Laurent Valla démontre que la « *Donation de Constantin* » est un faux antidaté de 4 ou 5 siècles
- Étude du réchauffement climatique (couche d'ozone, 1980)
- Première guerre du Golfe (1990-1991)
- 1999, guerre du Kosovo, bombardement de l'ambassade de Chine à Belgrade
- 2008, 2017 interactions entre données et marché financier : Google Finance, Yahoo! Finance, World-Check (2017)
- 2013, secteur administratif (Obamacare)
- 2016, USA : coûts de la « non-qualité » (3.100 milliards \$, 20% PIB USA, T. Redman)
- 2017, domaine des pipelines, de l'hydrologie, ...
- 2018-2024, bases de données en matière de terrorisme, médicales (Implant files, Covid19, ...), Registre National, Censure & ML en Chine, Causal AI, ...

Thématiques 2014 – 2023 (1)

- 2014

- « *Du stemma codicum au data tracking* » : vingt ans de recherche en matière d'évaluation et d'amélioration de la qualité des bases de données », par I. Boydens,

- « *Email Address Reliability*, » par V. Berten

Programme, slides et vidéo : <http://mastic.ulb.ac.be/2013/10/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique/>

- 2015 : « *Ouverture des données, gouvernance et engagement citoyen: questions de qualité* »

- « *Gouvernance et individus: réflexion sur un couple impossible* », Par F. de Smet

- « *Semantic Information Technology as a provider of tools for a productive and efficient organization of government* » par Agis Papantoniou

Programme et slides : <http://mastic.ulb.ac.be/2014/10/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-2/>

- 2016

- « *Analyzing the evolution history of data-intensive systems in support to software maintenance*», par A. Clève

Programme, slides et vidéo : <http://mastic.ulb.ac.be/2016/02/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-3/>

- 2017

- « *Temps réel, déterminisme et ordonnancement : trois défis majeurs des systèmes embarqués*», par J. Goossens

Programme, slides et vidéo : <http://mastic.ulb.ac.be/2017/02/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-4/>

Thématiques 2014 – 2023 (2)

• 2018

- « *La préservation du patrimoine scientifique à l'heure du numérique* », par A. Leroy
Programme, slides et vidéo : <http://mastic.ulb.ac.be/2017/09/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-5/>

• 2019

- « *Méthodes quantitatives, analyse critique de données et histoire médiévale : des données contemporaines aux sources anciennes* », par S. de Valeriola
Programme, slides : <http://mastic.ulb.ac.be/2018/08/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-6/>

• 2020

- « *I want to talk to a HUMAN!* » : *impact de la qualité des bases de connaissances sur les agents conversationnels* », par M. De Wilde
Programme, slides : <http://mastic.ulb.ac.be/2019/09/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-7/>

• 2021

- « *Avez-vous essayé l'ignorance ? Les défis du knowledge management aujourd'hui* », par F. Rossion
Programme, slides : <https://mastic.ulb.ac.be/2021/03/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-8/>

Thématiques 2014 – 2023 (3)

- **2022**

- « *Apprentissage automatique : les challenges de la qualité des données dans la perspective d'une adéquation aux usages* », par L. Dierickx,

Programme et slides : <https://mastic.ulb.ac.be/2022/02/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-%EF%BF%BC/>

- **2023**

- « *Les répertoires : pivots indispensables et méconnus des systèmes d'informations* », par P. Rivière,

Programme et slides : <https://mastic.ulb.ac.be/2023/04/reunion-du-groupe-de-contact-fnrs-analyse-critique-et-amelioration-de-la-qualite-de-linformation-numerique-10/>

Thématique 2024

“Approches quantitatives de textes historiques : quelques (non-) problèmes et comment les aborder ?”

par Simon Hengchen

Simon Hengchen

Simon Hengchen est consultant en TAL et IA chez iguanodon.ai et chargé de cours en TAL à l'Université de Genève, en Suisse.

Titulaire d'un doctorat en Sciences et Technologies de l'Information et de la Communication de l'Université libre de Bruxelles, ses recherches se concentrent sur la détection du changement sémantique dans les textes historiques.

Il a également été chercheur postdoctoral à Göteborgs universitet, en Suède, et à Helsingin yliopisto, en Finlande. Pour obtenir un aperçu complet de son parcours académique et de ses publications : <https://hengchen.net>

Programme 2024

- 13H30 Introduction groupe de contact FNRS (I. Boydens)
- 13H35 « Approches quantitatives de textes historiques : quelques (non-) problèmes et comment les aborder ? » (S. Hengchen)
- 14H35 Table ronde (modérateurs : M. De Wilde et G. Quintin)
- 15H10 Drink

Les slides des présentations seront disponibles en ligne sur le site de l'invitation à la fin de la rencontre FNRS.