

中文新聞內容標誌初階研究

以可延伸標誌語言(XML)

標示科學新聞為例

謝瀛春、黃學碩、維習安、瑞克·傑立夫、謝清俊、雷約翰

香港

2004年5月24日

研究目的

- 建立中文新聞內容正式通用的標誌方式
- 未來中文新聞內容資訊交換之參考
- 新聞寫作教學(像如何寫新聞)
- 新聞業界修改、編輯新聞(像新聞六要素是否遺漏)
- 使用者深入檢索新聞內容(像新聞事件之關係)
- 研究用途(像內容分析，探究新聞事件之開始、過程至結束)

研究期間

- 本研究於2000年開始
- 嘗試研究階段
- 中斷一年、斷斷續續進行
- 目前(2004)嘗試以數位典藏通訊電子報(NDAP Newsletter)內容為實驗測試對象
- 人力(缺乏懂XML及新聞之雙重專長者)是最大困難

NDAP: National Digital Archives Program

實驗案例的選取

- 二十則中文科學新聞，且限於純淨新聞 (straight news)
- 最符合制式新聞寫作的要求，而純淨新聞也是最普遍通用的寫作模式

分析依據及參照

- 新聞寫作之學理與實務
- 文件製碼協定(TEI; Text Encoding Initiative)及後設資料(Metadata)
- 圖書館界都柏林核心欄位協定(Dublin Core)及報業資訊交換格式協定(NITF)

新聞寫作之學理與實務

- 新聞六要素(即5W1H; who, what, when, where, why, and how)
- 新聞事件(event)
- 新聞要素與新聞事件彼此關係(含event之間及其與5W1H之關係)
- 新聞寫作結構(如導言、主體、結尾)

詳見新聞寫作結構圖(word檔)

新聞六要素 (*5W1H*)

人 who--subject (*person, organization*) of news
event

事 what--happened, happening

地 where--place

時 when--time thing(s) happened

為何 why--cause

如何 how--situation & process

E.L.Shuman, 1894

*(from Frank Luther Mott, News in America,
Harvard University Press, Mass. 1952, p.158)*

新聞事件之意涵

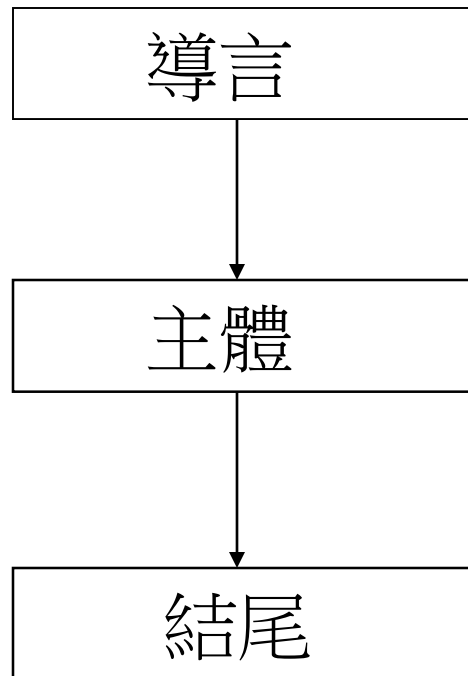
The occurrence which gives rise to media coverage will have fulfilled one or more, or an amalgam of **NEWS VALUES**(新聞價值).

- key event(關鍵事件)
- similar event(類似事件)
- thematically related event(主題關聯事件)

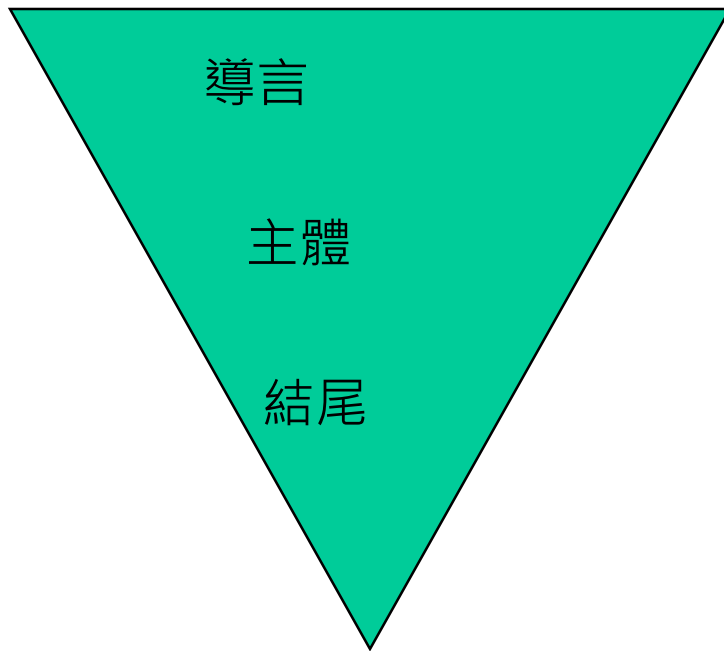
A Dictionary of Communication and Media Studies

Arnold (Fourth edition), 1997, pp.78-79

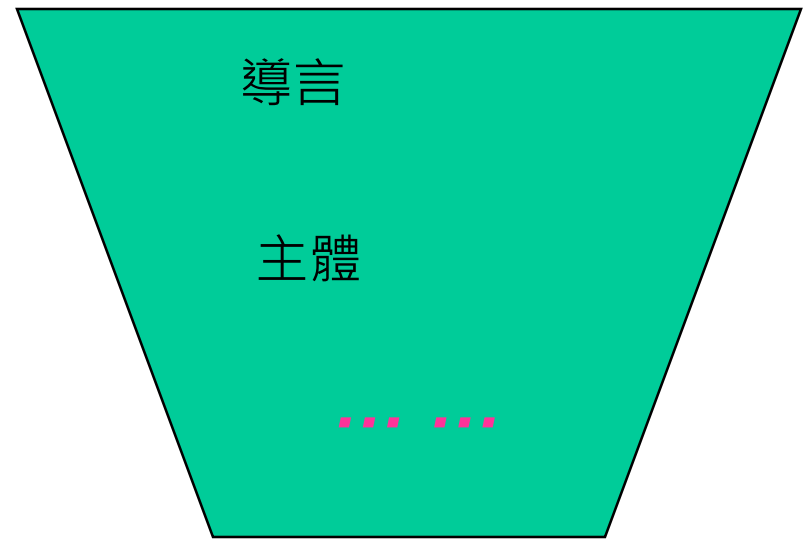
新聞寫作結構



純淨新聞寫作 (倒寶塔、倒金字塔)



記者

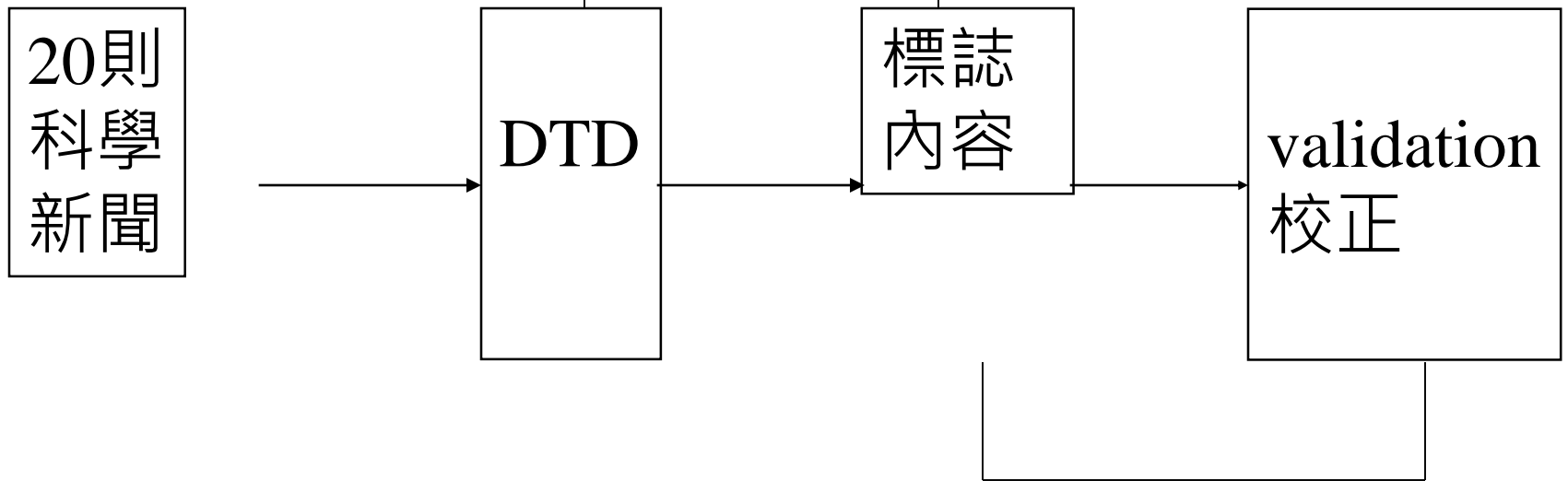


無結尾

編輯

分析程序

XML



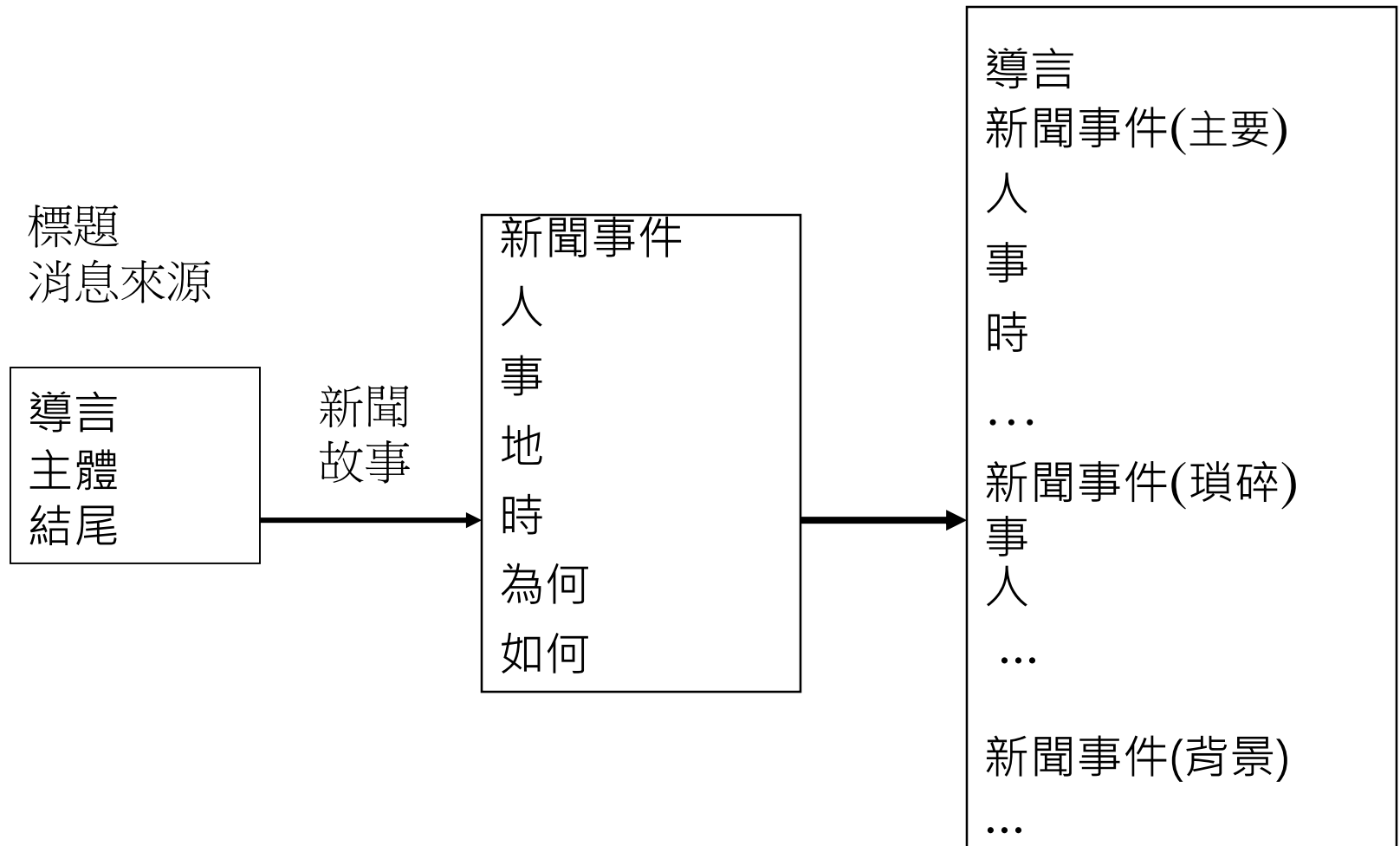
SP

Ultra Edit Textpad

分析程序及標誌內容

- 先分析中文新聞內容
 - 以科學新聞，且限於純淨新聞
- 再以XML標誌中文新聞內容
 - 限於科學新聞、純淨新聞
 - 不包括特寫、專題報導、深度報導、調查報導及評論
- 以語意單位為標誌之依據
 - 標誌之基本單位是以新聞內容前後文關係之語意為準

標誌那些內容?



中文DTD

```
<?xml version="1.0" encoding="big5"?>  
<!--NEWS XML encoding date: 2000-1-4-->  
<!DOCTYPE 科學新聞 [  
<!ELEMENT 科學新聞 (#PCDATA |科學新  
聞內容)* >  
<!--          寫作結構          -->  
<!ELEMENT 科學新聞內容 ( 標題 | 來源 | 導言  
| 主體 | 結尾)*>  
<!ELEMENT 標題          (#PCDATA)>
```

中文DTD

```
<!ELEMENT 來源      (#PCDATA|記者名|報紙  
名|通訊社名|外電日期|刊載日期|發稿地點|  
版面位置)*>  
<!ELEMENT 記者名    (#PCDATA)>  
<!ELEMENT 報紙名    (#PCDATA)>  
<!ELEMENT 通訊社名  (#PCDATA)>  
<!ELEMENT 外電日期  (#PCDATA)>  
<!ELEMENT 刊載日期  (#PCDATA)>  
<!ELEMENT 發稿地點  (#PCDATA)>
```

中文DTD

```
<!ELEMENT 版面位置      (#PCDATA)>
<!ELEMENT 導言          (#PCDATA |事件)*>
<!ELEMENT 事件          (#PCDATA |人|事|時|
地|如何|為何)*>
<!ELEMENT 人            (#PCDATA)>
<!ELEMENT 時            (#PCDATA)>
<!ELEMENT 地            (#PCDATA)>
<!ELEMENT 事            (#PCDATA |人|時|地|如
何|為何)*>
```


中文DTD

<!ELEMENT 如何 (#PCDATA |人|事|時|地)*>

<!ELEMENT 為何 (#PCDATA |人|事|時|地)*>

<!ELEMENT 主體 (#PCDATA |事件)*>

<!ELEMENT 結尾 (#PCDATA |事件)*>

<!ATTLIST 科學新聞內容 id ID
#REQUIRED>

<!ATTLIST 事件 id ID #REQUIRED
類型 (主要|次要|其他|背景說明|細節)
“細節”

中文DTD

陳述方式 (事實|評論|夾敘夾議) “事實”
內容性質 (新聞消息|科學消息) “新聞消息”
相關事件 CDATA #IMPLIED

關係類型 CDATA #IMPLIED>

<!ATTLIST 人 id ID #REQUIRED>

<!ATTLIST 事 id ID #REQUIRED>

<!ATTLIST 時 id ID #REQUIRED>

<!ATTLIST 地 id ID #REQUIRED>

<!ATTLIST 如何 id ID #REQUIRED>

<!ATTLIST 為何 id ID #REQUIRED>

<!--End of NEWS DTD-->

]>

(參見中文新聞內容標誌實例)

研究未竟之處

- 本研究尚未在其他複雜系統測試
- 未在中文報紙資料庫測試
- 新聞事件(event)之分析，尚在初步階段(primitive)
- 新聞事件與新聞六要素之關係，尚待更深入分析
- 語意單位之標誌宜再和語言學者商議

進一步研究

- 純淨新聞之結構分析(如倒寶塔寫作)，需繼續深化，以求窮盡
- 已做的語意單位標誌，宜重新評估、修正，並力求和語言學者已做之斷詞、斷句等核對、確認
- 逐步(從已做實例)建人名檔(含職銜等)——如同義辭典、字典(thesaurus、dictionary)

研究建議

- 需要團隊合作(Collaboration)新文化
- 跨學門、跨領域: 資訊科技、新聞、語言、寫作、標準、管理等
- 傳播溝通能力、移情心理
- 超越政治、社會之侷限
- 以大中華同文同種之文化考量優先
- Communication , Collaboration and Digitalization

特別感謝

- 香港中文大學中文新聞標示語言協會之邀；黃錦輝教授、盧嘉亮經理、譚秀儀小姐以及其他各位會議工作人員
- 台灣這次與會專家熱烈共襄盛舉；大家在急促的時間內，且無旅費支援之情形下，應允參加並發表研究成果。特別感謝朱四明、莊道明、傅心家、謝清俊四位教授報告，林蘋、王美玉兩位教授熱烈參加，更要特別謝謝我的研究助理賈馨潔小姐幫忙各項繁雜聯絡事宜。
- 期望未來我們能獲經費支持，在台灣接辦下次會議，邀請大家參與中文新聞標示語言之進一步研發成果。