

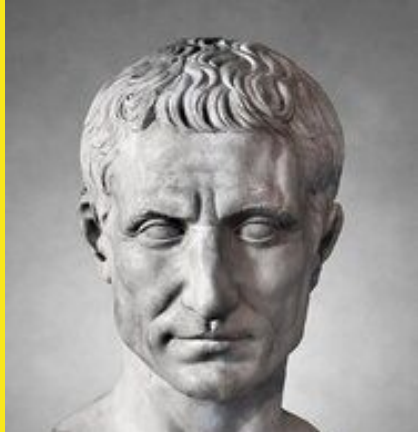
Linguistics-Reasoned NLP

語言、知識與運算

language

knowledge

computation



"Divide and Conquer"

- Julius Caesar

語言

→ 對結構的敏感度

父母把屎把尿地
把孩子餵飽。
父母真辛苦。

to feed in the ways of:
把屎把尿

父母一把屎一把尿地
把孩子餵飽。
孩子真辛苦。

to feed with:
一把屎、一把尿

知識

→ 對語境的敏感度

- 在 Pub 裡...

A:「先來一瓶紅酒」

B:「那我要一杯歐洽塔就好」

Q:「歐洽塔」最可能是什麼？

a. 食物

b. 飲料

c. 潤滑油

知識 → 對語境的敏感度

- 在機車行裡...

A:「我要全合成油」

B:「那我要歐洽塔」

Q:「歐洽塔」最可能是什麼？

a. 食物

b. 飲料

c. 潤滑油

不知道的時候，
你會「問」！

不問的時候，你會靠
語言 和 **知識** 去「猜」！

知識圖譜

Artificial Intelligence

計劃能力

邏輯歸因能力

搜尋

知識圖譜

ML/DL

基因演算法

LSTM

貝氏機率網路

統計模型

SVM

Transformer
(BERT/GPT)

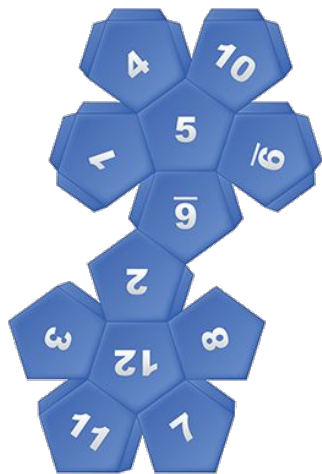
什麼是知識圖譜？

- 知識圖譜(Knowledge Graph, KG)
 - 本質上是一種語意網路。是「用以表達意義」的一種網路結構。
 - 網路結構可視為一個多面體，每一面是一個「節點 (Point)」，而面和面之間則以「邊 (Edge)」相連結
 - 在知識圖譜里，每個節點表示現實世界中存在的「實體 (Entity)」，每條邊為實體與實體之間的「關係 (Relation)」
-

Point/Edge a.k.a. Entity/Relation

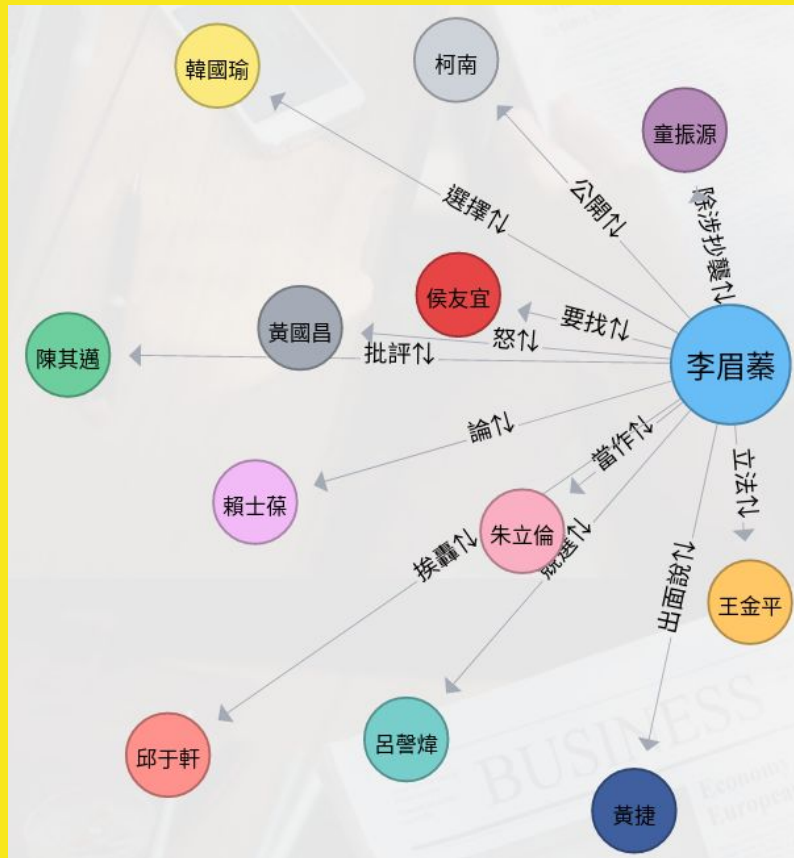


多面體



展開的多面體

圖譜化多面體



Rule-Based Systems

```
% Background knowledge:
isFather("John", "Dan") % John is Dan's father
isFather("Paul", "John") % Paul is John's father
isWife("Alice", "Paul") % Alice is Paul's wife

isGrandmother(X,Y):-isWife(X,Z),
                    isGrandfather(Z,Y) % X is Y's grandmother
                                         % if X is Z's wife and
                                         % Z is Y's grandfather

isGrandfather(X,Y):-isFather(X,Z),
                    isFather(Z,Y) % X is Y's grandfather
                                   % if X is Z's father and
                                   % Z is Y's father

% Logic programming query:
?-isGrandmother("Alice","Dan") % Is Alice Dan's grandmother?
```

以前

https://www.researchgate.net/figure/Logic-program-example_fig1_362908000

Enterprise
AI to t

以前


opencyc
Public
Watch 14

master
1 Branch
0 Tags

Add file
Code

	asanchez75 Update README.md	3f04417 · 4 years ago	6 Commits
	KB+inference_engine	added files	6 years ago
	.DS_Store	added files	6 years ago
	.gitattributes	added files	6 years ago
	README.md	Update README.md	4 years ago
	opencyc-2008-06-10.owl.gz	added files	6 years ago
	opencyc-2009-04-07-readable.owl.gz	added files	6 years ago
	opencyc-2009-04-07.owl.gz	added files	6 years ago
	opencyc-2010-08-15-readable.owl.gz	added files	6 years ago
	opencyc-2010-08-15.owl.gz	added files	6 years ago
	opencyc-2012-05-10-readable.owl.gz	added files	6 years ago
	opencyc-2012-05-10.owl.gz	added files	6 years ago
	opencyc-latest.owl.gz	added files	6 years ago

Cyc中的概念被稱為「常量 (constants)」。常量以"\$"開頭並區分大小寫。常量主要分為以下幾類：

- 個體，即*individuals*：例如#\$BillClinton又如#\$France。
- 集合，即*Collections*：例如#\$Tree-ThePlant（包含所有的樹），又如#\$EquivalenceRelation（包含所有的等價關係）。集合中的個體被稱為該集合的實例（*instance*）。
- 真值函數，即*Truth Functions*：該函數可被應用於一個或多個概念，並返回「真」或「假」。例如#\$siblings表示兄弟姐妹關係，若兩個參數對應的內容為兄弟姐妹關係，則該概念返回真值。約定真值函數以小寫字母開頭，並且可以被拆分為若干個邏輯連接詞（例如#\$and、#\$or、#\$not、#\$implies）、量詞（#\$forAll, #\$thereExists等等）以及謂詞。
- 函數，即*Functions*：用於以現有術語為基礎產生新的術語。例如#\$FruitFn具有以下作用：若接收到用於描述一種（或一個集合）植物的聲明，則會返回其果實。約定函數常量以大寫字母開頭，並以「Fn」作為結尾。

Cyc中的謂詞最重要的是#\$isa以及#\$genls。#\$isa表示某個對象是某個集合的個體（instance），#\$genls表示某個集合是另外一個集合的子集合。由概念構成的事實採用CycL語言描述的「句子」表示。謂詞則寫在與其相關的對象之前，並以括號括起來：（#\$isa #\$BillClinton #\$UnitedStatesPresident）表示「Bill Clinton屬於美國總統集合」；（#\$genls #\$Tree-ThePlant #\$Plant）表示「所有的樹都是植物」；（#\$capitalCity #\$France #\$Paris）表示「巴黎是法國的首都」。

句子中可以包含變量，變量字符串以"?"開頭，這些句子被稱為「規則」。與#\$isa謂詞有關的一條規則如下所示：

```
(#$implies
  ($and
    ($isa ?OBJ ?SUBSET)
    ($genls ?SUBSET ?SUPERSET))
  ($isa ?OBJ ?SUPERSET))
```

以前

以前

https://www.youtube.com/watch?v=gD2CNI_xLNQ

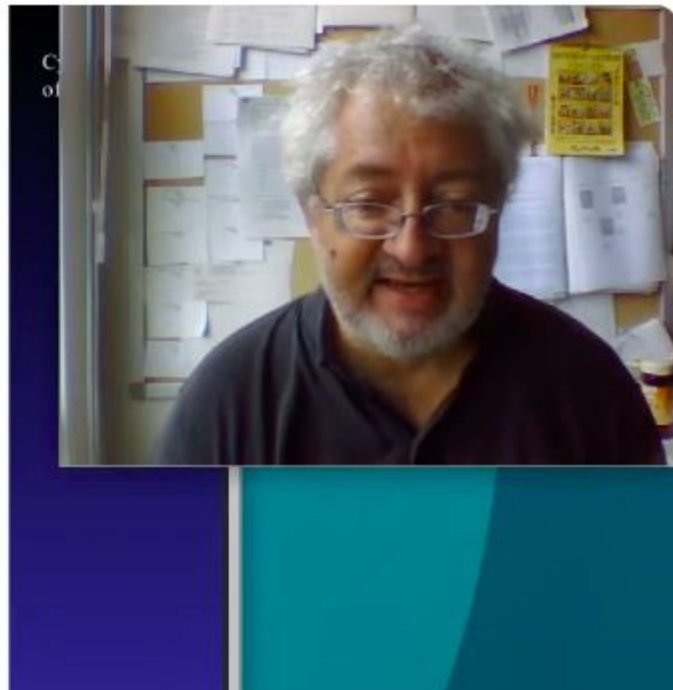
CycL query language

不，你不是這樣輸入！

“A frightened person” in CycL:

```
(#$and  
  ($isa ?x #Person)  
  ($feelsEmotion ?x #Fear #High))
```

你得用它自己的語法輸入！



這些 CS-AI 專家大概從來沒有想過「自然語言」有自己的句法解析器的存在！
相較之下，語意的「**相依樹**」是一般人比較好理解與想像的。

現況...

廣義知網知識本體架構2.0版線上瀏覽 (本介面呈現約二分之一的常用詞彙供參考查詢)

大廣義知網知識本體架構線上瀏覽系統

TopNode

- entity|事物 [事物]
 - event|事件
 - object|物體 [對象, 客體]
 - thing|萬物 [東西, 萬物, 萬有, 東東]
 - time|時間
 - 時間 [時間]
 - 日子 [日子]
 - 歲月 [歲月]
 - 年華 [年華]
 - space|空間 [空間, 空隙, 間隙]
 - relation|關聯 [關係]
 - OtherWord(object|物體)
 - Other(entity|事物)
 - instance(事物) [國家重點美術館(事物), 中華人民共和國國家重點京劇院團(事物), 國家國防教育示範基地(事物), 中範基地(事物), 中華人民共和國國家重點實驗室列表(事物) 準(事物), 黨和國家領導人(事物), 中華人民共和國國家重慶歷史文化街區(事物), 中華人民共和國國家實驗室列表(事
 - relation
 - FeatureConcept

TopNode

- entity|事物 [事物]
 - event|事件
 - object|物體 [對象, 客體]
 - thing|萬物 [東西, 萬物, 萬有, 東東]
 - physical|物質
 - animate|生物
 - AnimalHuman|動物 [動物]
 - plant|植物 [植物, 花木, 植株, 草木]
 - bacteria|微生物 [微生物]
 - BodySubstance|身體部件 [切片]
 - inanimate|無生物
 - 剩餘物|remnant [殘骸, 殘留物]
 - 導體|conductor [半導體]
 - 媒介|intermedium [媒介, 介質]
 - 禮物|gift [禮, 禮物, 禮品, 儀, 贈品, 恩賜]
 - Miscellaneous(physical|物質)
 - abstract|抽象物
 - part|部分 [部分, 份, 部份, 角, 成分, 成份, 片段, 部件, 片斷]
 - group|群體
 - 基礎|foundation [基本, 基礎, 根本, 基, 宗, 主體, 根基, 基石, 母體, 本體, 基底]
 - 障礙|obstacle [障礙, 障礙物]
 - Miscellaneous(thing|萬物)
 - time|時間
 - space|空間 [空間, 空隙, 間隙]
 - relation|關聯 [關係]
 - Miscellaneous(object|物體)
 - relation
 - FeatureConcept

現況...

默认搜索 ▾

猫



猫 实例

猫

猫，属于[猫科动物](#)，分[家猫](#)、[野猫](#)，是全世界家庭中较为广泛的[宠物](#)。

实例

[宠物](#)

[古埃及](#)

[猫科动物](#)

[野猫](#)

[波斯猫](#)

猫 实例

世界四大音乐剧之一

《猫》是安德鲁·劳埃德·韦伯根据T·S·艾略特为[儿童](#)写的诗改编的[音乐剧](#)。

实例

[音乐剧](#)

[英国](#)

[百老汇](#)

[伦敦](#)

[儿童](#)

猫 实例

郑振铎散文作品

《猫》是现代作家[郑振铎](#)于1925年创作的一篇散文。

概念类型

[作品](#)

实例

[郑振铎](#)

現況...



[Main page](#)
[Community portal](#)
[Project chat](#)
[Create a new Item](#)
[Create a new Lexeme](#)
[Recent changes](#)
[Random Item](#)
[Query Service](#)
[Nearby](#)
[Help](#)
[Donate](#)

[Tools](#)

[What links here](#)
[Related changes](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Cite this page](#)
[Concept URI](#)

Item [Discussion](#)

house cat (Q146)

domesticated feline

[edit](#)

[cat](#) | [domestic cat](#) | [housecat](#) | [Felis silvestris catus](#) | [cats](#)

▼ [In more languages](#)

[Configure](#)

Language	Label	Description	Also known as
English	house cat	domesticated feline	cat domestic cat housecat Felis silvestris catus cats
Chinese	家貓	小型猫科动物	貓 貓咪
Traditional Chinese	家貓	小型貓科動物，野貓的亞種	貓 貓咪
Taroko	No label defined	No description defined	

[All entered languages](#)

<https://www.wikidata.org/wiki/Q146>

現況：知識圖譜的特點

- **知識獲取**
 - Wikipedia、WikiData、Crawler、不好說...
- **知識表示**
 - 普遍採用的方法是專家法和歸納法
- **知識存儲**
 - 各種 SQL 和各種 NoSQL

現況：知識圖譜的弱點

- 知識獲取
 - 需要人工維護，逐條編寫或持續微調模型
 - 利用資料模型，[Other] 是什麼？**(高品質)資料不足**
- 知識表示
 - 需要人工介入，提供專家意見和人工歸納
- 知識存儲
 - 基本上沒有什麼問題

人工不好嗎？

- 貴

- 台灣的「優勢」正在往非洲移動 (貴在便宜)。

- 慢

- 語料庫一做 30 年，退休了都還沒上線 (改版 2.0，繼續申請科技部/國科會/教育部計劃。反正當初審查的人也退休了，現在剛好換一批...)

- 不一致

- 你的判斷不是我的判斷。(Conll2003, **God** as B-Person)

基於語言學的「知識圖譜擷取系統」解決以下痛點：

- 知識獲取方法不再需要人工維護
 - 給它優良的文本，讓它「在家自學」
- 知識表示方法不再需要人工介入
 - 以透明可讀的方式儲存，在「視覺化」和「操作性」之間求取平衡

陳義過高
難以推廣

Rule-based =>

Labor-based =>

ML-Driven =>

Linguistics-Reasoned

貼近地面
落地應用

1. 輸入斷詞後的 POS/NER 標記資料



2. 抽取資訊關係 (只看實詞，忽略功能詞。以動詞、助動詞及情態動詞為軸心)

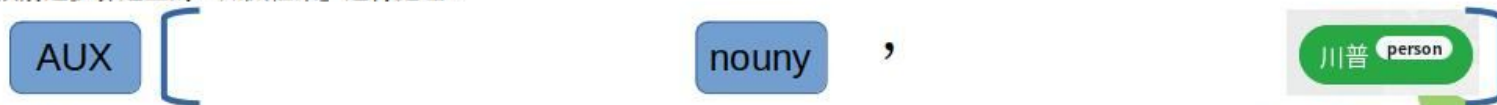


3. 只取 POS/NER 做進一步「抽象化」。建構知識框架結束。

1. 輸入新的句子。



2. 依前述步驟建立的「知識框架」進行處理。



3. 得到推論結果：

- i. 川普是人
- ii. 川普是候選人
- iii. 川普是共和黨
- iv. 川普是共和黨候選人

用 Loki 建立「知識的學習機制」

3. 文本內容 (一句一列)

∞ 無限字數

1

美國總統川普是一個身價數十億美元的企業家

▶ 單句分析

💬 生成同義句

⊕ 增加一列

⊖ 減少一列

🔄 移除空白列

▶ 全句分析

4. 勾選意圖所需參數

💡 使用小撇步

※通常是會變動的值或詞彙

※紅底請重新勾選參數

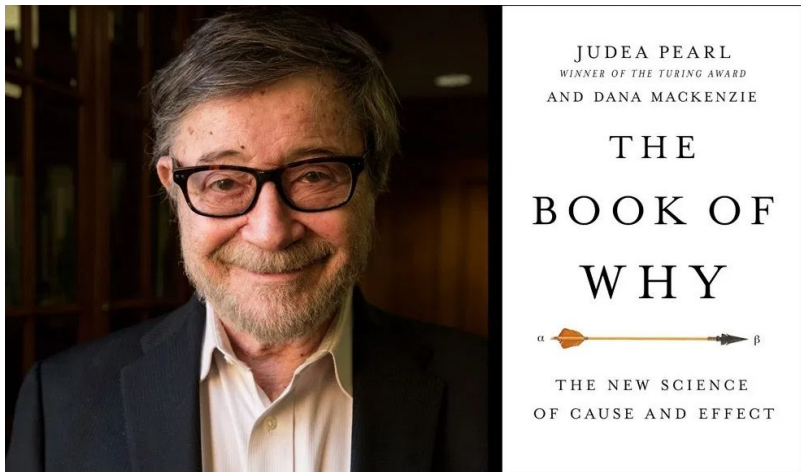
🗑️ 1 ☒ 美國 / ☒ 總統 / 川普 / 是 / 一個 / ☒ 身價 / ☒ 數十億美元 / 的 / ☒ 企業家

AI 的知識怎麼建？因果網路！

Intelligence = instincts + experiences

貝氏網路的發明人：Judea Pearl

因果網路的發明人：Judea Pearl



Training: 開關按下後，房間就亮了。

$P(\text{房間亮} | \text{開關按下}) \times P(\text{開關按下})$

Testing: 開關按下後，房間就 {__} 了。

Training: 公雞啼叫以後，天就亮了。

$P(\text{天亮} | \text{公雞啼叫}) \times P(\text{公雞啼叫})$

Testing: 公雞啼叫以後，天就 {__} 了。

知識由「因果網路」而來

命題：蘋果熟了，蘋果就會掉下來。

$P(\text{蘋果掉下來} \mid \text{蘋果熟了}) \times P(\text{蘋果熟了})$

推論：

蘋果掉下來的話，大概是因為它熟了。

$P(\text{蘋果熟了} \mid \text{蘋果掉下來}) \times P(\text{蘋果掉下來})$

命題裡的貝氏機率式無法產生推論裡的貝氏機率式

人類只要聽到命題，自然就會產生推論。

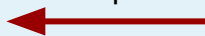
故，人類的智力運作不是貝氏機率式。

How our minds work?

命題：蘋果熟了，蘋果就會掉下來。

$P(X \mid \text{do}(Y))$

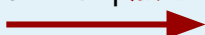
$P(\text{蘋果掉下來} \mid \text{熟了}(\text{蘋果}))$



推論：

蘋果掉下來的話，大概是因為它熟了。

$P(\text{蘋果掉下來} \mid \text{熟了}(\text{蘋果}))$

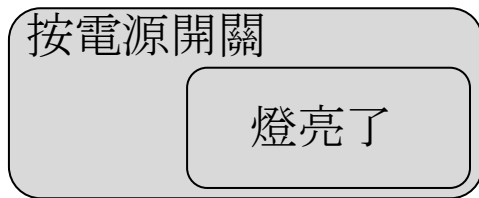


人類只要聽到命題，自然就會產生推論。

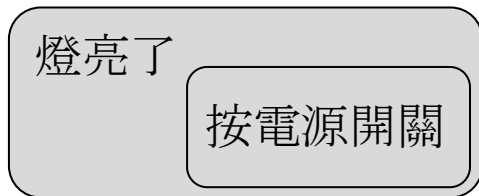
故，人類的智力運作是因果邏輯式。

機率網路

$$P(\text{燈亮了} \mid \text{按電源開關}) \times P(\text{按電源開關})$$



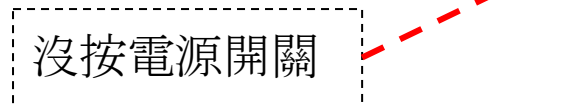
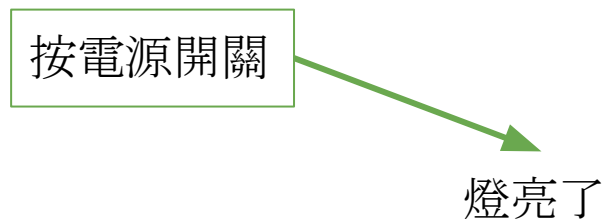
$$P(\text{按電源開關} \mid \text{燈亮了}) \times P(\text{燈亮了})$$



無法推論(無法建立知識):
燈亮不亮, 和有沒有按電源開關**高度相關**! 誰因誰果? 不知道...

因果網路

$$P(\text{燈亮了} \mid \text{do}(\text{按電源開關})) = \text{True}$$



$$P(\text{燈亮了} \mid \text{!do}(\text{按電源開關})) = \text{False}$$

推論(知識):
燈亮不亮, 和有沒有按電源開關是
因果關係



Language:



Knowledge:

語意理解引擎 4.0

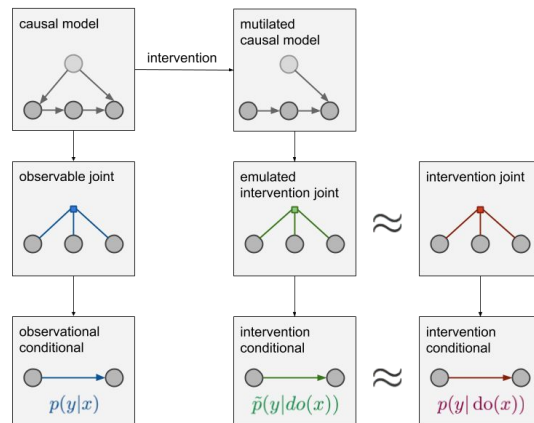
> Atlas_POC [意圖分析 (國語)]

建立基本意圖 建立進階意圖

設定生成模型

啟用	編號	意圖名稱	類型	句型
☒	1	A_sub_of_B	基本	1
☒	2	cause	進階	1
☒	3	effect	進階	1

Computation:



[美國][總統][川普]是[一個][身價][數十億美元]的[企業家]
 ['cause': [],
 'effect': [],
 'facts': ['朴槿良 是 韓國啦啦隊', '朴槿良 在 韓國', '朴槿良 是 新人', '朴槿良 的 月薪 為 30萬韓元']}]
 [cause]
 當你走進一個黑暗的房间，手按電源開關，燈就亮了，室內大放光明，帶給你許多方便，這就是福。====> 按電源開關
 [effect]
 當你走進一個黑暗的房间，手按電源開關，燈就亮了，室內大放光明，帶給你許多方便，這就是福。====> 燈就亮了
 cause: 按電源開關
 effect: 燈就亮了
 [cause] 買了美超微的股票，錢就賠掉了。====> 按電源開關
 [effect] 買了美超微的股票，錢就賠掉了。====> 燈就亮了
 cause: 買了美超微的股票
 effect: 錢就賠掉了
 [cause] 買了美超微的股票，錢就賠掉了。====> 按電源開關
 [effect] 買了美超微的股票，錢就賠掉了。====> 燈就亮了
 cause: 買了美超微的股票
 effect: 錢就賠掉了

Assignment

1. 依課堂中的建議調整你的 Workflow、Dataflow
2. 為你的專案準備一個 5 頁以內 (含封面) 的說明簡報
3. 為你的專案申請一個 Discord bot, 加入課程 Discord Server