Motivation

ระบบการศึกษาถือว่าเป็นรากฐานสำคัญในการขับเคลื่อนประเทศชาติ ซึ่งในตอนนี้ก็มีความพยายามจากหลายๆ ฝ่ายที่จะพัฒนาระบบการศึกษาในประเทศไทยด้วยวิธีการต่างๆ โดยข้อมูลเหล่านั้นจะถูกบันทึกไว้ในรูปแบบเอกสารต่างๆ แต่เอกสารเหล่านั้นยังไม่ถูกนำมารวบรวมและจัดเก็บเพื่อการนำไปใช้อย่างแพร่หลายให้เกิดประโยชน์สูงสุดต่อระบบการศึกษาทั้งประเทศ จึงเป็นเหตุให้การศึกษาในประเทศไทยยังไม่เกิดความเท่าเทียมกันอย่างที่ควรจะเป็น ดังนั้น ทีมของเราจึงเกิดแนวคิดที่จะพัฒนา community สำหรับแลกเปลี่ยนเอกสารในการพัฒนาการเรียนสอนนักเรียน (pedagogy) โดยผู้ที่สนใจ ได้แก่คุณครูและบุคลากรที่ต้องการจะค้นหาเอกสารเหล่านี้ และผู้ที่มีเอกสารที่เกี่ยวข้องกับเรื่องเหล่านี้ก็สามารถอัพโหลดเอกสารที่ตนเองมีอยู่เพื่อแบ่งปันความรู้และประสบการณ์ที่ได้เรียนรู้มาให้กับคนอื่นๆ ใน community นี้ แต่เนื่องจากเอกสารที่จะถูกแชร์มาในระบบนั้นมีทั้งจากทางโรงเรียนที่ทำการสนับสนุนและบุคลากรที่ต้องการจะแบ่งปันความรู้เป็นจำนวนมาก ทำให้สิ้นเปลืองทรัพยากรในการจำแนกเอกสารเหล่านั้น ทางทีมเราจึงคิดที่จะพัฒนาระบบที่สามารถจำแนกเอกสารเหล่านั้น โดยใช้เทคนิค Machine Learning ในการอ่านเอกสารและทำการจำแนกไปยังกลุ่มต่างๆ เมื่อข้อมูลการเรียนการสอนเหล่านี้ สามารถเข้าถึงและนำไปใช้งานจริงในโรงเรียนต่างๆ ทั่วประเทศแล้ว ความการศึกษาของโรงเรียนต่างๆ ทั่วประเทศไทยก็จะพัฒนาขึ้น และเกิดความเท่าเทียมในโรงเรียนต่างๆ มากขึ้นเช่นเดียวกัน

โดยแนวคิดดังกล่าวข้างต้นนั้น จะมีรูปแบบคล้ายๆ กับโปรแกรม DeepDive ของ Stanford University ที่สามารถวิเคราะห์เอกสารด้าน Geology ในภาษาอังกฤษ และทำการจัดกลุ่มเนื้อหาเหล่านั้นเข้าไปยังหมวดต่างๆ โดยอัตโนมัติ เช่น หมวดภูมิประเทศ หมวดแร่ธาตุในดิน เป็นต้น และมีการทำหน้า UI ที่เปิดให้ผู้ที่ต้องการค้นหาข้อมูลเกี่ยวกับ Geology เข้ามาสืบค้นเอกสารในเรื่องที่ต้องการค้นคว้า และนำไปศึกษาเพิ่มเติมได้ ทางทีมจึงนำรูปแบบการทำงานของ DeepDive มาเป็นต้นแบบในการสร้างระบบ machine learning ที่สามารถวิเคราะห์เอกสารด้าน pedagogy ทั้งภาษาไทยและภาษาอังกฤษ และทำการจัดกลุ่มเนื้อหาในเอกสารเหล่านั้นโดยอัตโนมัติ และเปิดให้ผู้ที่สนใจจะหาข้อมูลด้าน pedagogy เข้ามาค้นหาเอกสารในเรื่องที่สนใจได้

Schematic

โดยเครื่องมือที่จะนำมาใช้แก้ปัญหาเหล่านี้คือเทคโนโลยี Machine Learning โดยจะใช้วิธี Supervised Classification โดยในช่วงแรกจะใช้มนุษย์ในการจำแนกเอกสารและนำเอกสารเหล่านั้นมาสอนคอมพิวเตอร์ของเรา หลังจากนั้นจะให้คอมพิวเตอร์ทำการจำแนกเอกสารว่าเอกสารนี้เกี่ยวข้องกับอะไรด้วยตนเอง ซึ่งคอมพิวเตอร์นั้นจะจำแนกเอกสารตามที่เราได้สอนไปในขั้นตอนแรกเท่านั้น โดยขั้นตอนการทำงานจะถูกแบ่งเป็น 2 ขั้นตอน คือ

1. Training phase โดยทางทีมจะได้เอกสารต่างๆ ที่ถูกคัดเลือกมาอย่างเหมาะสมจากทางมูลนิธิสดศรี-สฤษดิ์วงศ์ ซึ่งเป็นมูลนิธิที่ทำงานเกี่ยวข้องกับการพัฒนาระบบการศึกษาในประเทศ โดยทางมูลนิธิจะทำการจัดหาผู้ที่เชี่ยวชาญในด้าน pedagogy ในรูปแบบต่างๆ มาอ่านเนื้อหาภายในเอกสารแต่ละชุดที่เตรียมไว้ และทำการ tag เนื้อหาในแต่ละส่วนของเอกสาร ว่าเนื้อหาในส่วนนี้ จะมีความเกี่ยวข้องกับการเรียนการสอนในแบบใดบ้าง เช่น Problem-based learning หรือ Self-learning เป็นต้น โดย tag เหล่านี้จะมีลักษณะทั้งเป็น tag ที่ถูกกำหนดประเภทมาแล้ว (Structured tag) หรือเป็น tag ที่ผู้เชี่ยวชาญเพื่มเข้ามาด้วยตนเอง (Free form tag) โดยข้อมูลข้างต้นจะเป็นต้นแบบในการสร้าง model ของระบบ machine learning ที่จะสามารถแยกแยะเนื้อหาในเอกสารได้ โดย algorithm ที่จะใช้ในการสร้าง model ขึ้นนี้ จะประกอบไปด้วย algorithm หลายๆ อย่าง เช่น TF-IDF ที่ใช้ในการแยกแยะคำแต่ละคำในเนื้อหา หรือ Classification ที่จะแยกแยะว่าเนื้อหาส่วนนี้มีความเกี่ยวข้องกับ tag ใดบ้าง เป็นต้น แล้วนำข้อมูลทั้งหมดจัดเก็บลงใน database
2. Prediction phase เมื่อตัว model ของ machine learning ถูกสร้างขึ้นมาแล้ว ระบบก็จะสามารถทำการวิเคราะห์เนื้อหาแต่ละส่วนในเอกสารด้าน pedagogy ได้โดยอัตโนมัติ โดยเอกสารต่างๆ จะถูกส่งขึ้นมายังระบบผ่านหน้าเว็บไซต์โดยกลุ่มครูตามโรงเรียนต่างๆ ที่มีเอกสารเกี่ยวกับ pedagogy เก็บไว้เป็นไฟล์เอกสารรูปแบบ PDF หรือ Microsoft Word document และหลังจากเอกสารถูกอัพโหลดขึ้นมาในระบบแล้ว ระบบจะทำการแบ่งเนื้อหาในเอกสารดังกล่าวตามย่อหน้าและทำการส่งไปยัง model เพื่อทำการติด tag เนื้อหาที่เกี่ยวข้องในแต่ละย่อหน้าโดยอัตโนมัติ โดยในแต่ละย่อหน้าจะมี tag หนึ่งอัน หลายอัน หรือไม่มี tag เลยก็ได้ หลังจากเนื้อหาทั้งหมดถูกติด tag เรียบร้อยแล้ว ข้อมูลทั้งหมด ทั้งเนื้อหาและ tag จะถูกส่งไปเก็บใน database เพื่อใช้ในการสืบค้นต่อไป และข้อมูลที่ถูกทำการติด tag โดยอัตโนมัติและถูกนำไปเผยแพร่ในเว็บไซต์แล้ว เมื่อมีผู้ใช้งานเข้ามาแสดงความคิดเห็นหรือให้ rating เอกสารมาเป็นระยะเวลาหนึ่ง ข้อมูลความเห็นต่างๆ จะถูกนำมารวบรวม วิเคราะห์ และนำไปใช้ในการ train ตัว model ใหม่อีกครั้ง ทำให้ระบบ machine learning สามารถพัฒนาการทำงานได้ในอนาคต

เมื่อเอกสารต่างๆ ถูกจัดแบ่งกลุ่มและจัดเก็บไว้ใน database เรียบร้อยแล้ว ผู้ที่สนใจศึกษาข้อมูล ทั้งคุณครูที่ต้องการจะหาข้อมูลด้าน pedagogy เพื่อนำไปปรับใช้ในการสอนในชั้นเรียน และผู้ปกครองที่ต้องการค้นหาว่าโรงเรียนต่างๆ มีรูปแบบการจัดการเรียนการสอนในลักษณะใด จะสามารถสืบค้นข้อมูลที่ถูกจัดเก็บไว้ได้ ผ่านหน้าเว็บไซต์ที่ทางทีมได้จัดทำขึ้นมา โดยผู้สืบค้นจะทำการค้นหาเนื้อหาที่ต้องการผ่าน tag ที่ถูกกำหนดไว้ และหน้าเว็บจะแสดงผลเนื้อหาในส่วนที่ถูกติด tag ที่ผู้ค้นหาเลือกไว้ โดยผู้ค้นหาจะสามารถคลิกเข้าไปดูเนื้อหาต่างๆ ได้ และสามารถดาวน์โหลดเอกสารฉบับเต็มของแต่ละเนื้อหาได้ นอกจากนั้น หน้าเว็บไซต์จะมีส่วนที่เปิดให้ผู้เข้าชมสามารถให้ rating เอกสารและแสดงความคิดเห็นเกี่ยวกับเอกสารนั้นๆ ได้ด้วย

Technologies

ตัวระบบ machine learning จะประกอบไปด้วยส่วนประกอบ 2 ส่วนด้วยกัน ได้แก่

1. Azure SQL Database และ Azure Machine Learning ซึ่งเครื่องคอมพิวเตอร์ในส่วนนี้จะทำหน้าที่ในการประมวลผลที่เกี่ยวกับ machine learning ทั้งหมด ทั้งด้านการ train model และการ classification เอกสารใหม่ๆ ที่นำเข้ามาด้วย และยังทำหน้าที่ในการเก็บข้อมูลทางด้านเอกสารทั้งหมดไว้ในฐานข้อมูลทั้งในรูปแบบไฟล์ต้นฉบับ และในรูปแบบ SQL database ซึ่งประกอบไปด้วย เนื้อหาที่ถูกแบ่งไว้แล้ว และ tag ทั้งหมดของเนื้อหาในส่วนนั้น โดยในส่วนของ Server ทั้งสองอย่างนี้ ทางทีมจะใช้บริการ Azure ML Service ในการรันโปรแกรมด้าน machine learning ต่างๆ ซึ่งเขียนขึ้นมาในภาษา Python , Azure Storage ในการเก็บไฟล์เอกสารต้นฉบับ และ Azure SQL Database ในการจัดเก็บข้อมูลที่อยู่ในรูปแบบ SQL ทั้งหมด
2. Azure Web Service ซึ่งจะทำหน้าที่เป็น server ในการเปิดหน้าเว็บไซต์สำหรับผู้ใช้งานที่เข้ามาสืบค้นข้อมูล และครูที่ต้องการแชร์เอกสารมายังระบบ และจัดการฐานข้อมูลต่างๆ ที่เกี่ยวข้องกับการจัดการเว็บไซต์ เช่น User database เป็นต้น โดยทางทีมจะใช้บริการ Azure Virtual Machine + Public IP เพื่อทำการตั้งเซิฟเวอร์สำหรับรันเว็บไซต์ และเป็นตัวเชื่อมต่อกันระหว่าง Azure SQL Database, Azure Machine Learning กับผู้ใช้งาน

Commercialization

แผนการระยะสั้นสำหรับการดำเนินการธุรกิจ จะเป็นการร่วมมือกันกับมูลนิธิที่ดูแลเกี่ยวกับด้านการพัฒนาระบบการศึกษาในประเทศไทย เช่น มูลนิธิสดศรี-สฤษดิ์วงศ์ หรือมูลนิธิการศึกษาไทย เป็นต้น โดยจะเปิดโรงเรียนต่างๆ ในเครือมูลนิธิดังกล่าวเข้าถึงข้อมูลทั้งหมดได้ฟรี เพื่อที่จะเสริมสร้าง community ให้เกิดความแข็งแกร่ง หลังจากนั้น เมื่อมีผู้ใช้งานและเอกสารอยู่ในระบบเป็นจำนวนหนึ่งแล้ว ก็จะเริ่มขยายผลเพื่อให้ community มีประสิทธิภาพมากยิ่งขึ้น โดยการขอความร่วมมือจากกระทรวงศึกษาธิการให้ช่วยเผยแพร่ community ไปยังโรงเรียนต่างๆ ทั่วประเทศ และประชาสัมพันธ์ไปยังโรงเรียนต่างๆ และผู้ปกครองที่สนใจ โดยจะใช้ระบบสมาชิก 3 แบบ ได้แก่แบบ Standard user (ฟรี) โดยผู้ใช้กลุ่มนี้จะเน้นไปที่ผู้ปกครองเป็นหลัก ซึ่งจะมี Feature ที่จะสามารถสืบค้นเอกสาร ดาวน์โหลดข้อมูล และให้ rating กับแสดงความคิดเห็นกับเอกสารต่างๆ ได้ทั้งหมด แบบ Teacher (ฟรี) โดยผู้ใช้กลุ่มนี้จะเพิ่มความสามารถในการ upload เอกสารด้าน pedagogy เข้าสู่ระบบได้ โดยคุณครูที่จะทำการอัพโหลดเอกสารจะต้องถูกคัดเลือกจากผู้ดูแลเพื่อแสดงว่า User นี้มีคุณสมบัติที่เหมาะสมที่จะสามารถนำเอกสารมาเข้าในระบบได้ และแบบ Subscription ซึ่งจะมีค่าใช้จ่ายรายเดือนเพิ่มเติม โดยผู้ใช้ในกลุ่ม Subscription นี้จะมีบริการเสริมเพิ่มเติมเข้ามา เช่น การ favorite หัวข้อที่สนใจเป็นพิเศษ แล้วจะมีการส่ง E-mail แจ้งเตือนเมื่อมีเอกสารใหม่ที่เกี่ยวข้องกับเรื่องที่ favorite ไว้เพิ่มเข้ามาในระบบ