



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

Title: बायोमेडिकल अनुसंधान तक बेहतर पहुंच के माध्यम से COVID-19 के कारणों और उपचारों की खोज

[Alan Porter CIC Profile](#)

NSF Award #: [2029673](#)

[YouTube Recording with Slides](#)

[January 2022 CIC Webinar Information](#)

Transcript Editor: Saanya Subasinghe

प्रतिलिपि

एलन पोर्टर:

स्लाइड 1

सभी को धन्यवाद, मैं एन.एस.एफ. रैपिड प्रोजेक्ट से कुछ परिणाम साझा करना चाहता था जो हमने किया। यहां शीर्षक के साथ - और यह मूल रूप से बायोमेडिकल साहित्य को सुविधाजनक तरीके से और कुछ अलग दृष्टिकोणों तक पहुंचने का प्रयास करने के तरीकों से खनन कर रहा है।

स्लाइड 2

यहाँ मेरी पृष्ठभूमि है. मैं उससे नहीं गुजरूंगा. मूल रूप से हम प्रौद्योगिकी पर नजर रखने वाले हैं - डेटा इकट्ठा करने और उससे कुछ उपयोगी ज्ञान निकालने का प्रयास कर रहे हैं।

स्लाइड 3

यहां एक विशेष परियोजना अनुसंधान तक बेहतर पहुंच के माध्यम से सीओवीआईडी के कारणों और इलाज की खोज कर रही थी और मैं केवल तीन तत्वों पर बात करना चाहता हूँ। एक है [जिसे] हम इन अमूर्त पाठों से जानकारी निकालने का प्रयास करने के लिए बुद्धिमान बिब्लियोमेट्रिक्स कह रहे हैं। दूसरा, जिस डेटा सेट का हम वास्तव में विश्लेषण कर रहे हैं उससे परे कुछ साहित्य आधारित खोज करने के लिए अनुशंसा प्रणाली पर एक नज़र डालना है। और तीसरा परिणाम साझा करने के लिए एक डैशबोर्ड है।

स्लाइड 4

इसके लिए हमारा शॉर्टहैंड 'टेक माइनिंग' है। यहां उपलब्ध डेटा, मुख्य नेशनल लाइब्रेरी ऑफ मेडिसिन की, एक तरह की, कोर सीओवीआईडी खोज रणनीति का उपयोग कर रहा है। हमने इसे सीधे डेढ़ साल तक इस्तेमाल किया। और मैं दो अलग-अलग उपसमूहों के बारे में बात करूंगा। एक अक्टूबर 2020 की शुरुआत तक लगभग 60,000 रिकॉर्ड हैं, जिस पर हमने बहुत सारे बुद्धिमान ग्रंथसूची प्रकार का काम किया है और दूसरा 1 जनवरी से डैशबोर्ड का सबसे हालिया संस्करण है - आपको यह दिखाने के लिए कि यह अभी कैसा दिखता है।

स्लाइड 5

ठीक है, तत्व नंबर एक: तकनीकी खनन विज्ञान प्रौद्योगिकी और नवाचार डेटा संसाधनों का पाठ विश्लेषण है, जो आमतौर पर शोध प्रकाशन या पीएटी के सार रिकॉर्ड के रूप में होता है। कुछ चीजें जो हमने यहां की हैं - एक उस शोध के मुख्य भाग के भीतर देखना है कि किन विषयों पर ध्यान तेजी से बढ़ता है और उसे समय पर करने का प्रयास करें। मैं आज उस बारे में बात नहीं करूंगा। दूसरा है समय के साथ विषयों पर नज़र रखना।

स्लाइड 6

प्रौद्योगिकी विश्वविद्यालय, सिडनी में हमारे सहयोगी यी झोंग के नेतृत्व में इसकी एक झलक यहां दी गई है। एक उपयोग शब्द को क्लस्टर करके निकाला जाता है या समय के साथ रुचि के कुछ सामयिक समूह क्या हैं?

स्लाइड 7

प्रक्रिया, संक्षेप में, शब्दों की सह-घटना के आधार पर विषयों को एक समय अवधि में समूहित करना है, और फिर क्रमिक अवधियों में ऐसा करना है, और फिर उन विषयों के बीच लिंक को देखना है। और मुझे यहां कुछ संदर्भ मिले हैं, मैं ज्यादा मेहनत नहीं करूंगा, लेकिन उन्हें प्रेजेंटेशन के पीछे टैग किया गया है।

स्लाइड 8

यहां इसके एक छोटे से टुकड़े पर जूम इन किया गया है। 2020 में हमारे नवीनतम विश्लेषणों से पता चला कि तेजी से प्रसार रुचि का विषय था। और इसमें संक्रामक रोगों के हाल के उपचार में रुकावट थी - समय को ध्यान में रखते हुए अध्ययन से थोड़ा पहले। और फिर पिछले कुछ कोरोनाविरस पर वापस जा रहे हैं - कुछ 2006 एकाग्रता में पीसीआर के साथ काम करते हैं। 2003 में SARS का पता लगाया जा रहा है। इसलिए कोई भी विषयों और समय के आधार पर कुछ लिंक कर सकता है।

स्लाइड 9

हमारा दूसरा तत्व - साहित्य-आधारित खोज - और यहां हमने इसे एक बहुत ही सरल अनुशंसा प्रणाली के लिए उपयोग किया है। देखें कि क्या COVID साहित्य में कोई ऐसा विषय है जिसमें आपकी रुचि है। क्या उस साहित्य के बाहर आपके लिए कुछ उपयोगी शोध हो सकता है?

स्लाइड 10

और यहां प्रक्रिया - हम लगभग 30 या अधिक विषयों से सह-रुग्णता वाले एक साधारण चित्रण के रूप में एक डेटासेट तैयार करते हैं, जिसे हमने अभी यहां चुना है। और फिर कोविड साहित्य के भीतर, इस मामले में यह 60,000 रिकॉर्ड है, उससे संबंधित अधिकांश दस्तावेजों की तलाश की जा रही है। फिर उस साहित्य को साफ करने के लिए वहां कुछ पाठ विश्लेषण करें जो इसका पता लगाता है, और उन दस्तावेजों में किन शब्दों की विशेष रूप से उच्च आवृत्ति है, और इसके विपरीत विशेष रूप से कम आवृत्ति है, इसे ज्ञान मॉडल कहते हैं। फिर बाहर जाकर संपूर्ण पबमेड मेटलाइन डेटासेट में उन शब्दों के लिए फर्म आवृत्ति [और] व्युत्क्रम दस्तावेज आवृत्ति की गणना को देखना। वे समान नहीं हैं

[अश्रव्य]। और उन लेखों के सार की पहचान करना जो सह-रुग्णता जैसे किसी विशेष विषय पर काम करने वाले किसी व्यक्ति के लिए रुचिकर हो सकते हैं।

स्लाइड 11

यहां सहरुग्णताओं के तीन विषयों का एक छोटा सा चित्रण दिया गया है जहां हमने दूसरे मानदंड के रूप में उद्धरण गणना में जोड़ा है। इसलिए यदि आप उद्धृत किए जाने में उच्च हैं और आप सीओवीआईडी साहित्य में सह-रुग्णता के उपयोग के समान शब्दों का उपयोग कर रहे हैं, तो [यह] आपके लिए रुचिकर हो सकता है। शायद इन पर एक नज़र डालें और यह तीन लेख हो सकते हैं, या हम 33 मिलियन पर लोडिंग तक जा सकते हैं, जो मुझे नहीं लगता कि कोई भी चाहेगा।

स्लाइड 12 - <https://sites.google.com/searchtech.com/covidproject/home>

आखिरी तत्व जिस पर मैं गौर करना चाहता था वह हमारे डैशबोर्ड बिल्ड-अप पर नज़र डालना था। और यहां प्रोजेक्ट डैशबोर्ड है और उसके भीतर मैं बस एक वैक्सीन बुल्सआई का उल्लेख कर रहा हूं जिसे बिज़िंट में हमारे सहयोगियों ने एक साथ रखा है, जिसे समय के साथ ट्रैक करना दिलचस्प है। मैं हमारे पबमेड डैशबोर्ड पर एक नज़र डालने जा रहा हूँ। और वह अब 1 जनवरी तक के 153 - 150,000 अमूर्त रिकॉर्ड पर प्रकाश डाल रहा है। और वहां एक डेमो है, कुछ मिनिटों में, यह समझाने के लिए कि इसका उपयोग कैसे करना है। यहां डेटा के कुछ अलग-अलग क्षेत्र उपलब्ध हैं। और मुझे बस एक त्वरित पॉप-इन करने दीजिए, यहां एक संस्करण है जहां हम जा सकते हैं और हम कुछ दिलचस्प चीजें देख सकते हैं और मैं बस करने वाला हूँ - स्पॉटलाइट करने का प्रयास करें। और मुझे लगता है कि हमें यहां कुछ सुस्त प्रतिक्रिया मिल रही है। ईरान में अनुसंधान - हम वहां जा सकते हैं और देख सकते हैं कि वहां कुछ लोग सह-रुग्णताओं के संबंध में क्या कर रहे हैं। और हमें यहां कहीं कोई ऐसा लेख भी मिल सकता है जो हमारी रुचि को आकर्षित करता है। और यहां हम उस अमूर्त रिकॉर्ड को पॉप अप कर सकते हैं। तो चीजों को जल्दी से करने के लिए, बस एक और उदाहरण है। यदि हम सबसे हालिया काम पर प्रकाश डालना चाहते हैं और वह पिछले 2021 के दिसंबर से है। यह अधूरा है क्योंकि कोडिंग, वर्गीकरण, कैटलॉगिंग इत्यादि अभी भी चल रहा है, लेकिन हम यहां फिर से जा सकते हैं और जो रुचिकर हो सकता है उसे चुन सकते हैं। [फ्रीज़िंग वीडियो] कोलंबिया में - और पिछले महीने में कोलंबिया के एक लेखक से जुड़े 46 लेखों में से कुछ हैं। तो यह हमारा डेमो है- अगर कोई डैशबोर्ड पर जाना चाहता है तो हमें अच्छा लगेगा। इस पर अभी भी काम चल रहा है।

स्लाइड 13

और मेरा मूलतः काम पूरा हो गया है। हम वहां जाते हैं - संसाधन और समापन। ध्यान देने के लिए आपको धन्यवाद!