

FUTURS FANTASTIQUES

Vendredi 10 décembre 2021

Grand Auditorium

Session de questions

Question:

Yves Citton:

I was wondering, in the different projects, what would be the part of what I didn't have time to address in my presentation, the notion of heteromation? I talked a lot about automation which is delegating to machines work that used to be made by men, but some people proposed the notion of heteromation to reveal the human work which is involved in the artificialization of operation. We saw that in all the projects part of it: you had to check, you had to do things, you had to correct, you had to put things... How much of it would be done through mechanical turk or through this type of platform, and how much of it would be done by otherwise students or researchers themselves etc. so the sort of the proportion of mechanical turn of platform capitalism versus people's devotion or people being paid by a research grant to do that work?

Answers:

Aurélia Rostaing:

Je peux tenter une première réponse. Je pense que pour beaucoup de projets il n'y aurait pas les financements : c'est aussi pour cela qu'on recourt à l'intelligence artificielle. Au départ, il faut de l'huile de coude, enfin il faut du travail humain important pour produire les vérités terrain, et ensuite ce travail va en diminuant, il consiste uniquement à corriger des données obtenues grâce à des modèles et l'idée c'est qu'à terme les modèles soient suffisamment robustes, suffisamment fiables, pour qu'on puisse se passer d'interventions humaines.

Jolan Wuyts :

Yeah I think my answer would be that it's very dependent on the task itself. Of course you know what do you actually want to achieve we did look at the option of as you said the platform capitalism just you know outsourcing this to another group of people. Our conclusion there was basically that it left us with little to no control over knowing exactly where this came from. Of course you would know it's been outsourced and you can then do some testing on the what you get back to see where the biases are how this was created but as long as you keep the annotators as close to your own knowledge base as possible you really have a lot more control over how exactly these datasets are being created. And I think it also just speaks to the

transparency of what you're doing and being able to make this open source, being able to show your work as much as possible that ideally you would keep annotators as close to your own field of work as possible I do of course appreciate that's just not possible for everyone and as you said there's a lot of finances involved in that and time and resources. So yeah, thank you.

Catherine Cronin:

I'd like to speak for heritage made digital. When we were embarking on this project we knew that it had to be actually low impact on staff resource, we already had staff members working with collection items and identifying these fake fly sheets, so as part of their work it was quite easy to then get them to feed the images they were finding into the training model, and then there was some resource spent on looking at the results, you know, and seeing how good they were, how well they were predicting things and then actually bringing that tool just back into the workflow really quickly, so in terms of setting this flyswot tool up for us, we did want it to be low impact on the work that was being done but it already fitted into our workflows and I think at the moment we're still working with the tool and improving it and we do want to gather some statistics on the difference between how much time now is this tool saving in staff resource compared to if you had a staff member who had to just work with image content and didn't have this tool. Yeah I hope that answers the question from our point of view.

Peter Broadwell:

Okay. And I would also just say the piano rolls is kind of a special case because it is, like I said, a prehistoric not prehistoric but a pre-digital computational technology and most of the approaches we're using work quite well up to a very close asymptote of perfect, so the question becomes then do we build an entire training apparatus and engage either more experts or possibly web workers to refine that to get that last five percent of accuracy and that's still a question where we're struggling with.

Question:

On va prendre une dernière question pour ne pas trop amputer sur le temps de pause.

J'avais une question, bonjour, pour les collègues qui travaillent autour des archives notariales dans le minutier central. Quand on étudie des archives de notaires en approche diplomatique on se dégage des modèles et ensuite en revanche, on parlait ce matin d'autorité bibliographique, on a des problématiques de noms qui varient y compris donc d'orthographe mais également de langue, c'est-à-dire qu'un même acte va être rédigé en latin puis dans une langue un peu plus locale etc. en fonction du notaire qui agit, alors c'est peut-être pas le cas au minutier central mais plus généralement dans les archives de notaires, donc est-ce que dans l'approche de l'apprentissage de la machine vous parvenez à faire ces rapprochements

linguistiques et à faire un travail autour des autorités de personne parce que c'est ce à quoi on se confronte beaucoup y compris dans les inventaires quand les gens ne sont pas nommés de la même manière, d'un centre d'archive à l'autre, ou même au sein d'un inventaire d'une même archive. Merci beaucoup.

Answer:

Aurélia Rostaing :

J'ai un peu envie de botter en touche face à cette question difficile... En fait pour la reconnaissance des entités nommées nous n'avons fait qu'une expérimentation, le projet s'est surtout concentré sur l'HTR. Mais le service interministériel des archives de France a lancé avec l'entreprise Teklia un projet qui s'appelle Socface et qui porte sur des registres de recensement de population de toute la France conservés dans des archives départementales et municipales et l'un des défis auxquels Teklia devra répondre sera d'aligner de 5 ans en 5 ans, période des recensements, les noms des personnes. Alors on est sur des archives contemporaines essentiellement du 20ème siècle sur une centaine d'années, à cette époque-là les patronymes, la transcription est codifiée donc on n'a pas les problèmes que vous évoquez pour l'Ancien Régime par exemple. Mais effectivement c'est un vrai problème parce qu'on a énormément d'homonymes sur 150 ans d'archives notariales et sur une ville aussi peuplée que Paris, après il faudra trouver des moyens de désambiguïsation avec notamment les adresses qu'on a dans les répertoires.

Je ne veux pas répondre trop longuement pour laisser un temps de pause. Voilà.