B. Data Cleaning of Relations in ConceptNet

Different relations have different methods to correct. Here are the detailed data cleaning information of each relation. The process of data cleaning can divided into modify, move (to other relations) and delete. Each relation may contain modification of SurfaceText and description of relation, which is more detailed than appendix A.3. [A] and [B] in SurfaceText refer to Start and End fields in ConceptNet. Text quoted by " and " is relation or SurfaceText.

AtLocation

    Original SurfaceText

        "你可以在 [B] 找到 [A]"

        "[A] 在 [B] (裡、外、下)"

        "[A] 有 [B]"

    Modified SurfaceText

        "可以在 [B] 找到 [A]"

        "[A] 位在 [B] (裡、外、上面、下面)"

    Description

        The concept in End field is not always a physical location. It could be a virtual place, an organization, an object or a place that something can in, out, on, under or beside it.

    Modify

        Revert assertions when SurfaceText is "你可以在 [B] 找到 [A]" and "[A] 有 [B]".

        Move localizer in concept to SurfaceText to increase concept degree if

concept is separated, e.g., "[私房錢] 位在 [衣櫥 底下]" → "[私房錢] 位在 [衣櫥] 底下". ([case-dough] is under [the wardrobe])

Move localizer in SurfaceText to concept if it is in vocabulary table [1] after moving, e.g., "[廚房] 在 [家] 裡" → "[廚房] 在 [家裡]" ([kitchen] is at [home]).

Move

Move to relation "HasA" when the SurfaceText is "[A] 有 [B]" or [B] is an abstract conception. For example, The church has windows. Although, windows are at the location church, it's more precise that windows are part of the church or the church has the windows.

Delete

Delete concepts in End field if they're not a location. Concept is a location if its category in Cilin [2] is "Cb" (C-時間與空間, Cb-空間). If specific term like [康是美], a store sells cosmetic and medicine, not in Chinese WordNet or Cilin. Check whether specific term in Start field in ConceptNet by relation "IsA" is a location or not.

CapableOf

Original SurfaceText

"[A] 能做的事情有 [B]"

"[A] 會 [B]"

Modified SurfaceText

"[A] 會 [B]"

[1] vocabulary table of word embedding
[2] HIT IR-Lab Tongyici Cilin (Extended), 哈工大信息检索研究室同义词词林扩展版

Description

· Organism can do something

· Non-living object can do something or have certain functions.

Move

· Move to relation "HasSubevent" when SurfaceText is "[A] 能做的事情有 [B]" and concept in Start field is verb, e.g., "[單戀] 能做的事情有 [等待]" (you can [wait] when [one-sided love]). Concept in Start field in relation "CapableOf" should be an organism which can take actions or an object has certain functions. Hence, verb in Start field would be inappropriate.

· Move to relation "Causes" when SurfaceText is "[verb] 會 [B]". For example, "[被罵] 會 [難過]" ([being_scolded] can [sad]). The Surface-Text should be "[A] 會令人 (make someone feel) [B]", not "[A] 會 (can) [B]".

Delete

Concept in Start field should be non-living object or can't act actively.

Causes

Original SurfaceText

"因為 [A] 所以 [B]"

"[A] 會讓你 [B]"

"[A] 可能會引起 [B]"

"[A] 可能會帶來 [B]"

"[A] 之後可能會發生的事情是 [B]"

Modified SurfaceText

"因為 [A] 所以 [B]"

"[A] 會令人 [B]"

"[A] 會引發 [B]"

"[A] 會帶來 [B]"

"[A] 後會 [B]"

Description

· "因為 [A] 所以 [B]": causality between A and B. It's a generalized SurfaceText of relation "Causes". Most of the assertions with causal relation can use this. In contrast, it is not precise to describe the relation between A and B.

· "[A] 會令人 [B]": A makes someone feel B.

· "[A] 會引發 [B]": A brings about B, which B is usually something bad.

· "[A] 會帶來 [B]": A causes an effect on B or A will bring in B

· "[A] 後會 [B]": A causes an effect on B. After doing A, event or action B will be executed. One with causality and the other without causality. For example, "[吃] 後會 [飽]" and "[吃] 後會 [聊天]" (you will [feel stuffed] after [eating] and you will [chat] after [eating]).

Modify

Most of the assertions are not precise if SurfaceText is "因為 [A] 所以 [B]". Modify these data to SurfaceText "[A] 會令人 [B]" and "[A] 後會 [B]".

Move

Move to relation "HasSubevent" when SurfaceText is "因為 [A] 所以

116

[B]", e.g., "因為 [跳舞] 所以 [聽音樂]" → "[跳舞] 的時候，你會 [聽音樂]" (Because [dance], [listen to the music] → [listen to the music] when [dance]). Move to relation "MotivatedByGoal" if concept in Start field starts with "要", e.g., "因為 [要遺產] 所以 [結婚]" → "[結婚] 是為了 [要遺產]" (Because you [want heritage], you [marry] → you would [marry] because you [want heritage]).

CausesDesire

Original SurfaceText

"[A] 會讓你想要 [B]"

Modified SurfaceText

"[A] 會令人想要 [B]"

"[A] 的時候會想要 [B]"

Description

‧"[A] 會令人想要 [B]": A makes someone wants to do B.

‧"[A] 的時候會想要 [B]":

Someone wants to do B while doing A.

During a period of situation or event A, someone wants to do B.

Modify

"[A] 會讓你想要 [B]" → "[A] 的時候會想要 [B]" if A doesn't cause B to happen, e.g., "[孤單] 會令人想要 [睡覺]" → "[孤單] 的時候會想要 [睡覺]" ([lonely] makes someone wants to [sleep] → you want to [sleep] when you are [lonely]).

Desires

Original SurfaceText

Original SurfaceText

"[A] 想要 [B]"

"[A] 喜歡 [B]"

"[A] 痛恨 [B]"

"[A] 懼怕 [B]"

"[A] 不想要 [B]"

Modified SurfaceText

"[A] 想要 [B]"

"[A] 喜歡 [B]"

Description

A wants or wants to B.

A likes or likes to B.

Move

· Move to relation "NotDesires" when SurfaceText contains 痛恨、懼怕、不想要.

· Move to relation "CausesDesire" if concept in Start field is verb, e.g., "[畢業] 想要 [工作]" → "[畢業] 的時候會想要 [工作]" ([graduate] want [a job] → you want [a job] when [graduate]).

Delete

Delete non-living objects in Start field.

HasFirstSubevent

Original SurfaceText

"[A] 的時候，首先要 [B]"

Modified SurfaceText

"[A] 時，要優先 [B]"

Description

Someone does B first when A.

Move

Some of the assertions in relation "HasFirstSubevent" should be relation "HasPrerequisite". For example, "[煮飯] 時，首先要 [買食材]" → "[煮飯] 前，首先要 [買食材]" (you will [buy ingredients] first while [cooking] → you will [buy ingredients] before [cooking]). The meaning of relation "HasFirstSubevent" is you will do B first while doing A, not before. There is no relation "HasPrerequisite", which means doing something before something, in Chinese ConceptNet. Move these data to relation "MotivatedByGoal" and revert assertions, such as "[買食物] 是為了 [煮飯]" ([buy ingredients] in order to [cook]). It means you need to buy ingredients first to achieve the goal of cooking.

HasProperty

Description

B is a property of A.

Move

There are two meanings about SurfaceText "[A] 是 [B] 的" in Chinese. One is A has property B and another is A is belonged to B, e.g., "[小孩] 是 [聰明] 的" and "[書] 是 [他] 的" ([child] HasProperty [smart] and [the book] HasProperty [him]). In relation "HasProperty", the latter is wrong. B should be an adjective which can describe A, not an organism who can own something. Move to relation "HasA" if B is an organism.

HasSubevent

Original SurfaceText

"[A] 的時候，你會 [B]"

"在 [A]，你會 [B]"

"[A] 可能代表 [B]"

Modified SurfaceText

"[A] 的時候會 [B]"

"在 [A] 會 [B]"

Description

・"[A] 的時候會 [B]": Do B while A

・"在 [A] 會 [B]": In location A, you will do B.

Move

・Move to relation "SymbolOf" when SurfaceText is "[A] 可能代表 [B]".

・Move to relation "HasFirstSubevent" if concept in End field contains "先".

HasA

Original SurfaceText

There is no relation "HasA" in earlier version of Chinese ConceptNet.

Modified SurfaceText

"[A] 擁有 [B]"

MadeOf

Original SurfaceText

"[A] 可以用 [B] 製成"

"[A] 由 [B] 組成"

"[A] 的原料是 [B]"

"[A] 可以拿來做成 [B]"

Modified SurfaceText

"[A] 可用 [B] 製成"

"[A] 可用 [B] 組成"

"[A] 的原料是 [B]"

"[B] 可以做成 [A]"

Description

· "[A] 可用 [B] 製成": need to be processed

· "[A] 可用 [B] 組成": combination of different components

· "[A] 的原料是 [B]": food ingredients

· "[B] 可以做成 [A]": need to be processed

Modify

· Revert assertions when SurfaceText is "[B] 可以拿來做成 [A]".

· Modify the SurfaceText to be more precise. Some concepts are physical combination, and some concepts need to be processed.

MotivatedByGoal

Original SurfaceText

"[A] 是為了 [B]"

"當你想要 [B] 的時候你可能會 [A]"

"想要有 [B] 應該要 [A]"

"你會 [A] 因為你 [B]"

"[A] 的時候會想要 [B]"

Modified SurfaceText

"[A] 是為了 [B]"

"想要 [B] 的時候會 [A]"

"想要有 [B] 應該要 [A]"

Description

　·"[A] 是為了 [B]": Someone does A in order to B

　·"想要 [B] 的時候會 [A]": Someone does A in order to B

　·"想要有 [B] 應該要 [A]": Someone should do B if he/she wants B

Modify

　·Revert assertions when SurfaceText is "想要 [B] 的時候會 [A]" and "想要有 [B] 應該要 [A]".

　·The SurfaceText "想要 [B] 的時候會 [A]" contains two meanings. The first one is someone does A because they want result B, e.g., "想要 [讀書] 的時候會 [翻開書]" (You'll [open the book] when you want to [read]). The second one is someone does A because they don't want result B, e.g., 想要 [睡覺] 的時候會 [買咖啡] (You'll [buy a coffee] when you want to [sleep]). B is usually a physiological needs like sleeping, crying, sneezing or yawning If B is going to happen, you will take actions to deal with or prevent it. However, the meaning of relation "MotivatedByGoal" is doing A to achieve the goal B. Therefore, the second one is inaccurate, and one of the concept pair need to be modified.

Move

　·Move to relation "CausesDesire" when SurfaceText is "A 的時候會 想要 B".

・Move to relation "HasSubevent" when SurfaceText is "你會 A 因為你 B", e.g., "你會 [喝飲料] 因為你 [口渴]" → "[口渴] 的時候會 [喝飲料]".

NotDesires

Original SurfaceText

"[A] 懼怕/痛恨/厭惡 [B]"

Modified SurfaceText

"[A] 懼怕/痛恨/厭惡 [B]"

"[A] 的時候，懼怕/痛恨/厭惡 [B]"

Description

・"[A] 懼怕、痛恨、厭惡 [B]": A scares/hates/disgusts B

・"[A] 的時候，懼怕、痛恨、厭惡 [B]": Someone scares/hates/disgusts B while doing A

Modify

Create a new SurfaceText "A 的時候，懼怕、痛恨、厭惡 B" when the concept in Start field is a verb or a period of time.