



Journée d'étude ITEM / EMAN / THALIM

Textes mathématiques et transcription en XML-TEI

Vendredi 31 mai 2024, 10h-16h30

Journée d'études organisée par Emmylou Haffner et Richard Walter

Journée en hybride :

Salle 319, Campus CNRS, Pouchet – <http://www.pouchet.cnrs.fr/acces> – 59/61 rue Pouchet, 75849 Paris cedex 17. Entrée libre dans la mesure des places disponibles.

Inscription par mail à <eman@groupe.renater.fr>

Visioconférence. Pour recevoir le code zoom envoyer un mail à <eman@groupe.renater.fr>

~~~~~

Nous organisons une journée d'étude sur les pratiques d'encodage et de visualisation des textes mathématiques à partir des différentes expériences de projets sur la plateforme EMAN (Dedekind/DWR, Manuscrits Leibniz et BANANA).

Si l'édition numérique de textes est aujourd'hui bien balisée, notamment grâce au langage XML/TEI, le traitement de certains objets complexes, comme les symboles et formules mathématiques, reste problématique. Pour les projets souhaitant éditer numériquement des textes scientifiques, se pose la question de gérer de la cohabitation – la coexistence – entre un système textuel « standard » et un système pouvant encoder et éditer des symboles mathématiques.

Le langage LaTeX est majoritairement utilisé dans les communautés scientifiques pour pouvoir générer des symboles mathématiques. Pour l'édition critique dans des bibliothèques numériques ou des bases de données en ligne, il ne semble, à première vue, pas toujours être le plus approprié. Par ailleurs, du côté des éditeurs de textes du patrimoine notamment, l'initiative TEI (Text Encoding Initiative) s'est imposée pour éditer, exploiter et archiver des documents textuels numériques. Elle a aussi approché d'autres standards d'écriture. Elle a ainsi proposé un formalisme pour pouvoir encoder et visualiser des notations musicales tout en restant dans l'univers de la TEI : la MEI (Music Encoding Initiative). À notre connaissance, la TEI n'a pas fait de même pour traiter les formules mathématiques. Les nouveaux projets d'édition de textes scientifiques demandent l'ouverture de ce chantier afin de mieux comprendre quels rôles pourraient avoir LaTeX et XML-TEI comme format pivot – s'il s'agit effectivement de la solution la plus appropriée.

Pour permettre l'édition et l'exploitation de ces textes sur la plateforme EMAN, nous avons développé un outil pour faire de la transcription en XML-TEI, le module Transcript, et nous utilisons la bibliothèque JavaScript MathJax pour visualiser les équations saisies en LaTeX dans une page HTML.

Chaque système est indépendant, mais pour réaliser une édition critique, nous aurions besoin d'insérer des balises XML-TEI à l'intérieur même du code LaTeX, ce que les outils actuels ne savent pas gérer.

L'objectif de cette journée d'étude est d'offrir un état de l'art sur la question de l'encodage et de la visualisation des mathématiques à partir de l'usage du plugin Transcript sur la plateforme EMAN.

Nous aborderons essentiellement trois sujets :

- les différentes solutions aux questions posées par cette cohabitation entre deux langages d'encodage et de visualisation ;
- les pistes possibles pour l'édition critique numérique de textes contenant des mathématiques, en particulier pour l'insertion d'apparat critique dans les formules mathématiques (idéalement encodées en LaTeX) ;
- comment, en TEI, intégrer les formules mathématiques, non seulement pour une visualisation correcte des textes, mais aussi pour que ceux-ci soient exploitables et indexables.

---

---

## Programme

**10h-10h15** : Accueil

**10h15-10h30** : La plateforme EMAN et la transcription, une histoire complexe, Richard Walter (EMAN, Thalm)

**10h30-11h30** : Encoder des maths et du texte en XML-TEI sur Transcript : état de l'art, guidelines et pistes, Candice Lecoffre (projet *DWR*)

**11h30-12h30** : Retour d'expériences par les projets *Dedekind*, *DWR* et *Manuscrits Leibniz*

### Déjeuner

**14h-14h45** : Encodage savant en TEI de variantes d'un texte, Julie Giovacchini (Centre Jean Pépin)

**14h45-16h** : Travaux pratiques et discussions : quels encodages pour quels besoins.

Amenez vos exemples à encoder.

**16h-16h15** : Conclusion et perspectives