



บันทึกข้อความ

กองนโยบายและแผนการใช้ที่ดิน
เลขที่รับ.....๒๔๕๐
วันที่.....๒๔ พ.ค. ๖๗
เวลา.....๑๑.๐๗

ส่วนราชการ กองนโยบายและแผนการใช้ที่ดิน กลุ่มวิเคราะห์สภาพการใช้ที่ดิน โทร. ๒๒๐๓

ที่ กษ ๐๘๓๗.๐๒/ ๕๗๗ วันที่ ๒๔ พฤษภาคม ๒๕๖๗

เรื่อง ขอส่งสรุปการอบรมครั้งที่ ๒/๒๕๖๗

เรียน ผู้อำนวยการกลุ่มวิเคราะห์สภาพการใช้ที่ดิน

ในปีงบประมาณ ๒๕๖๗ ข้าพเจ้าได้เข้าร่วมอบรมหลักสูตร เรื่อง “การสร้างความรู้ความเข้าใจ ภาษาศาสตร์คอมพิวเตอร์เพื่อนำไปประยุกต์ใช้กับปัญญาประดิษฐ์” เมื่อวันที่ ๑๘ พฤษภาคม ๒๕๖๗ และ เข้าร่วมตอบแบบสอบถามในงานวันสถาปนากรมพัฒนาที่ดิน “๖๑ ปี พัฒนาที่ดิน เกษตรกรทำกินอย่างยั่งยืน” เมื่อวันที่ ๒๓ พฤษภาคม ๒๕๖๗ เสร็จเรียบร้อยแล้ว จึงขอส่งสรุปรายงานเรื่อง “การสร้างความรู้ความเข้าใจ ภาษาศาสตร์คอมพิวเตอร์เพื่อนำไปประยุกต์ใช้กับปัญญาประดิษฐ์” รายละเอียดตามเอกสารที่แนบมา พร้อมนี้

จึงเรียนมาเพื่อโปรดทราบ และนำเรียนผู้อำนวยการกองนโยบายและแผนการใช้ที่ดินต่อไป

นางสาวสมจิตต์ เลิศดิษยวรรณ

(นางสาวสมจิตต์ เลิศดิษยวรรณ)
นักวิชาการเกษตรชำนาญการพิเศษ

เรียน ผอ.กนผ.

เพื่อโปรดพิจารณา

นางสาวอมรรัตน์ สระเพ็ชร

(นางสาวอมรรัตน์ สระเพ็ชร)
นักวิชาการเกษตรชำนาญการพิเศษ
ผู้อำนวยการกลุ่มวิเคราะห์สภาพการใช้ที่ดิน

ลงนามแล้ว

- ศก. / วกก. รวบรวม

(นายชาคริต อินนระ)

ผู้อำนวยการกองนโยบายและแผนการใช้ที่ดิน
๒๕ พ.ค. ๒๕๖๗

รายงานสรุปการอบรม/สัมมนา/พัฒนาความรู้/ประชุมเชิงปฏิบัติการ/และเป็นวิทยากร
กองนโยบายและแผนการใช้ที่ดิน กรมพัฒนาที่ดิน

ส่วนที่ ๑ ข้อมูลทั่วไป

ชื่อ นางสาวสมจิตต์ นามสกุล เลิศศิษยวรรณ
ตำแหน่ง นักวิชาการเกษตรชำนาญการพิเศษ กลุ่ม/ฝ่าย กลุ่มวิเคราะห์สภาพการใช้ที่ดิน
หลักสูตร/หัวข้อเรื่องอบรม/สัมมนา/พัฒนาความรู้
การสร้างความรู้ความเข้าใจภาษาศาสตร์คอมพิวเตอร์เพื่อนำไปประยุกต์ใช้กับปัญญาประดิษฐ์
สถานที่อบรม/สัมมนา/พัฒนาความรู้
ระบบการฝึกอบรมผ่านสื่ออิเล็กทรอนิกส์ (TDGA e-Training)
หน่วยงานที่จัดฝึกอบรม/สัมมนา/พัฒนาความรู้
สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล
ตั้งแต่วันที่ ๑๘ เดือน พฤษภาคม พ.ศ. ๒๕๖๗ ถึงวันที่ ๑๘ เดือน พฤษภาคม พ.ศ. ๒๕๖๗
เพื่อ ☒ อบรม ☐ สัมมนา ☐ อื่นๆ ระบุ _____

ส่วนที่ ๒ สิ่งที่ได้รับจากการอบรม/สัมมนา/พัฒนาความรู้

๒.๑ รายงานสรุปเนื้อหาสาระสำคัญในการอบรม/สัมมนา/พัฒนาความรู้

ภาษาศาสตร์ คอมพิวเตอร์ (Computational Linguistics) หรือ การประมวล
ภาษาธรรมชาติ(Natural Language Processing) เป็นการรวมเทคนิคทางด้าน Computer Science
โมเดลทางคณิตศาสตร์ และคลังข้อมูลภาษา เพื่อใช้ในการทำความเข้าใจธรรมชาติของภาษา และใช้ในการทำ
แอปพลิเคชันต่าง ๆ แต่แตกต่างกันที่จุดมุ่งหมาย โดยจุดมุ่งหมายของภาษาศาสตร์คอมพิวเตอร์
(Computational Linguistics) เพื่อเข้าใจธรรมชาติของภาษาผ่านคลังข้อมูลและโมเดล ในขณะที่จุดมุ่งหมาย
ของการประมวลภาษาธรรมชาติ (Natural Language Processing) ทำให้คอมพิวเตอร์ทำหน้าที่ทางภาษา
ได้ดีขึ้น

๑ การศึกษาคุณสมบัติทางสถิติของภาษา คือ การศึกษาว่าหน่วยภาษาต่าง ๆ เกิดขึ้นมากน้อยแค่ไหน
ซึ่งเรียกอีกอย่างว่า การกระจายตัวของหน่วยทางภาษา (Distribution) ความรู้เหล่านี้ทำให้เราเข้าใจ
ธรรมชาติของภาษามากขึ้น อีกทั้งนำไปประยุกต์ได้ในการประมวลภาษาธรรมชาติได้ ซึ่งสามารถใช้ได้กับ
ทุกภาษา

๒ กฎของชิพฟ์ Zipf และโค้ง Zipf

ขั้นตอนแรกในการศึกษาภาษาศาสตร์คอมพิวเตอร์ คือ การเลือก representative text หรือ
คลังข้อมูล(Corpus) ที่เหมาะสมกับจุดประสงค์ของการศึกษา ถ้าต้องการศึกษาภาษาที่คนใช้จริงในปัจจุบัน
ทั่ว ๆ ไป ตัวอย่าง representative text ที่เหมาะสมได้แก่ หนังสือพิมพ์ นิตยสาร หรือเว็บไซต์ทั่ว ๆ ไป
และตัวอย่าง representative text ที่ไม่เหมาะสมได้แก่ จดหมายราชการ ประกาศจากราชการ เป็นต้น

กราฟของ Zipf จะพล็อตจากแกน X เป็นความถี่ และแกน Y เป็นอันดับของความถี่ของหน่วยทางภาษาหรือคำที่ทำการตัดไว้ก่อนแล้ว และโค้งของ Zipf จะเป็นเส้นโค้งลง โดยมีความชันของเส้นในช่วงแรกสูงมากหรือมีการตกลงอย่างรวดเร็ว และด้านปลายมีความชันที่ค่อนข้างต่ำ และเมื่อใช้ข้อมูลที่ได้จากกฎของ Zipf สามารถนำมาสร้างเป็นกราฟเชิงเส้น แบบ Logarithm ที่มีความชันเท่ากับ -๑

กฎของ Zipf เป็นจริงสำหรับภาษาทุกภาษา ข้อความทุกขนาด ประเภทของข้อความทุกประเภท หัวข้อเรื่องทุกแบบ และความซับซ้อนยากง่ายของเนื้อหาทุกรูปแบบ

๓ การประยุกต์ใช้กฎของ Zipf เพื่อการพัฒนา Search Engine อย่างง่าย

เริ่มจากการสร้าง inverted index ก่อน (จากคำไปถึงเอกสาร) แล้วนำกฎของ Zipf มาพล็อตข้อมูลนั้น โดยตัดข้อมูลในช่วงเริ่มต้นออก เพราะเป็นข้อมูลที่ใหญ่ ทำให้ประมวลผลช้า และไม่มีประโยชน์ และเก็บส่วนที่เหลือไว้ เพราะเป็นคำที่มีประโยชน์ ทำให้ค้นหาและประมวลผลได้เร็วขึ้น ทำให้ขนาด inverted index เล็กลงและเก็บได้ง่ายขึ้น

๔ การประยุกต์ใช้ Language Model

๔.๑ เครื่องแปลงเสียงเป็นอักษร (Speech Recognizer) ประกอบด้วย Acoustic model Pronunciation model และ Language Model

๔.๒ Language Model เป็นตัวที่ระบุได้ว่ามันเป็นภาษาหรือเปล่า มีความเป็นภาษาขนาดไหน ซึ่งจะมีหลายแบบ เช่น Unigram model (วิเคราะห์ทีละหนึ่งคำ) Bigram model (วิเคราะห์ทีละสองคำที่ติดกัน) Trigram model (วิเคราะห์ทีละสามคำที่ติดกัน) และ Fourgram (วิเคราะห์ทีละสี่คำที่ติดกัน) และต่อไปเรื่อย ๆ แล้วแต่กรณี

โดยวิเคราะห์จากข้อมูลในคลังข้อมูลที่เลือกมาเป็นตัวแทนที่ดี จากนั้นเราก็จะตัดคำที่ไม่มีประโยชน์ออกไป พิจารณาเฉพาะสถิติของคำที่ต้องการเปรียบเทียบกับ ว่าคำใดหรือวลีใดที่มีการใช้งานมากน้อยเพียงใด คำใดหรือวลีใดที่มีการพบมากที่สุดหรือมีสถิติสูงสุดเป็นคำที่ถูกต้อง

๔.๓ Language Model สามารถประยุกต์ใช้ในการตรวจไวยากรณ์ การตรวจการสะกดคำผิด หรือคำที่ผิดบริบทของประโยค

Language Model เป็นสิ่งที่นักภาษาศาสตร์คอมพิวเตอร์คิดขึ้นมา ไม่ต้องใช้กฎหรือไวยากรณ์ แค่ใช้คลังข้อมูลที่เหมาะสมกับหน้าที่ทางภาษาที่เราต้องการจะทำ

ในภาษาศาสตร์คอมพิวเตอร์(Computational Linguistics) การใช้ Language Model ทำให้เราเข้าใจภาษามากขึ้น สามารถแยกแยะสิ่งที่กำกวมได้ดีขึ้น

ส่วนการประมวลภาษาธรรมชาติ (Natural Language Processing) ถ้าเรามี context หรือบริบทที่มากขึ้น ก็ทำให้สามารถตัดสินใจได้ดีขึ้นว่าประโยคไหนถูกต้อง ประโยคไหนผิด แต่การที่จะเทรนโมเดลที่มี context ที่กว้างขึ้นก็ต้องใช้คลังข้อมูลขนาดใหญ่มากขึ้นเช่นกัน

๒.๒ ประสพการณ์/ประโยชน์ที่ได้รับการประยุกต์ใช้กับหน่วยงาน

☒ ต่ตนเอง

ทำให้ได้รับความรู้เกี่ยวกับภาษาศาสตร์คอมพิวเตอร์ ขั้นตอนกระบวนการในการคิด การวิเคราะห์ และการประยุกต์ใช้ Language Model

☒ ต่อหน่วยงาน/การนำมาประยุกต์ใช้กับหน่วยงาน

สามารถนำความรู้ความเข้าใจภาษาศาสตร์คอมพิวเตอร์เพื่อนำไปประยุกต์ใช้กับ
ปัญญาประดิษฐ์ การสร้างแอปพลิเคชันต่าง ๆ ที่สามารถนำไปใช้ปฏิบัติงานที่มีประสิทธิภาพมากยิ่งขึ้น

๒.๓ ปัญหาและอุปสรรคในการอบรม/สัมมนา/พัฒนาความรู้

-

๒.๔ ข้อคิดเห็นและข้อเสนอแนะ

-

ลงชื่อ

นางสาวสมจิตต์ เลิศดิษยวรรณ

(นางสาวสมจิตต์ เลิศดิษยวรรณ)

ตำแหน่ง

นักวิชาการเกษตรชำนาญการพิเศษ

ผู้รายงาน

วันที่

๒๐ เดือน พฤษภาคม พ.ศ. ๒๕๖๗

ส่วนที่ ๓ ความเห็นของผู้บังคับบัญชา

() ทราบ

ลงชื่อ

(นายชาคริต อินณะระ)

ตำแหน่ง ผู้อำนวยการกองนโยบายและแผนการใช้ที่ดิน

วันที่

๒๔ เดือน พ.ค. พ.ศ. ๖๗

ประกาศนียบัตร

ให้ไว้เพื่อแสดงว่า

สมจิตต์ เลิศดิษยวรรณ

ได้ผ่านการอบรมด้วยระบบการเรียนออนไลน์ในบทเรียน
การสร้างความรู้ความเข้าใจภาษาศาสตร์คอมพิวเตอร์
เพื่อการนำไปประยุกต์ใช้กับปัญญาประดิษฐ์

รวมระยะเวลาทั้งสิ้น 1 : 30 ชั่วโมง

โดยสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล
ภายใต้การดำเนินงานของสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)
ให้ไว้ ณ วันที่ 20 พ.ค. 2567



(นายชินทร์ ชีรฐิตยางกูร)

ผู้อำนวยการสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล



0ac12eeb

Signed by สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) (สพร.)
Digital Government Development Agency (Public
Organization) (DGA)
Date: 2024-05-20T18:02:00.757+07:00