

Transcript of a Presentation by Carlos Badenes-Olmedo (Universidad Politecnica de Madrid), July 26, 2023



Title: [ड्रग्स4कोविड: कोरोनावायरस के नैदानिक नियंत्रण में उपयोग की जाने वाली दवाओं के बारे में ज्ञान ग्राफ](#)

NIH Publication:

<https://pubmed.ncbi.nlm.nih.gov/37156393/>

[YouTube Recording with Slides](#)

[Summer 2023 CIC Webinar Information](#)

Transcript Editor: Shikhar Johri

प्रतिलिपि

स्लाइड 1

कार्लोस बेडेनेस-ओल्मेडो:

ठीक है, धन्यवाद, लॉरेन। और मुझे आमंत्रित करने के लिए धन्यवाद, फ्लोरेंस। ठीक है, मैं अपनी स्क्रीन साझा करने जा रहा हूँ। तो, इस वेबिनार में भाग लेने के लिए मुझे आमंत्रित करने के लिए धन्यवाद। मैं कार्लोस बाडेनोस-ओल्मेडो, ओन्टोलॉजी इंजीनियरिंग ग्रुप का एक शोधकर्ता हूँ और मैं यूनिवर्सिटी डे पोलिटेकनिका डी मैड्रिड में एक सहायक लाभ प्रोफेसर भी हूँ। मैं ड्रग्स4कोविड प्रोजेक्ट प्रस्तुत करने जा रहा हूँ।

यह एक परियोजना है जिसमें हम एक ज्ञान ग्राफ बनाने का प्रस्ताव रखते हैं। यह उन दवाओं का प्रतिनिधित्व करने का प्रयास कर रहा है जिनका उपयोग हमने महामारी के दौरान कोरोनावायरस के नैदानिक नियंत्रण को परिभाषित करने के लिए किया था।

स्लाइड 2

विचार यह है कि कोविड-19 महामारी के दौरान, दुनिया भर के संस्थानों ने वैज्ञानिक अनुसंधान लेखों के साथ डेटा सेट की पहचान करने या बनाने की कोशिश की जो कोरोनावायरस से संबंधित हैं। इस प्रकार की जानकारी अस्पताल फार्मेशियों, नैदानिक चिकित्सकों के लिए उपयोगी हो सकती है। हमारा अनुसंधान समूह, ओन्टोलॉजी इंजीनियरिंग ग्रुप, यह प्रचार करने का प्रयास कर रहा था कि हम इस प्रकार के डेटा से जानकारी कैसे प्राप्त कर सकते हैं - ज्ञान प्राप्त कर सकते हैं। पहला कदम यह पहचानना था कि सबसे महत्वपूर्ण डेटासेट कौन सा है जिसका हम उपयोग कर सकते हैं। इस मामले में, यूरोपीय संघ से कोविड-19

यूरोपीय डेटा पोर्टल, साथ ही [अश्रव्य] कोविड डेटा सेट भी प्रदान किया गया था। मुख्य रूप से, कोविड-19 ओपन प्रीसेट डेटा सेट सबसे महत्वपूर्ण डेटा सेट था जो पब मेड BioRxiv, MedRxiv और arXiv जानकारी से जानकारी को जोड़ता है। यह विश्व स्वास्थ्य संगठन की जानकारी को भी जोड़ता है और 400,000 से अधिक लेख प्रदान करता है जिनकी हमें आवश्यकता है - हमने सोचा कि हम ज्ञान प्रदान करने का लाभ उठा सकते हैं। इस प्रकार के ज्ञान को मैट्रिड की स्वयं-सेवा के साथ जोड़ना, [अश्रव्य]। हम ज्ञान ग्राफ बनाने के लिए तंत्र प्रदान करते हैं जिसका हम लाभ उठा सकते हैं और इस प्रकार के लोगों से ज्ञान प्रदान कर सकते हैं।

स्लाइड 3

इसलिए, सबसे पहले, हम अलग-अलग चरणों के साथ एक वर्कफ़्लो को परिभाषित करते हैं जो न केवल चरण प्रदान करता है बल्कि उन चरणों की अनुशंसा भी करता है जिनका पालन करके आप अंतिम विश्लेषण ग्राफ़ बना सकते हैं और शोषण की सुविधा भी प्रदान कर सकते हैं। हम छह चरणों वाले वर्कफ़्लो को परिभाषित करते हैं। पहला कदम कटाई है। इस चरण में विचार यह है कि आपको उस डेटा सेट की पहचान करने की आवश्यकता है जो कोरोनावायरस से संबंधित है, और यह भी मूल्यांकन करना है कि डेटा पूरी तरह से उपलब्ध है या नहीं। दूसरा चरण प्री-प्रोसेसिंग है क्योंकि आपको प्रदान किए गए डेटा को व्यवस्थित करने की आवश्यकता है, इस मामले में एक डेटा सेट, क्योंकि यह डेटा कुल नहीं है क्योंकि हम परीक्षणों के साथ काम कर रहे हैं। हमें डेटा के [मूल्यांकन] के एक निश्चित तरीके को मॉडल बनाने की आवश्यकता है। फिर, हम तीसरे चरण की ओर बढ़ते हैं, जो सूचना निष्कर्षण है। इस चरण में हमें इस प्रकार के डेटा के मुख्य तत्वों की खोज करने की आवश्यकता है जिसका उपयोग हम विश्लेषण ग्राफ़ बनाने के लिए कर सकते हैं। हमारे मामले में, मुख्य तत्व बायोमैडिकल अवधारणाएं हैं (उदाहरण के लिए, दवाएं, रोग और आनुवंशिक जानकारी)। फिर, हमें डोमेन के औपचारिक विवरण को परिभाषित करने की आवश्यकता है, जो कि चौथा चरण है, जो कि शब्दार्थ है। इस मामले में, हमें बायोमैडिकल अवधारणाओं के बीच संबंधों को परिभाषित करने और सभी अवधारणाओं, सभी तत्वों को परिभाषित करने के लिए एक ऑन्कोलॉजी बनाने की आवश्यकता है, जो अंततः ज्ञान ग्राफ में दिखाई देते हैं। अगला चरण ज्ञान ग्राफ निर्माण है। इस चरण में हमें ज्ञान ग्राफ में उदाहरण बनाने के लिए नियमों को परिभाषित करने की आवश्यकता है। और अंत में हम ज्ञान ग्राफ में मौजूद जानकारी के उपयोग को सुविधाजनक बनाने के लिए - शोषण के लिए तंत्र प्रदान कर सकते हैं।

स्लाइड 4

इसलिए हम पहले चरण पर ध्यान केंद्रित करते हैं। इसका उद्देश्य प्रासंगिक डेटा स्रोतों की पहचान करना और डेटा की उपलब्धता का मूल्यांकन करना भी है। हमारा प्रस्ताव यह है कि आपको कोरोना वायरस की मुख्य अवधारणाओं को ध्यान में रखते हुए एक व्यवस्थित साहित्य समीक्षा करने की आवश्यकता है, और फिर डेटा सेट को परिभाषित करना होगा। उदाहरण के लिए, डिजिटल रिपॉजिटरी, पब मेड, BioRxiv इत्यादि से, लेकिन अन्य स्रोतों के साथ संयोजन भी, उदाहरण के लिए, समन्वय कोष से नैदानिक संग्रह, प्रमुख कोविड डेटासेट और अतिरिक्त संसाधन भी। उदाहरण के लिए, विकिपीडिया से पेटेंट, विश्वकोश लेख। और यह सारा डेटा इस पहले चरण में व्यवस्थित किया गया है।

स्लाइड 5

अगले चरण में, हमें इस असंरचित डेटा, जो कि टेक्स्ट है, को तालिकाओं में स्थानांतरित करना होगा, जो संरचित डेटा हैं। फिर, हम जो पद्धति प्रस्तावित करते हैं वह आपके लिए आवश्यक न्यूनतम जानकारी की पहचान करना है। टेक्स्ट को संरचित तरीके से बदलने का सबसे आसान तरीका लेख के पूर्ण टेक्स्ट को डेटा के रूप में परिभाषित करना है। हमारी राय में, ऐसा करने का यह सबसे अच्छा तरीका नहीं है और हमारा

प्रस्ताव है कि आपको जितनी न्यूनतम जानकारी की आवश्यकता हो, उसे परिभाषित किया जाए: लेखों का पैराग्राफ। उस क्षेत्र में आप सभी संदर्भों या बायोमैडिकल अवधारणाओं के बीच संबंध की खोज कर सकते हैं।

स्लाइड 6

फिर अगला चरण सूचनात्मक संरचना है। इस मामले में, विचार उन पैराग्राफों के आधार पर एनोटेशन बनाने का है जो दवाओं, बीमारियों और आनुवंशिक जानकारी की खोज करते हैं। अपने अनुभव में, हम प्रत्येक अलग-अलग बायोमैडिकल अवधारणा के लिए अलग-अलग भाषा मॉडल को ठीक करते हैं। विचार यह है कि आपको दवाओं की पहचान करने के लिए और विभिन्न शब्दावली के अनुसार दवाओं को सामान्य बनाने के लिए एक विशिष्ट भाषा मॉडल को परिभाषित करने की आवश्यकता है क्योंकि विभिन्न देशों में हम विभिन्न मानक कोड का उपयोग करते हैं।

स्लाइड 7

एक बार जब हमारे पास संस्थाओं के साथ एनोटेशन और कोड भी हो जाएं तो हम इस सारी जानकारी का वर्णन करने के लिए औपचारिक स्थान को परिभाषित कर सकते हैं। यह वह कदम है जिसकी हमें ऑन्टोलॉजी बनाने के लिए आवश्यकता है। बायोमैडिकल डोमेन में बहुत सारे ऑन्कोलॉजी मौजूद हैं, इसलिए विचार शुरू से ही ऑन्टोलॉजी बनाने का नहीं है। इसका उद्देश्य सिस्टम ऑन्टोलॉजी को कम करना और नई ऑन्टोलॉजी में गुम जानकारी प्रदान करना है। हमारे मामले में ऑन्कोलॉजी ईबीओसीए थी और गायब जानकारी ऐसे साक्ष्य प्रदान करने के लिए थी जो दवाओं के बीच, बीमारियों के बीच और आनुवंशिक जानकारी के बीच संबंधों का समर्थन करते हैं। हमारे मामले में हमने एकीकृत चिकित्सा भाषा प्रणाली और डी.आई.एस.एन.ई.टी. प्लेटफॉर्म का भी उपयोग किया। ये सारी जानकारी संयुक्त है। हम यह भी प्रदान करते हैं, यह बैंगनी क्षेत्र है, साक्ष्य के बारे में जानकारी। कौन सा सबूत है? इसका प्रमाण वह पैराग्राफ है जहां वैज्ञानिक लेख में इन तत्वों के बीच संबंध बताया गया है।

स्लाइड 8

एक बार जब हम औपचारिक डोमेन, ऑन्कोलॉजी को परिभाषित कर लेते हैं, तो हमें ज्ञान ग्राफ में उदाहरण बनाने के लिए वैज्ञानिक लेखों से प्राप्त उदाहरणों, दावों, बयानों की पहचान करने की आवश्यकता होती है। यह ज्ञान ग्राफ निर्माण चरण है इसलिए हमारी कार्यप्रणाली में पिछली मॉडल भाषा का उपयोग करके संस्थाओं और उनके बीच संबंधों की पहचान करने के लिए नियम बनाने का प्रस्ताव है। अंत में, एक बार जब हमारे पास ऑन्कोलॉजी हो जाती है, तो हमारे पास उदाहरण होते हैं, हम ग्राफ में नोड्स की पहचान करने में सक्षम होते हैं। उदाहरण के लिए, नीले वाले तत्व हैं, नारंगी वाले उनके बीच के संबंध हैं, और बैंगनी वाले सबूत हैं जो इस प्रकार के संबंधों का समर्थन करते हैं। साक्ष्य न्यूनतम सूचना इकाई है जो पैराग्राफ और लेख भी है।

स्लाइड 9

एक बार जब हमारे पास ज्ञान ग्राफ हो जाता है, तो अंततः, हम जानकारी के दोहन की सुविधा के लिए उपलब्ध होते हैं। सबसे अच्छा - पहला कदम निश्चित रूप से हम एस.पी.ए.आर.क्यू.एल. क्वेरीज़ का उपयोग कर सकते हैं - यह एक विशिष्ट भाषा मॉडल है जिससे आप भाषा, विश्लेषण ग्राफ का फायदा उठाने के लिए क्वेरीज़ बना सकते हैं। इसके लिए इस तरह के डोमेन में विशेषज्ञ की आवश्यकता होती है।

स्लाइड 10

हमारी दूसरी पद्धति एक प्रश्न-उत्तर इंटरफ़ेस बनाना है जो न केवल ज्ञान ग्राफ से जानकारी प्रदान करता है बल्कि बाहरी स्रोतों से, दूसरों से और दूसरों से ज्ञान ग्राफ, दस्तावेज़ कनेक्शन और फिर उत्तर प्रदान करने के लिए प्राकृतिक भाषा में प्रश्नों का समर्थन भी करता है। प्राकृतिक भाषा में तो हमारा प्लेटफ़ॉर्म एक कोविड प्लेटफ़ॉर्म के लिए है।

स्लाइड 11

यह सारी जानकारी नीति के अनुसार उपलब्ध है: ज्ञान ग्राफ, मॉडल, डेटा सेट और सेवाएँ भी - ये संसाधन पूरी तरह से मुफ्त हैं, पूरी तरह से सार्वजनिक हैं। ये इन यूआरएल से उपलब्ध हैं। आपके ध्यान के लिए धन्यवाद और मैं आपके किसी भी प्रश्न का उत्तर देने में सक्षम हूँ।