

1. (50%) Regex and Constructions

※ 注意：請於試卷內之「非選擇題作答區」標明題號依序作答。

正規表示式 (regular expressions, regex) 是使用單個字串來描述、符應一系列符合某個句法規則的字串表示。比方說，我們可以用 `H(a|ä|ae)ndel` 這個 regex 來描述 Handel、Händel 和 Haendel 這三個字串。

1.1 (20%)

有一個能夠涵蓋以下所有姓名的正規表示寫成這樣：`"(\\w+),\\s(Mr|Ms|Mrs|Dr).?\\s(\\w+)"`
請解釋其中每個字元的意義。

| |
|------------------------|
| Smith, Mr. John |
| Davis, Ms Nicole |
| Robinson, Mrs. Rebecca |
| Armstrong, Dr Sam |
| Downey, Mr. Robert |

1.2 (30%)

寫出一個正規表示法，能用來抽取語料庫中所有的【一 X 一 Y】的構式實例。所謂的構式在這裡的定義是，一種約定俗成的「形式-意義」結合體，其整體的意義不是由個別組成成分的語意所組合的。例如：‘一心一意’、‘一來一往’。但注意，X 與 Y 可以是超過一個漢字，例如‘一步一腳印’。

如果要避免寫出的 regex 有過度蔓生 (overgenerate) 的問題，有什麼作法？例如，如何濾掉‘一天一餐’，‘一張一塊錢’、‘一天又一天’的偽構式？

見背面

題號： 37

國立臺灣大學 111 學年度碩士班招生考試試題

科目： 語言與計算方法

題號：37

節次： 3

共 3 頁之第 2 頁

2. (50%) 機器的注意力 ※ 注意：請於試卷內之「非選擇題作答區」標明題號依序作答。

小明最近在研究機器學習，他發現有一個系統有五個小模組(甲乙丙丁戊)。只要輸入句子(標點符號不予考慮)，系統的五個模組就會分別給出字詞之間的連結訊息。表一則是他從語料庫中找出的 8 個句子，分別輸入系統得到的有字詞索引的輸出句，以及表格二的連結訊息。但是要注意：(1). 每個模組輸出的連結訊息可以只有一個，兩個，或沒有。(2). 表格二其中有一格的連結訊息是錯誤的。

舉例來說，輸入「我想讀語言學。」，會得到「我 1 想 2 讀 3 語言學 4。」，以及

3>1, 3>4, none, 3>2, none

你的任務是

1. 把句子 1-8 和 A-H 對應起來。
2. 填入表格二中空白的連結訊息(a-i)，並指出哪個訊息是錯的。

表一

| 句子編號 | 輸入句與輸出句 |
|------|---|
| 1 | 老師會在放假前送幾位用功的好學生一些禮物。 老師 1 會 2 在 3 放假 4 前 5 送 6 幾位 7 用功 8 的 9 好 10 學生 11 一些 12 禮物 13。 |
| 2 | 晚餐他已經吃完了，現在只想再喝點飲料而已。 晚餐 1 他 2 已經 3 吃 4 完 5 了 6，現在 7 只 8 想 9 再 10 喝 11 點 12 飲料 13 而已 14。 |
| 3 | 林莉的同事們都非常友善，她們總是願意給她很多建議。 林莉 1 的 2 同事們 3 都 4 非常 5 友善 6，她們 7 總是 8 願意 9 給 10 她 11 很多 12 建議 13。 |
| 4 | 我上週拜訪王董的時候，看到了他的作品。 我 1 上週 2 拜訪 3 王董 4 的 5 時候 6，看 7 到 8 了 9 他 10 的 11 作品 12。 |
| 5 | 下個週末是父親節，你和你的家人們打算怎麼過？ 下 1 個 2 週末 3 是 4 父親節 5，你 6 和 7 你 8 的 9 家人們 10 打算 11 怎麼 12 過 13？ |
| 6 | 這件事我以前沒聽說過，但之後會再仔細地思考。 這 1 件 2 事 3 我 4 以前 5 沒 6 考慮 7 過 8，但 9 之後 10 會 11 再 12 仔細 13 地 14 思考 15。 |

接次頁

題號： 37

國立臺灣大學 111 學年度碩士班招生考試試題

科目： 語言與計算方法

題號：37

節次： 3

共 3 頁之第 3 頁

| | |
|---|---|
| 7 | 他每晚十一點前都會把自己的每件衣物洗好。 他 1 每 2 晚 3 十一點 4 前 5 都 6 會 7 把 8 自己 9 的 10 每件 11 衣物 12 洗好 13。 |
| 8 | 王晴很用心地照顧她的父母，他們應該恢復得很好。 王晴 1 很 2 用心 3 地 4 照顧 5 她 6 的 7 父母 8，他們 9 應該 10 恢復 11 得 12 很 13 好 14。 |

表二

| | 甲 | 乙 | 丙 | 丁 | 戊 |
|------------|-------------|-----------|------|-------|-----------|
| sentence A | 1>6 | 6>11 | none | 6>2 | (a) |
| sentence B | 6>3; 10>7 | 10>11 | 3>1 | 10>9 | 7>3; 11>1 |
| sentence C | 4>2; 11>2 | (b) | none | 11>9 | none |
| sentence D | 13>6; 13>10 | 13>5 | 10>8 | 13>11 | none |
| sentence E | (c) | 13>12 | 12>9 | (d) | 9>1 |
| sentence F | 7>4; 15>4 | (e) | none | 15>11 | none |
| sentence G | 3>1; 7>1 | 3>4; 7>12 | (f) | none | 10>4 |
| sentence H | (g) | 5>8 | 8>6 | (h) | (i) |

試題隨卷繳回