



IEEE DataPort

กับชุดข้อมูล COVID-19



พ.ศ.ดร.ณัฐ ธีระวัฒน์

คณะวิศวกรรมศาสตร์ ภาควิชาวิศวกรรมไฟฟ้า
กลุ่มวิจัยระบบสารสนเทศการศึกษาคอมพิวเตอร์
และความเสี่ยง ภาควิชาวิศวกรรมไฟฟ้า
<http://natt.leelawat.com>

ใน เวลาวิกฤตสถานการณ์โรคระบาด COVID-19 ข้อมูล
ดิบต่างๆ นับว่าเป็นสิ่งจำเป็นสำหรับนักวิเคราะห์
ข้อมูล และนักวิจัย ด้วยเทคโนโลยีของวิทยาการคอมพิวเตอร์ และ
วิทยาการข้อมูลในปัจจุบันทำให้เรามีเครื่องมือมากมายให้ใช้ในการ
วิเคราะห์ และนำเสนอ ว่าแต่ข้อมูลอะไรจะสามารถนำมาใช้
วิเคราะห์ได้บ้าง และเราจะหาข้อมูลได้จากที่ไหนบ้าง บทความนี้
ผู้เขียนขอแนะนำเสนอหนึ่งในแหล่งข้อมูลดิบที่ท่านผู้อ่านสามารถ
ดาวน์โหลดมาวิเคราะห์กันได้เลย

DataPort

DataPort เป็นแพลตฟอร์มของ IEEE ที่ถูกออกแบบมา
สำหรับการเข้าถึงข้อมูลที่ทำให้ผู้ใช้งานสามารถเก็บชุดข้อมูล
เข้าถึงชุดข้อมูล วิเคราะห์ข้อมูล และนำไปสู่การสร้างงานวิจัยได้
(IEEE, 2021d)

สำหรับการเก็บข้อมูลนั้น ทางผู้ใช้งานสามารถเก็บข้อมูลได้
ทุกประเภท ทุกรูปแบบ โดยสามารถเก็บชุดข้อมูลได้มากที่สุดถึง 2TB
เลยทีเดียว ชุดข้อมูลที่ถูกเก็บไว้ใน DataPort จะสามารถถูกค้นหา
ผ่านคำสำคัญ และประเภทของชุดข้อมูลได้ เมื่อได้ชุดข้อมูลแล้ว
ผู้ใช้งานก็สามารถดาวน์โหลดชุดข้อมูล และอาจจะใช้เครื่องมือต่าง ๆ
ในการวิเคราะห์หรือรวมถึงการวิเคราะห์ข้อมูลบนคลาวด์ (IEEE,
2021d)

ประเภทหมวดหมู่ชุดข้อมูล

ในปัจจุบัน DataPort มีการแบ่งหมวดหมู่ของชุดข้อมูล
ทั้งหมด 27 หมวดหมู่ ได้แก่ ปัญญาประดิษฐ์, ดาราศาสตร์, ชีวเวช
และวิทยาศาสตร์สุขภาพ, สัญญาณชีวฟิสิกส์, การประมวลผล
แบบคลาวด์, การสื่อสาร, ปัญญาการคำนวณ, คอมพิวเตอร์
วิทัศน์, COVID-19, ข้อมูลประชากร, นิเวศวิทยา, สิ่งแวดล้อม,
การเงิน, ธรณีศาสตร์ และการสำรวจระยะไกล, การหลอมภาพ,
การประมวลผลภาพ, อินเทอร์เน็ตของสรรพสิ่ง, การเรียนรู้ของเครื่อง,
พลังงาน, ความน่าเชื่อถือ, ความปลอดภัย, เซ็นเซอร์, การประมวล
ผลสัญญาณ, สังคมศาสตร์, ข้อมูลวิจัยมาตรฐาน, การคมนาคม,
และหมวดอื่น ๆ (IEEE, 2021c) โดยการแบ่งหมวดหมู่เกิดจากการ
ใส่แท็กคำสำคัญนั่นเอง ดังนั้นจึงเป็นไปได้ว่าชุดข้อมูลเดียวกันอาจ
จะอยู่ในหมวดหมู่มากกว่าหนึ่งหมวดหมู่ก็เป็นได้

ชุดข้อมูล COVID-19

เมื่ออ่านถึงจุดนี้ ท่านผู้อ่านอาจจะสงสัยว่าแล้วชุดข้อมูล
COVID-19 จะมีหน้าตาเป็นอย่างไรกัน ในชุดข้อมูลหมวดหมู่นี้ก็มี
ทั้งข้อมูลภาพ ข้อมูลวิดีโอ ข้อมูลทวิต ข้อมูลสถิติ ข้อมูลข้อความ
ข้อมูลโพสต์ และอีกมากมาย (IEEE, 2021b) จะเห็นได้ว่าเป็น
ชุดข้อมูลทั้งแบบมีโครงสร้าง และไม่มีโครงสร้าง

ว่าแต่ชุดข้อมูลเหล่านี้หรือแม้แต่ชุดข้อมูลจากแหล่งอื่นๆ
นั้นจะมีการนำไปใช้งานได้อย่างไรกัน ผู้เขียนขอลองยกตัวอย่าง
ข้อมูลภาพ เช่น ข้อมูลภาพเอกซเรย์ปอดที่ได้รับเชื้อ COVID-19
ในลักษณะต่าง ๆ (Ahmed et al., 2021; Liu, 2021) ก็สามารถ
นำไปเป็นข้อมูลเรียนรู้สำหรับการเรียนรู้ของเครื่องเพื่อช่วยในการ
ระบุโรคหรืออาการสำหรับแพทย์ได้ ข้อมูลภาพเคลื่อนไหวของ
ละอองฝอย (Gore et al., 2020) อาจจะไปใช้ศึกษาการโอกาส
การสัมผัสเชื้อได้



สำหรับการวิเคราะห์ข้อมูลสื่อสังคมออนไลน์ก็สามารถนำไปเป็นข้อมูลตั้งต้นในการวิเคราะห์ความรู้สึกของผู้โพสต์ ผ่านอัลกอริทึมการเรียนรู้ของเครื่อง (Sontayasara et al., 2021) หรือการวิเคราะห์ข่าวปลอมด้วยการประมวลผลภาษาธรรมชาติ (Verma et al., 2020) ได้ เป็นต้น

และด้วยประโยชน์ที่กล่าวมาจึงทำให้ IEEE ได้ตัดสินใจเปิดหมวดหมู่ COVID-19 ขึ้นมาใน DataPort เพื่อที่จะทำให้นักวิจัยและสถาบันวิจัยสามารถเก็บ แบ่งปัน และเข้าถึงชุดข้อมูลเกี่ยวกับ COVID-19 ได้โดยไม่มีค่าใช้จ่าย ดังนั้นชุดข้อมูลในหมวดหมู่ COVID-19 จึงถูกเปิดให้เข้าถึงฟรีนั่นเอง (IEEE, 2021a)

บทสรุป

ถ้าท่านผู้อ่านสนใจอยากค้นหาชุดข้อมูลเพื่อนำไปวิเคราะห์สำหรับศึกษาหรือทำวิจัย ท่านผู้อ่านสามารถลงเข้าไปค้นหาชุดข้อมูล COVID-19 ได้ที่ <https://iee-dataport.org/access-covid-19-datasets> นะครับ

อ้างอิง

Ahmed, K. B., Goldgof, G. Paul, R., Goldgof, D., & Hall, L. (2021). COVID-19 and Pneumonia Xrays. IEEE Dataport. doi: <https://dx.doi.org/10.21227/cw6k-dc90>.

Gore, R., Saldana, C., Wright, D., & Klein, A. (2020). Intubation Aerosol Containment System Video. IEEE Dataport. <https://dx.doi.org/10.21227/f0qs-ct30>

IEEE. (2020a). Access COVID-19 Datasets. Retrieved May 2021, from <https://iee-dataport.org/access-covid-19-datasets>

IEEE. (2020b). COVID-19. Retrieved May 2021, from <https://iee-dataport.org/topic-tags/covid-19>

IEEE. (2020c). Dataset Categories. Retrieved May 2021, from <https://iee-dataport.org/datasets>

IEEE. (2020d). Why IEEE DataPort?. Retrieved May 2021, from <https://iee-dataport.org/why-ieee-dataport>

Liu, C. (2021). COVID-19Dataset. IEEE Dataport. <https://dx.doi.org/10.21227/16sk-kv49>

Sontayasara, T., Jariyapongpaiboon, S., Promjun, A., Seelpipat, N., Saengtattim, K., Tang, J., & Leelawat, N. (2021). Twitter sentiment analysis on Bangkok tourism during the COVID-19 situation using support vector machine algorithm. *Journal of Disaster Research*, 16 (1), 24-30.

Verma, S., Paul, A., Karayannavar, S. S., & Katarya, R. (2020). Understanding the Applications of Natural Language Processing on COVID-19 Data. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), (pp. 1157-1162) IEEE.