Report

算法解釋

• **Q1 – Logic Warm-up (2%):**

```
Question q1
==========

*** PASS: test_cases/q1/correctSentence1.test
***     PASS
*** PASS: test_cases/q1/correctSentence2.test
***     PASS
*** PASS: test_cases/q1/correctSentence3.test
***     PASS
*** PASS: test_cases/q1/entails.test
***     PASS
*** PASS: test_cases/q1/entailsLong.test
***     PASS
*** PASS: test_cases/q1/findModelSentence1.test
***     PASS
*** PASS: test_cases/q1/findModelSentence2.test
***     PASS
*** PASS: test_cases/q1/findModelSentence3.test
***     PASS
*** PASS: test_cases/q1/findModelUnderstandingCheck.t
est
***     PASS
*** PASS: test_cases/q1/plTrueInverse.test
***     PASS

### Question q1: 10/10 ###
```

Sentence1 和 sentence2 直接按照題目指示的 pseudocode 指示寫

Sentence3: PacmanAlive_1 PacmanAlive_0, PacmanBorn_0, and PacmanKilled_0, 名字代表了狀態，後面接的代表時間，同樣依照 pseudo code 指示完成。

findModelUnderstandingCheck()：

pycoSAT(expr) 將會返回一個字典，字典包含使得表達式為真的變數及其對應的 Bool。這個字典的鍵是 Expr 類型的變數，值是 Bool。所以創建一個 class 回傳 Expr 類型的變數名字

entails：

前提與結論的否定一起構成一個不可滿足的表達式，那麼前提就蘊含了結論

plTrueInverse：

pl_true(exp, model={}):Return True if the propositional logic expression is true in the model, and False if it is false.

• **Q2 – Logic Workout (2%):**

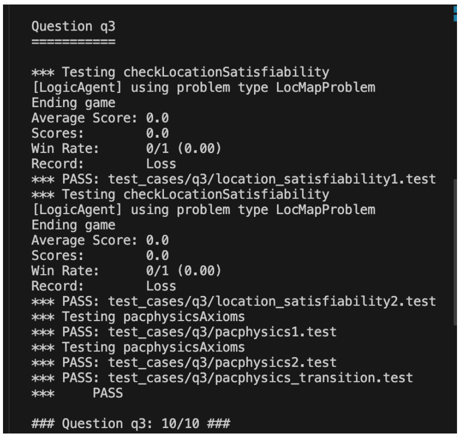
```
Question q2
===========

*** PASS: test_cases/q2/atLeastOne.test
***     PASS
*** PASS: test_cases/q2/atLeastOneCNF.test
***     PASS
*** PASS: test_cases/q2/atLeastOneEff.test
***     PASS
*** PASS: test_cases/q2/atMostOne.test
***     PASS
*** PASS: test_cases/q2/atMostOneCNF.test
***     PASS
*** PASS: test_cases/q2/atMostOneEff.test
***     PASS
*** PASS: test_cases/q2/exactlyOne.test
***     PASS
*** PASS: test_cases/q2/exactlyOneCNF.test
***     PASS
*** PASS: test_cases/q2/exactlyOneEff.test
***     PASS

### Question q2: 10/10 ###
```

atLeastOne：如果 A ｜ B ｜ C 為 true，則至少有一個是真的

atMostOne：如果有大於 1 個數為真，則(~A|~B)，(~A|~C)，(~B|~C)至少一個 pair 為假，則(~A|~B)＆ (~A|~C)＆(~B|~C)不可能為真

exactlyOne：同時符合 atLeastOne 和 atMostOne 則可以成立 exactlyOne。

- **Q3 – Pacphysics and Satisfiability (2%):**



pacmanSuccessorAxiomSingle：

如果 current 條件成立 successor 才會成立，反之亦然。

pacphysicsAxioms：

對於 all_coords 中的所有 (x, y)：

If a wall is at (x, y) --> Pacman is not at (x, y)(implication)

確保了在時間 t ,Pacman 只能出現在非外牆的一個位置還有只有一個行動。

調用 sensorModel(t, non_outer_wall_coords) 的結果，除非為 None。

調用 successorAxioms(t, walls_grid, non_outer_wall_coords) 的結果，描述了 Pacman 如何在這個時間步中以不同的位置結束。

最後，返回將所有句子連接起來的結果。

checkLocationSatisfiability：

將 Pacman 在時間 t=0 和 t=1 的物理規則添加到知識庫中。

將 Pacman 在時間 t=0 的位置和執行的動作添加到知識庫中。

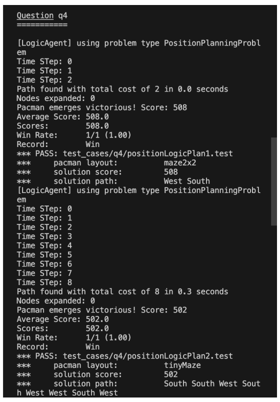
添加 Pacman 在時間 t=1 執行的動作到知識庫中。

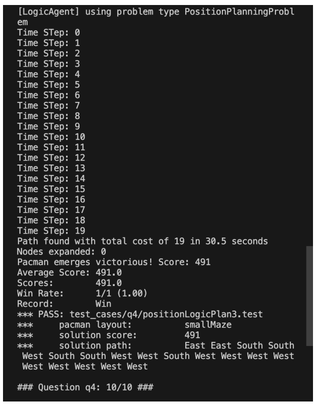
回傳兩個模型：

一個模型，其中 Pacman 在指定位置。

另一個模型，其中 Pacman 不在指定位置。

- **Q4 – Path Planning with Logic (2%):**





將起始位置添加到 KB 中。

for t in range(50)

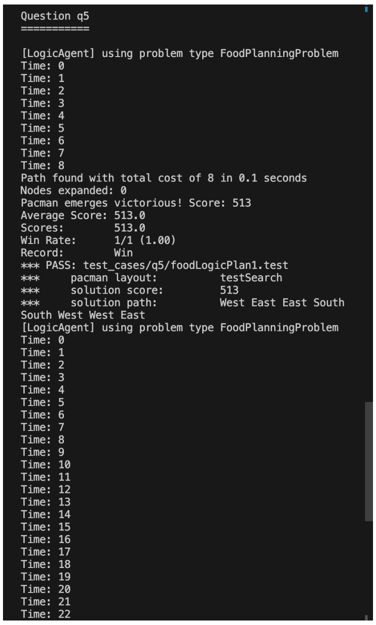
確保了在每個時間步 Pacman 只能出現在一個位置。

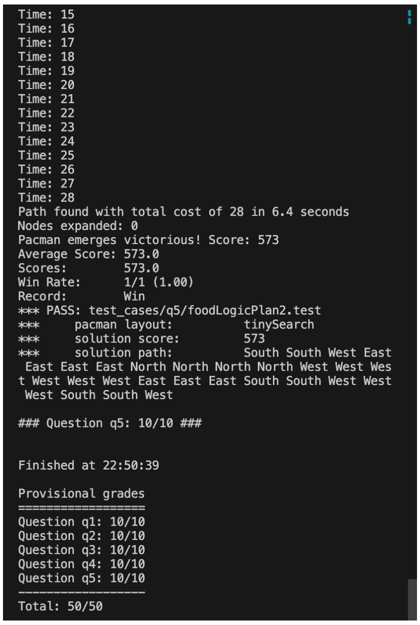
檢查 Pacman 是否到達了目標位置。如果找到了符合條件的模型，用 extractActionSequence 函數從模型中提取動作序列並返回。

確保在每個時間步 Pacman 只能採取一個動作

使用 pacmanSuccessorAxiomSingle 函數添加過渡模型句子。這個句子確保了 Pacman 在時間步 t+1 時的位置取決於時間步 t 的位置和動作

- **Q5 – Eating All the Food (2%):**





將 Pacman 的初始位置加入 KB 中。

將所有食物的初始位置加入 KB 中。

for t in range(50):

創建了一個 goal，目標是確保所有食物在時間 t 都被吃掉了。

確保在時間 t，Pacman 只能在非牆壁的位置之一。

檢查 Pacman 是否到達了目標。如果找到了符合條件的模型，用 extractActionSequence 函數從模型中提取動作序列並返回。

確保在時間 t，Pacman 只能採取其中一個動作。

確保 Pacman 在時間 t+1 的位置是合法的。

food successor axiom：如果在時間 t ，Pacman 在 (x, y) 位置，或是如果時間 t，(x,y)沒有食物，則在時間 t+1，(x, y) 位置沒有食物。

Q1(8%). According to AI Weekly in the lecture, some experts and scholars such as Karl Friston and Yann LeCun believe: "You can't get to AGI with LLMs ." Nowadays, the prospects of LLM are so optimistic. Why do you think these experts have such ideas? Please elaborate on your views.

Importance of World Models:

one of the experts in the article stated that the reason the future of AI is not generative is because you can not establish any truth or understanding of the way the world works through computer vision or LLMs. They argue that LLMs lack the ability to develop comprehensive world models necessary for true intelligence.

Limitations of Deep Learning:

While a powerful engineering approach to AI, is not sufficient to achieve AGI. The experts argued that Deep Learning systems are very limiting, and we need systems that can achieve more complex learning and decision making.

Energy Efficiency:

The experts compared LLMs to human learning. He pointed out even a four-month-old infant absorbs more information than the largest LLM in just a few months. They argue that current LLM-based systems require unsustainable amounts of energy and data, making them impractical for achieving AGI.

Q2(8%). According to the paper "CLIP-Event: Connecting Text and Images with Event Structures," in CVPR 2022, after the process of generating the event-centric structured data, how does this work implement contrastive learning? Specifically, how does this work choose the positive and negative samples for contrastive learning?

Positive Samples:

Positive samples consist of the original images paired with their corresponding event-centric structured text descriptions.

Negative Samples:

Negative samples are selected based on the confusion matrix for the event type classifier of the CLIP model, which do not match the primary event. The negative event types are the challenging cases in image event typing, the event types whose visual features are ambiguous with the primary event type. For example, if the primary event is "TRANSPORT," similar event types like "ARREST" will be chosen as negative samples.

For argument roles, the model manipulates the order of arguments by performing a right-rotation of the argument role sequence, reassigning each argument to a different role. If there is only one argument for the event, a negative role is sampled according to the argument confusion matrix of the text argument extraction system.

Q3(8%). Referring to the paper, what is the main problem with the current description of the VLM pre-training process? Please describe the steps to generate the structured graph-based data in this work.

The main problem with the current description of the VLM (Vision-Language Model) pre-training process is that it relies solely on unstructured descriptions, which consist of key entities and attributes that define a category. However, conventional descriptions fall short of structured information that effectively represents the interconnections among entities or attributes linked to a particular category

Steps to generate the structured graph-based data:

Linguistic Data Generation:

Use ChatGPT to generate descriptions with corresponding structured relationships.

adopt Nh question templates as the language instruction T for LLMs

Generate descriptions (D) using the language instruction (T) for each category.

Design an extra instruction (T) to leverage LLMs for producing structured knowledge, including entities, attributes, and relationships among them.

The structured knowledge generated from the descriptions (D) is denoted as R ,includes entity set, attribute set, entity-entity relationships, and entity-attribute relationships based on description.

Hierarchical Prompt Tuning (HPT):

HPT leverages both the descriptions (D) and structured knowledge (R) for learning prompts in a hierarchical manner.

1. Low-Level Prompts
2. High-Level Prompts
3. Global-Level Prompts

Hierarchical Tuning process:

It involves updating the prompts based on the structured knowledge and descriptions. The final text representation is acquired by projecting the text embeddings to a common latent embedding space.

Relationship-guided Attention Module:

This module is introduced to model structured knowledge (R) and capture pair-wise correspondences among entities and attributes.

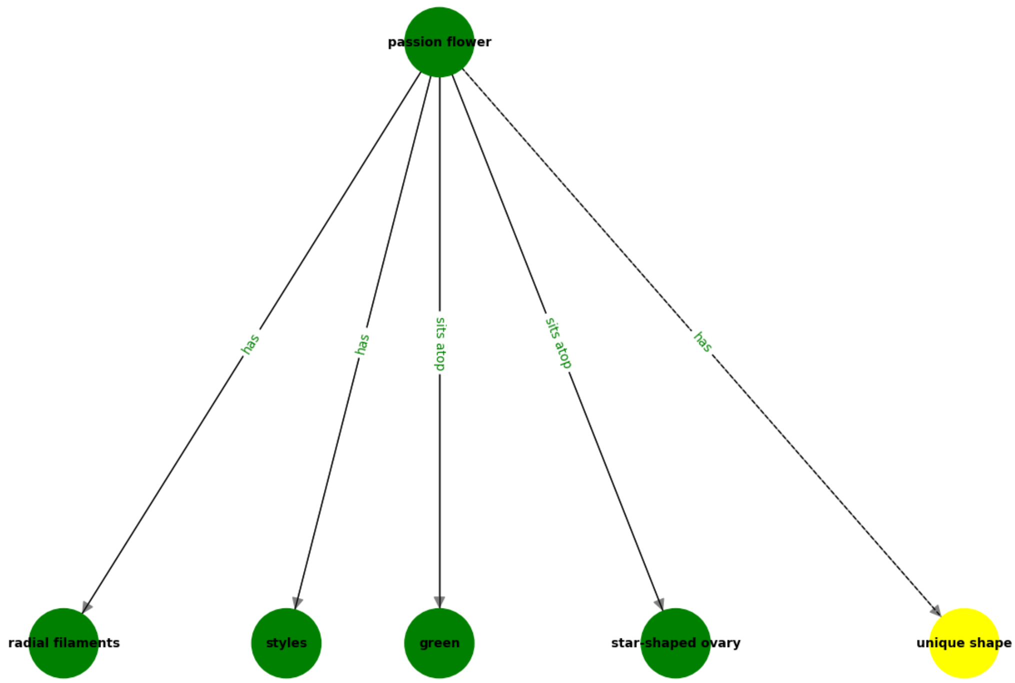
Q4(8%). Please select one of the datasets provided in this work and visualize two categories.

- **Path to the gpt-generated data: ./data/gpt_data**
- **You must show the corresponding description and two graph-based structured data components for each category as shown on the right.**

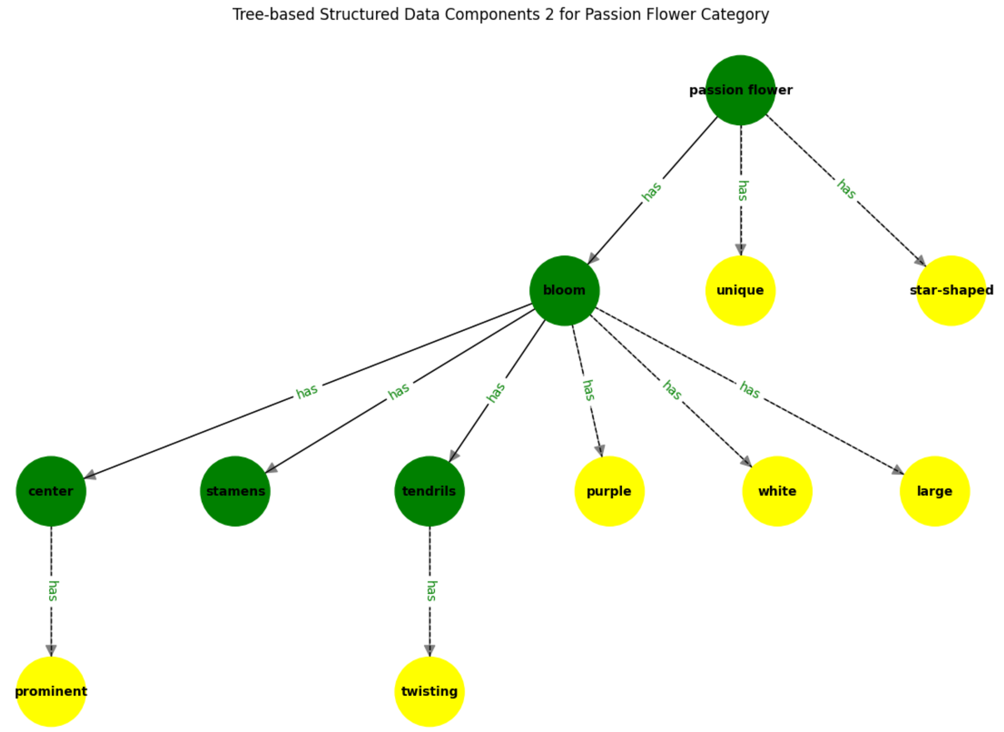
Category: passion flower

"The passion flower has a unique shape with radial filaments and three styles that sit atop a green, star-shaped ovary."

Tree-based Structured Data Components 1 for Passion Flower Category

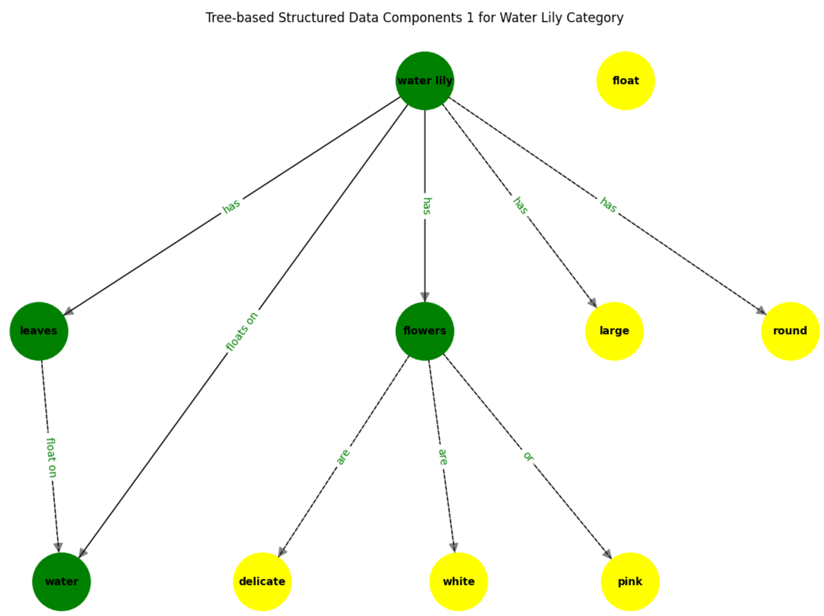


"The passion flower has a unique star-shaped purple and white bloom with a large center, prominent stamens, and twisting tendrils."



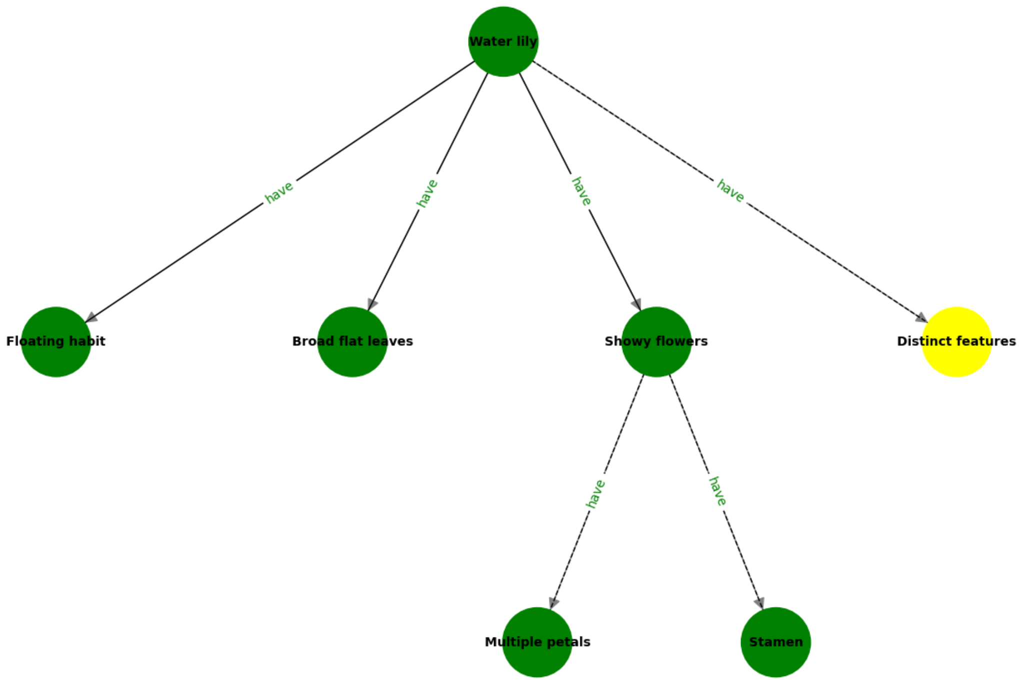
Category: water lily

"The water lily has large, round leaves that float on the water and beautiful, delicate, white or pink flowers."



"Some distinct features of a water lily are its floating habit, broad flat leaves, showy flowers with multiple petals and stamen."

Tree-based Structured Data Components 2 for Water Lily Category



Q5(8%). Based on current VLM auxiliary data improvement methods, such as the event-centric structure data in the lecture and the structured linguistic knowledge in this paper, what other deep semantic knowledge do you think humans possess that can be provided to VLM for learning

人類擁有情感，可以從語言理解、表達和回應情感。為 VLM 提供有關情感、共情和社交提示的知識可以幫助它們生成更具情感智慧和共情能力的回應。