

[COVID Information Commons \(CIC\) Research Lightning Talk](#)

Transcript of a Presentation by Alan Porter ( Search Technology, Inc. and Georgia Institute of Technology),  
January 2022



Title: [Exploring Causes and Cures for COVID-19 through Improved Access to Biomedical Research](#)

[Alan Porter CIC Profile](#)

NSF Award #: [2029673](#)

[YouTube Recording with Slides](#)

[January 2022 CIC Webinar Information](#)

Transcript Editor: Julie Meunier

---

Transcript

*Slide 1*

Alan Porter:

Merci à tous. Je souhaitais partager quelques résultats d'un projet NSF RAPID que nous avons mené. Dont le titre est ici - et il s'agit essentiellement d'exploiter la littérature biomédicale pour tenter d'obtenir un accès pratique et des perspectives différentes.

*Slide 2*

Voici un aperçu de mon parcours. Je n'y reviendrai pas. En gros, nous effectuons une veille technologique - nous essayons de rassembler des données et d'en extraire des connaissances utilisables.

*Slide 3*

Un projet particulier consistait à explorer les causes et les traitements du COVID en améliorant l'accès à la recherche et j'aimerais aborder trois éléments. Premièrement, nous faisons appel à une bibliométrie intelligente pour essayer d'extraire des informations de ces textes abstraits. Le second est un aperçu d'un système de recommandation qui permet de faire des recherches dans la littérature au-delà de l'ensemble des données que nous analysons actuellement. Le troisième est un tableau de bord pour partager les résultats.

*Slide 4*

Notre abréviation pour cela est "tech mining". Les données dont nous disposons ici sont issues de la stratégie de recherche principale COVID de la National Library of Medicine. Nous avons utilisé cette

stratégie pendant un an et demi. Je parlerai de deux sous-ensembles différents. Le premier comprend environ 60,000 enregistrements jusqu'au début du mois d'octobre 2020, sur lesquels nous avons effectué une grande partie du travail de bibliométrie intelligente, et le second est la version la plus récente du tableau de bord, à partir du 1er janvier, pour vous montrer à quoi il ressemble actuellement.

#### *Slide 5*

D'accord, premier élément : l'exploration technologique est une analyse textuelle des ressources de données relatives à la science, à la technologie et à l'innovation, généralement sous la forme de résumés de publications de recherche ou de PAT, le plus souvent. Certaines des choses que nous avons faites ici - l'une d'entre elles consiste à rechercher dans le corpus de cette recherche les sujets qui font l'objet d'une attention accrue et à essayer de le faire en temps opportun. Je n'en parlerai pas aujourd'hui. Deuxièmement, nous suivons les sujets dans le temps.

#### *Slide 6*

Voici un aperçu de cette étude menée par notre collègue Yi Zhong de l'Université de technologie de Sydney. L'un d'entre eux tire des conclusions en regroupant l'utilisation des termes, ou quels sont les groupes de sujets d'intérêt au fil du temps ?

#### *Slide 7*

Le processus consiste essentiellement à prendre une période et à regrouper les sujets sur la base de la cooccurrence des termes, puis à procéder de la même manière pour des périodes successives, et enfin à examiner les liens entre ces sujets. J'ai ici quelques références que je ne citerai pas, mais qui sont indiquées à la fin de la présentation.

#### *Slide 8*

Voici un zoom sur un petit morceau. Nos dernières analyses, datant de 2020, ont montré que la propagation rapide était un sujet d'intérêt. Et cela a eu des origines dans le traitement très récent des maladies infectieuses - un peu plus tôt que *les études qui notent le temps*. Ensuite, nous sommes revenus sur certains des coronavirus précédents - avec certains travaux avec la PCR de 2006. Jusqu'à remonter au SRAS de 2003. Il est donc possible d'établir des liens entre les sujets et le temps.

#### *Slide 9*

Notre deuxième élément - la découverte basée sur la littérature - est utilisé ici pour un système de recommandation très simple. Voyez s'il y a un sujet qui vous intéresse dans la littérature COVID. Y a-t-il des recherches utiles pour vous en dehors de cette littérature ?

#### *Slide 10*

Le processus consiste à générer un ensemble de données, à titre d'illustration simple, à partir d'une trentaine de thèmes de comorbidités, dont celui que nous venons de sélectionner. Ensuite, dans la littérature COVID, dans ce cas 60 000 enregistrements, nous recherchons les documents les plus liés à ce sujet. Nous procédons ensuite à des analyses de texte pour nettoyer les textes de la littérature sélectionnés, et nous recherchons les termes dont la fréquence est particulièrement élevée dans ces

documents et, inversement, ceux dont la fréquence est particulièrement faible, ce que nous appelons un modèle de connaissances. *Ensuite, nous calculons la fréquence de l'entreprise [et] la fréquence inverse du document pour ces termes dans l'ensemble des données PubMed Medline.* Ce ne sont pas les mêmes [inaudible]. Et pour identifier les résumés d'articles susceptibles d'intéresser quelqu'un qui étudie un sujet particulier, comme la comorbidité.

#### *Slide 11*

Voici une petite illustration de trois des thèmes des comorbidités pour lesquels nous avons ajouté le nombre de citations comme deuxième critère. Ainsi, si vous souhaitez être cité et que vous utilisez des termes similaires à ceux utilisés pour les comorbidités dans la littérature COVID, [ceci] pourrait vous intéresser. Jetez-y un coup d'œil et il pourrait s'agir de trois articles, ou nous pourrions aller jusqu'à des charges de 33 millions, ce que personne ne voudrait, je pense.

#### *Slide 12*

Le dernier élément que je voulais aborder est l'élaboration de notre tableau de bord. Voici le tableau de bord du projet et, à l'intérieur de celui-ci, je mentionne juste une cible de vaccination que nos collègues de BizInt ont mise au point, qui est intéressante à suivre dans le temps. Je vais maintenant jeter un coup d'œil à notre tableau de bord PubMed. Il met en évidence les 153 - 150 000 enregistrements de résumés au 1er janvier. Un tutoriel de quelques minutes explique comment l'utiliser. Voici quelques-uns des différents champs de données disponibles. Permettez-moi de faire une petite démonstration rapide - voici une version où nous pouvons aller et où nous pouvons repérer quelque chose d'intéressant et je vais juste - essayer de le mettre en évidence. Et je pense que nous obtenons une réponse lente, ici. Recherche en Iran - nous pourrions parcourir ce document et voir ce que font certaines personnes là-bas en ce qui concerne les comorbidités. Il se peut même que nous repérions quelque part un article qui suscite notre intérêt. Et ici, nous pouvons faire apparaître son résumé. Pour aller plus vite, il y a une autre illustration. Si nous voulions mettre en évidence le travail le plus récent, il s'agirait d'un article de décembre 2021. Ce n'est pas complet car le codage, la catégorisation, le catalogage, etc. sont encore en cours, mais nous pourrions aller ici et choisir ce qui pourrait nous intéresser. [Pause la vidéo] À Columbia - et voici quelques-uns des 46 articles associés à un auteur de Columbia au cours du dernier mois. Voilà donc notre démonstration - si quelqu'un veut aller sur le tableau de bord, nous en serions ravis. Il s'agit d'un travail en cours.

#### *Slide 13*

Et j'ai pratiquement terminé. Et voilà - Ressources et Finition. Merci de votre attention!