# 以詞彙表為基礎的知識本體 雛型建構研究

以「公司治理」領域知識為例

2018.10.09

洪振倫 陳易宣 王翎瑋



# 目錄 | CONTENTS



研究目的與動機 研究設計 在公司治理的應用 研究結論

## 研究目的與動機



## 研究設計



### 詞彙表建置

文本蒐集與處理 關鍵詞篩選 建立領域詞彙表

#### 文本處理並產生詞庫

進行中文斷詞; 產生本研究之公司治理詞庫

#### 語意網路建構

利用TFIDF為詞項建立特徵向量進行詞項間的相似度計算;並依據所篩選出之詞項關聯程度,形成語意網路

#### 語意結構分析

透過分群(clustering)將互有緊密關聯的詞彙聚集「群集」; 以詞彙共現性、同義、反義分析與句型分析等方法,以建構 出本體知識之概念階層

#### 雛型本體產出

將語意網路轉換為以 OWL (web ontology language)語言所表達 領域本體知識

### 雛型本體評估

設計適當問卷,進行妥適程度 之評估,以檢驗「雛型本體」 可否提供後續再利用。

## 在公司治理的應用

### 詞彙表建置

01

- 1. 期刊以會計研究月刊中有關公司治理等議題 的文章共4篇為主,並於全國博碩士論文網下載 公司治理議題論文摘要共150篇
- 2.輔以文字探勘技術 , 找出233個重要關鍵詞
- 3.全文搜尋詞彙定義 + 同詞異義的處理 + 異詞同義的處理

## 文本處理並產生詞庫

02

- 1.針對所建構之詞彙表,輸入中研院中文斷詞系統,協助建置公司治理領域詞庫
- 2.在進行中文斷詞後,會將斷詞結果 與本研究所制訂之詞彙表中的詞項進 行比對

### 語意網路建構

03

- 1.基於前一步驟詞庫所建立之詞彙向量空間,利 用TFIDF為詞項建立特徵向量,相似度計算總連 結筆數n為20,301筆
- 2.進行詞項間的相似度計算,計算出詞項間之關聯程度,經篩選後僅餘375筆
- 3. 並依據所篩選出之詞項關聯程度,進行詞彙間聯結,分別以公司治理、內部控制、企業風險管理為中心分別做呈現

## 在公司治理的應用

### 語意結構分析

04

- 1.透過分群(clustering)演算法,將互有緊密關聯的詞彙聚集「群集」,並建議出「群集中心」
- 2. 利用平面分群從語意網路中辨識結構相似的概念群,再針對各概念群進行階層
  - A.辨識結構相似概念群
  - B.調整「分群演算法」結果—定義 「中心—相關」概念
  - C.發展概念階層關係
  - D.組織與調整:從屬關係、同義關係

### 雛型本體產出

05

依據步驟三至五進一步將語意網路轉換 為以OWL (web ontology language) 語言所表達領域本體知識

### 雛型本體評估

06

設計適當問卷,並委由五名具有公司治理知識之專家進行問卷調查。結果顯示本研究提出的建置方法,五名專家對於該雛型本體的同意題數比例都非常高

## 研究結論



本研究主要的貢獻是將A-G方法,改良為可執行的本體建置程序,具體 以循序步驟及配套技術來實踐



此初步顯示領域專家應可根據此雛型本體為基礎,再繼續修正為最終之應用 本體 Thank you!

