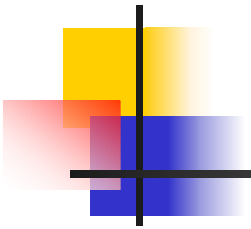
Chinese News Content Markup by XML

Ying-chun Hsieh, Mils Y. M. Chou

- Hsieh is a professor of Department of Journalism, College of Communication, National Chengchi University
- Chou is a Ph. D. candidate at Department of Information Management, National Taiwan University

September 22, 2002



OUTLINE

- Background
- Contributors
- Purposes of News Content Markup
- Basic Concept of News
- Structure of System Design
  - System Limitation at the Present Stage
- System Demo
- Conclusion

Background

- National Digital Archives Program (NDAP)
- Thematic group of News Data
  - ***1. Project of Local Archives in Taiwan***
  - ***2. Project of National Library's Collections of Periodicals and Newspapers***
  - ***3. Project of the Digitalization of CTS's TV News***
  - ***4. Project of the Digitalization of 'World Daily News'***
  - ***5. Project of NDAP Newsletter***

Contributors

- Ching-chun Hsieh, Institute of Information Science, Academia Sinica
- Mary Ku, Hsin-je Chia and I-lin Chen, NDAP Newsletter, Computing Centre, Academia Sinica
- Ruey-bin Wei and Ming-yean Lin, Department of Information Management, College of Gin Wen Technology

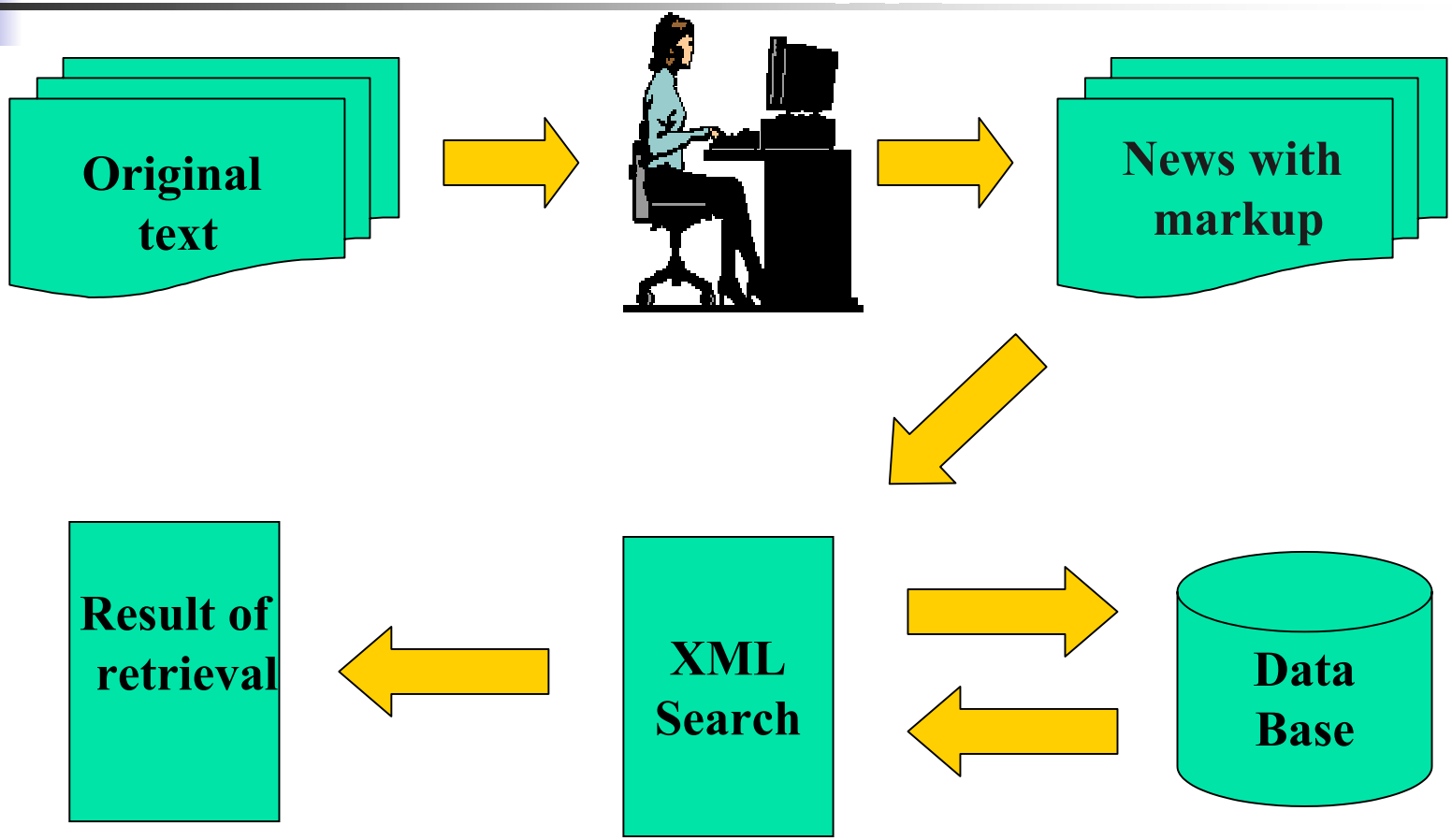
Purposes of News Markup

- Information Sharing
- Information Retrieval
- News Processing (Reporting, Editing, Printing,…)
- Journalism Education
- Journalism Research

Basic Concept of News

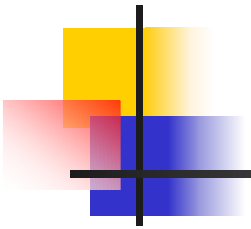
- News writing is a kind of standard format, especially on straight news
- News events; major, minor, sub-event
- News elements: who, what, when, where, why, and how

Structure of System Design



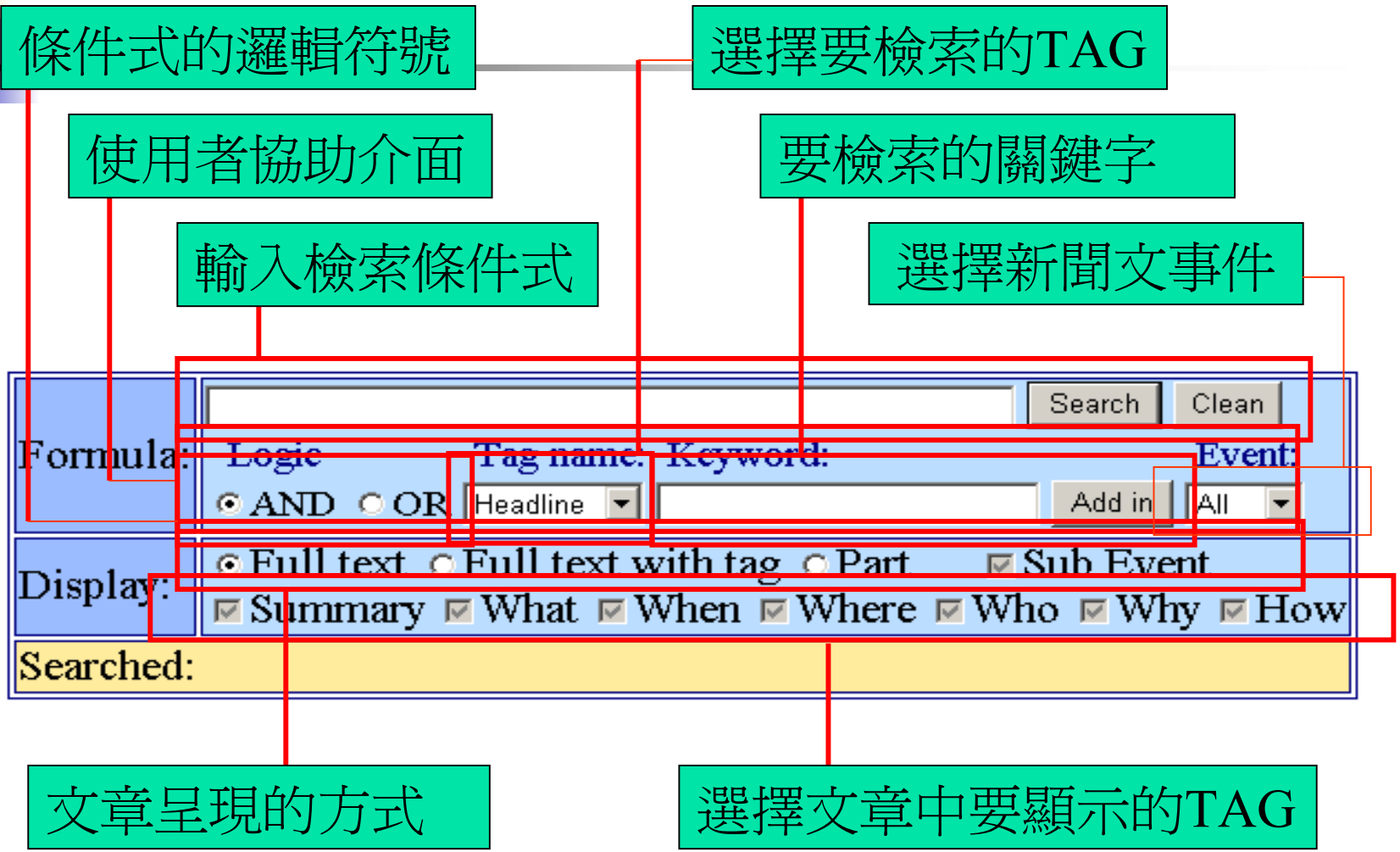
Limitation of the present system

- **Not for full text retrieval (**本系統非全文檢索**)**
- **No markup for tables yet (**本系統不處理表格問題**)**
- **No hyperlink for image and literatures yet (**本系統目前暫不處理圖片與相關文章連結**)**



System DEMO

介面說明


條件式的邏輯符號
選擇要檢索的TAG
使用者協助介面
要檢索的關鍵字
輸入檢索條件式
選擇新聞文事件
Search
Clean
Formula:
Logic
Tag name:
Keyword:
Event:
AND
OR
Headline
Add in
All
Display:
Full text
Full text with tag
Part
Sub Event
Summary
What
When
Where
Who
Why
How
Searched:
文章呈現的方式
選擇文章中要顯示的TAG

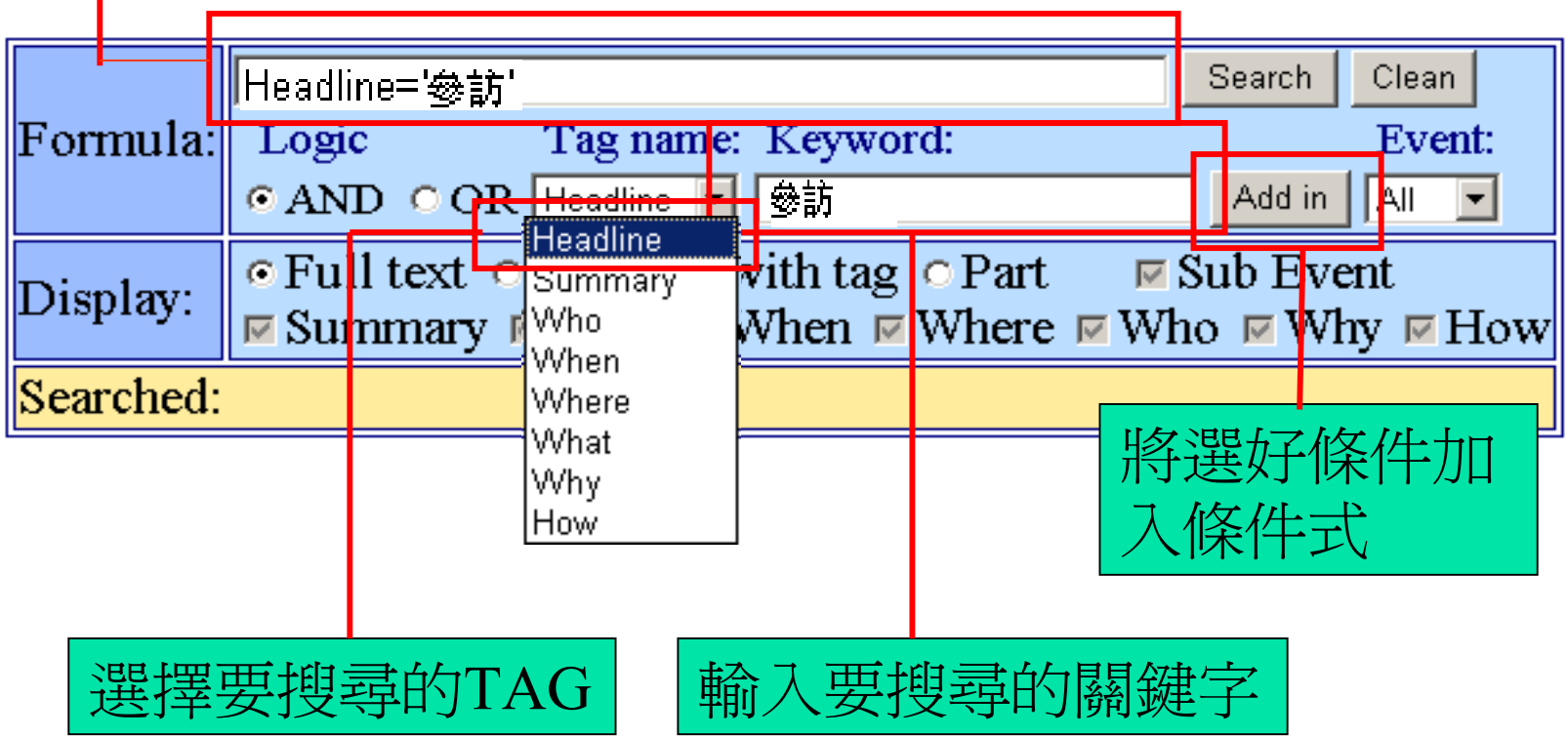
範例 一

- 魏瑞彬同學想要檢索電子報內所有有關參訪活動的新聞。

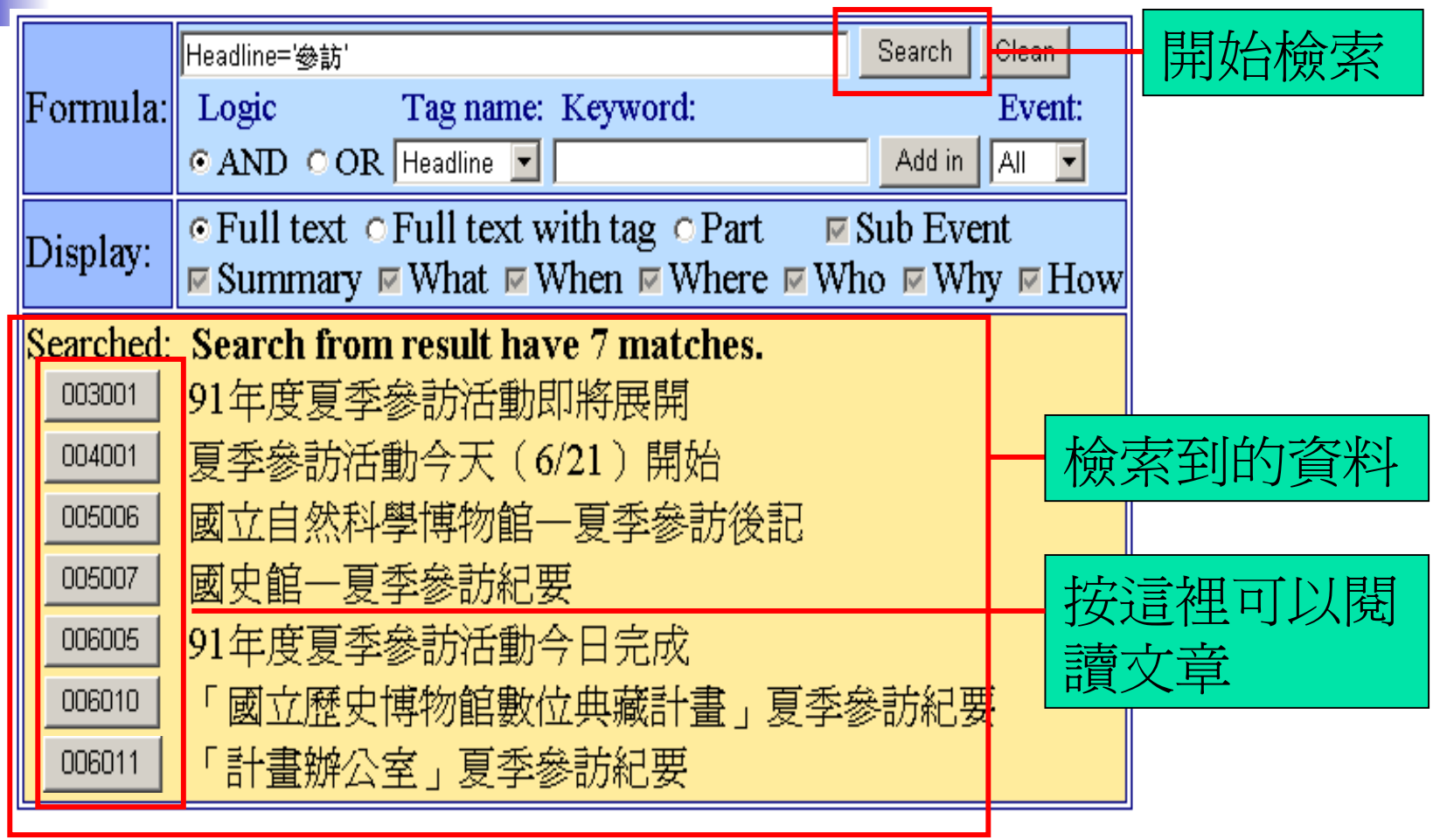
範例 1-輸入條件式

輸入完的條件式，檢索標題內含有‘參訪’的文章

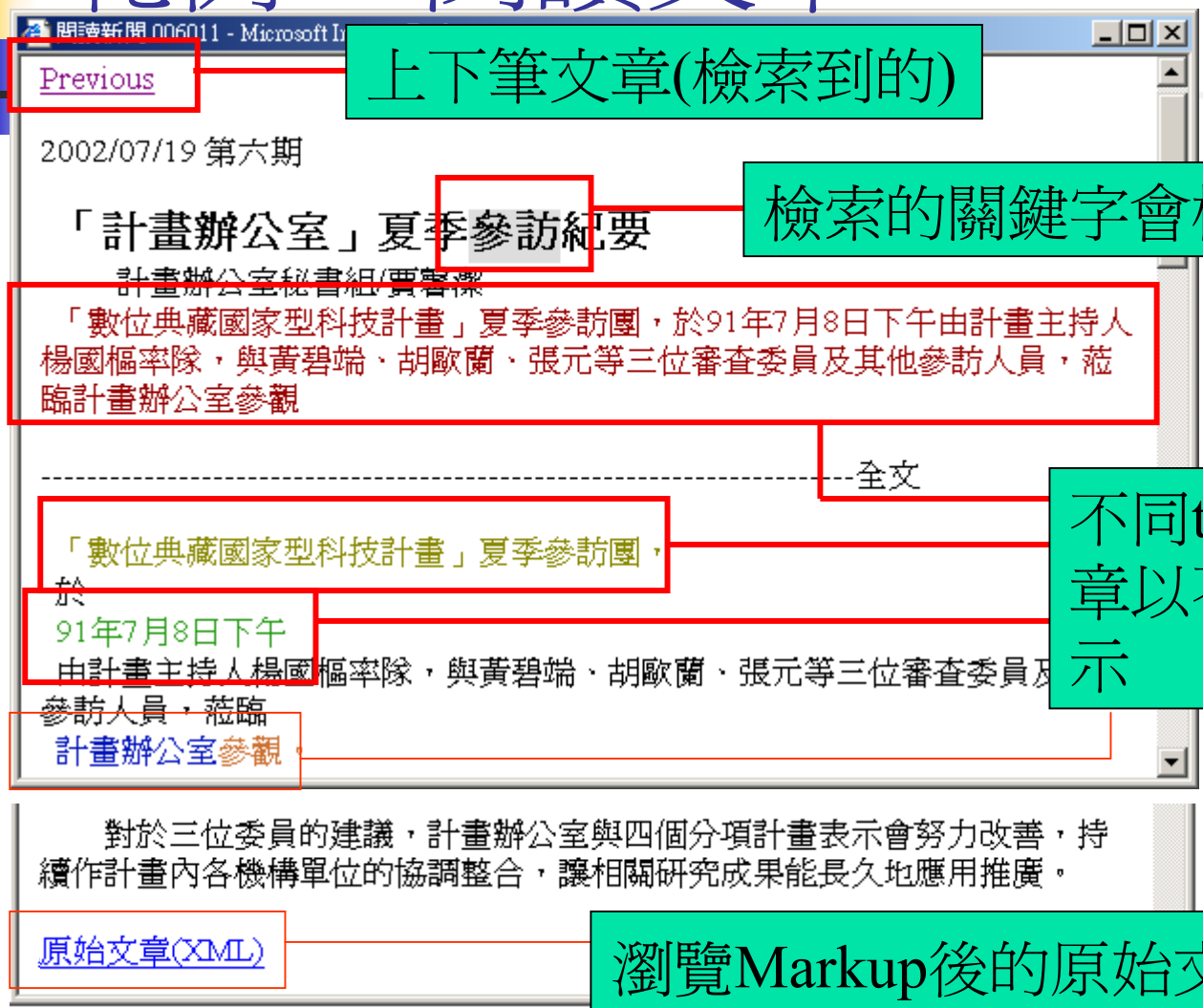
條件式: **Title = ‘參訪’**



範例 1-檢索資料


Formula:
Headline='參訪'
Search
Clean
Logic
Tag name:
Keyword:
Event:
AND
OR
Headline
Add in
All
Display:
Full text
Full text with tag
Part
Sub Event
Summary
What
When
Where
Who
Why
How
Searched: Search from result have 7 matches.
003001 91年度夏季參訪活動即將展開
004001 夏季參訪活動今天（6/21）開始
005006 國立自然科學博物館—夏季參訪後記
005007 國史館—夏季參訪紀要
006005 91年度夏季參訪活動今日完成
006010 「國立歷史博物館數位典藏計畫」夏季參訪紀要
006011 「計畫辦公室」夏季參訪紀要
開始檢索
檢索到的資料
按這裡可以閱讀文章

範例 1-閱讀文章



上下筆文章(檢索到的)

檢索的關鍵字會標示起來

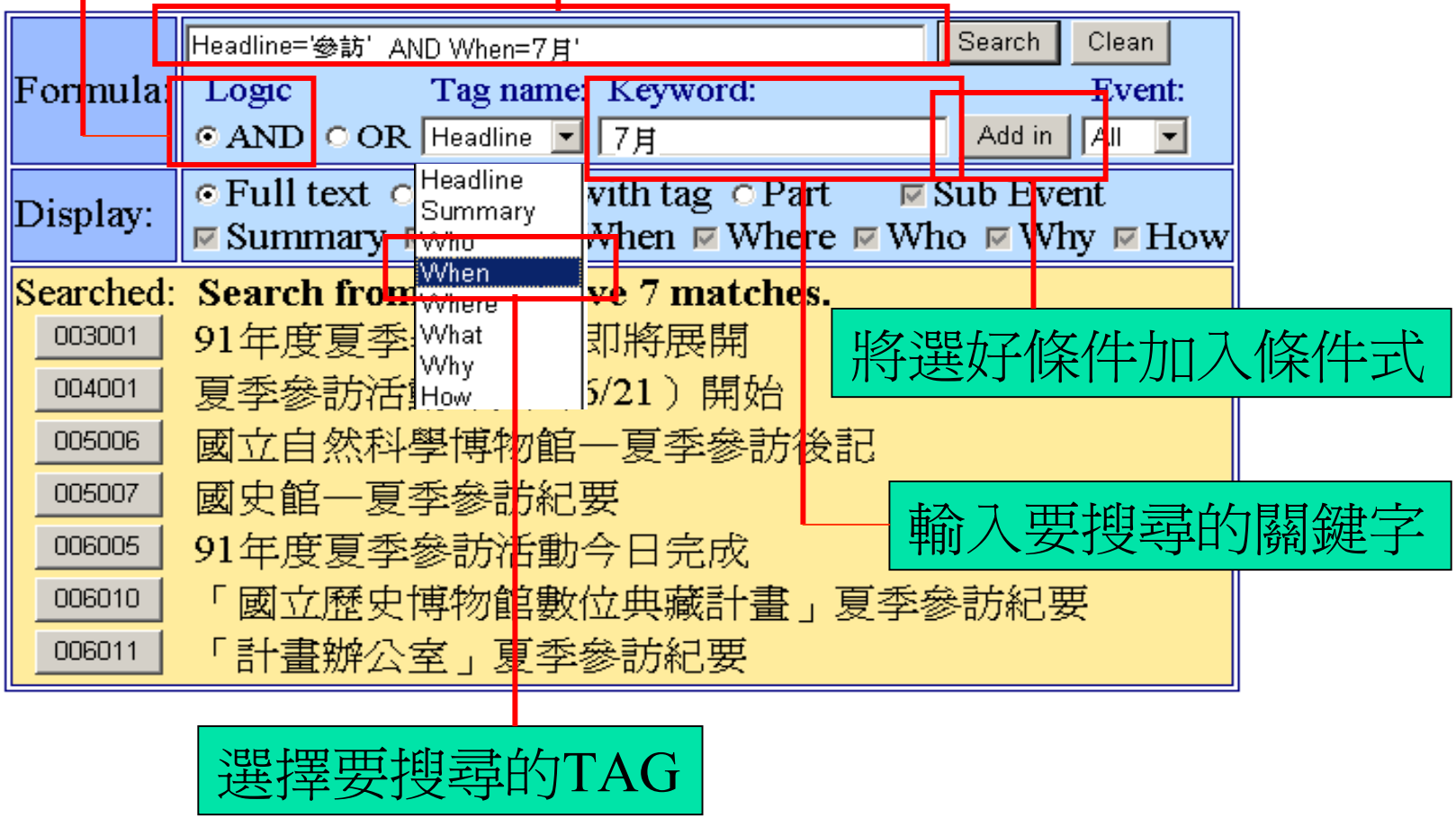
不同tag內容的文章以不同顏色標示

瀏覽Markup後的原始文章(XML)

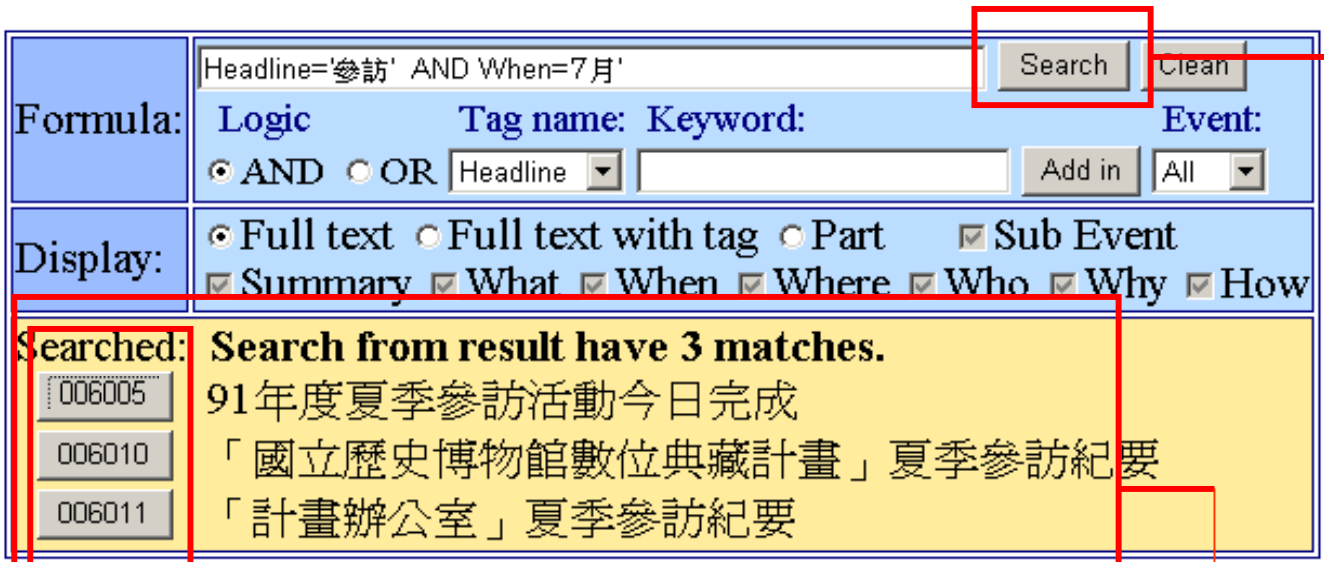
範例 二

- 魏同學想要將檢索範圍限定在七月份的參訪活動。

範例 2-輸入條件式


條件式的邏輯符號
條件式: Title = '參訪' AND When = '7月'
Headline='參訪' AND When=7月'
Search
Clean
Formula:
Logic
Tag name:
Keyword:
Event:
AND
OR
Headline
7月
Add in
All
Display:
Full text
Part
Sub Event
Summary
Where
Who
Why
How
Headline
Summary
When
What
Why
How
Searched:
Search from
matches.
003001
91年度夏季
即將展開
004001
夏季參訪活
/21）開始
005006
國立自然科學博物館—夏季參訪後記
005007
國史館—夏季參訪紀要
006005
91年度夏季參訪活動今日完成
006010
「國立歷史博物館數位典藏計畫」夏季參訪紀要
006011
「計畫辦公室」夏季參訪紀要
將選好條件加入條件式
輸入要搜尋的關鍵字
選擇要搜尋的TAG

範例 2-檢索資料



開始檢索

檢索到的資料

按這裡可以閱讀文章

範例 2-閱讀文章

上下筆文章(檢索到的)

檢索的關鍵字會標示起來

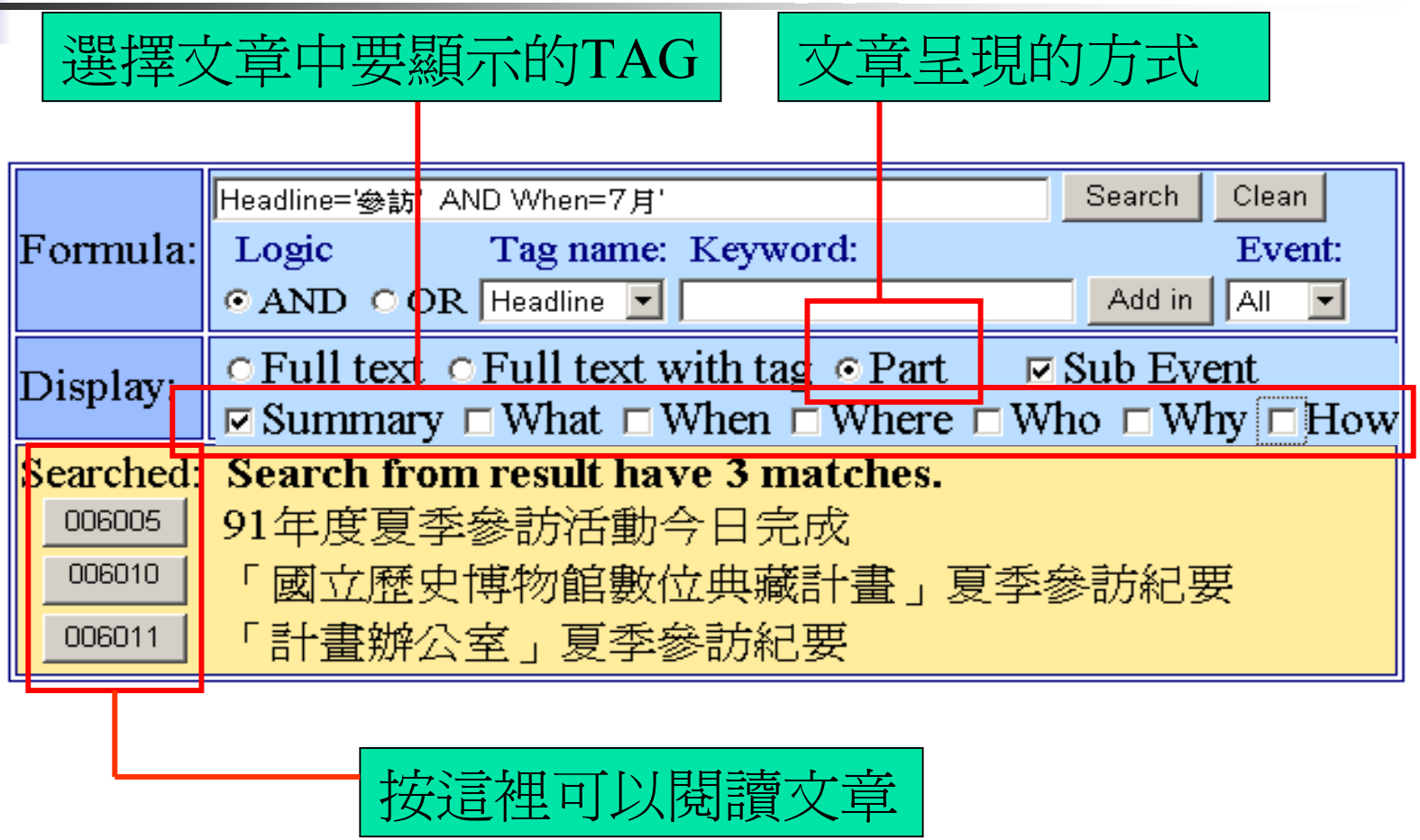
不同tag內容的文章以不同顏色標示

瀏覽Markup後的原始文章(XML)

範例 三

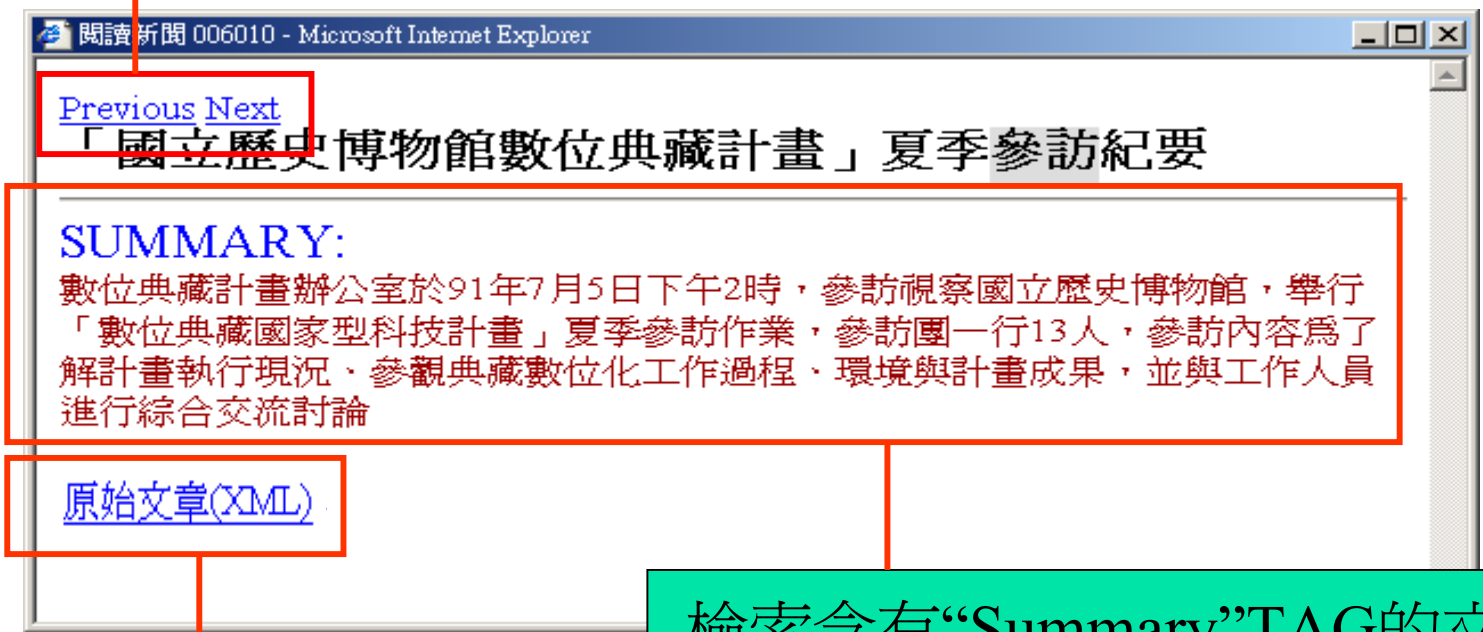
- 魏同學想要檢索所有含有摘要-“**Summary**”標籤的文章~~而他也只想閱覽摘要的部分**!!**

範例 3-輸入條件式


選擇文章中要顯示的TAG
文章呈現的方式
Formula:
Headline='參訪' AND When=7月'
Search
Clean
Logic
Tag name:
Keyword:
Event:
AND
OR
Headline
Add in
All
Display:
Full text
Full text with tag
Part
Sub Event
Summary
What
When
Where
Who
Why
How
Searched:
Search from result have 3 matches.
006005
91年度夏季參訪活動今日完成
006010
「國立歷史博物館數位典藏計畫」夏季參訪紀要
006011
「計畫辦公室」夏季參訪紀要
按這裡可以閱讀文章

範例 3-閱讀文章

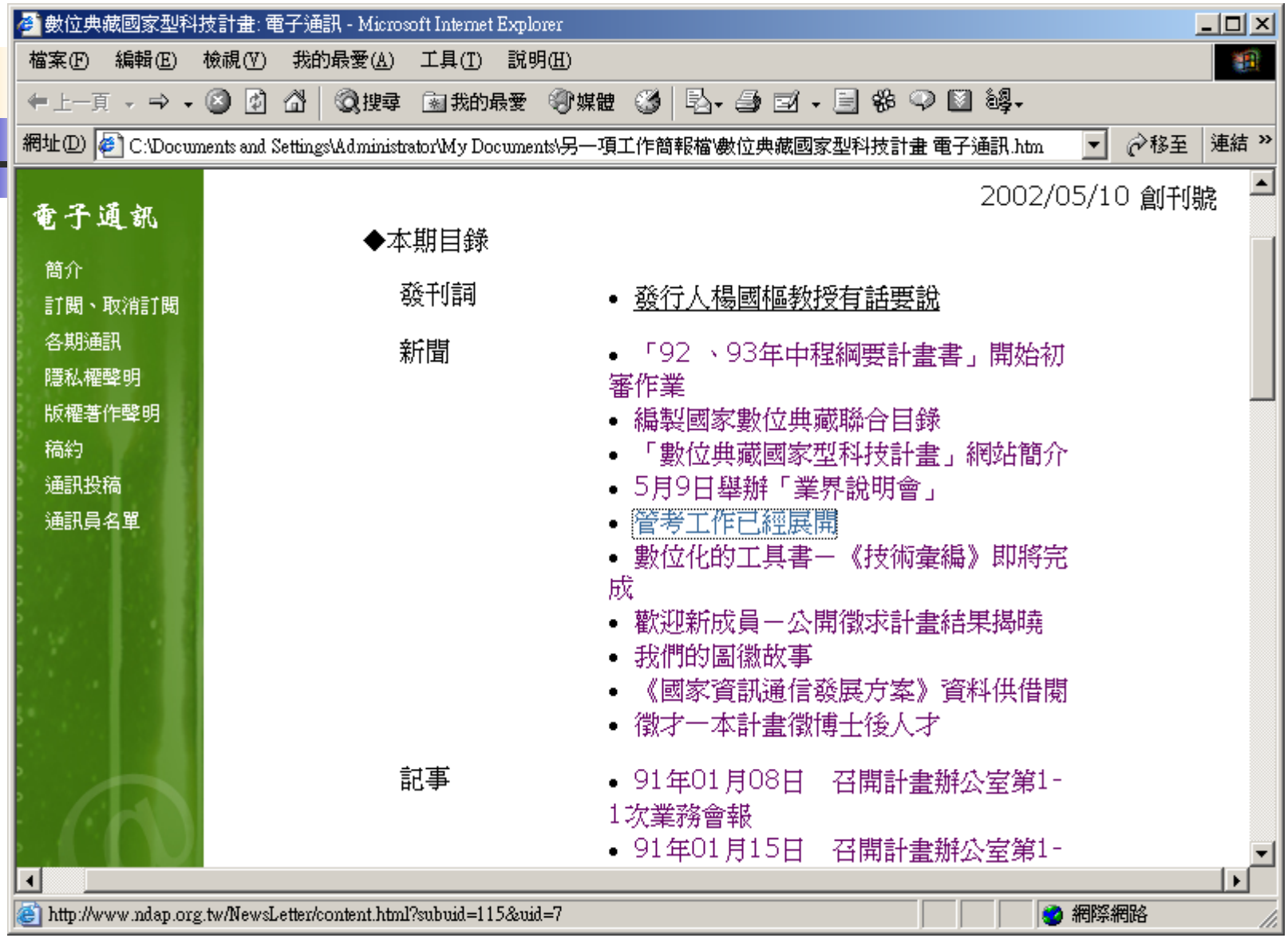
上下筆文章(檢索到的)



檢索含有“Summary”TAG的文章

瀏覽Markup後的原始文章(XML)

原始新聞文件


數位典藏國家型科技計畫: 電子通訊 - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛 媒體
網址(D) C:\Documents and Settings\Administrator\My Documents\另一項工作簡報檔\數位典藏國家型科技計畫 電子通訊.htm
移至 連結
電子通訊
簡介
訂閱、取消訂閱
各期通訊
隱私權聲明
版權著作聲明
稿約
通訊投稿
通訊員名單
2002/05/10 創刊號
◆本期目錄
發刊詞
發行人楊國樞教授有話要說
新聞
「92 、93年中程綱要計畫書」開始初審作業
編製國家數位典藏聯合目錄
「數位典藏國家型科技計畫」網站簡介
5月9日舉辦「業界說明會」
管考工作已經展開
數位化的工具書－《技術彙編》即將完成
歡迎新成員－公開徵求計畫結果揭曉
我們的圖徵故事
《國家資訊通信發展方案》資料供借閱
徵才－本計畫徵博士後人才
記事
91年01月08日 召開計畫辦公室第1-1次業務會報
91年01月15日 召開計畫辦公室第1-
http://www.ndap.org.tw/NewsLetter/content.html?subuid=115&uid=7
網際網路

加入Tag的新聞文件

001001 - 記事本

檔案(F) 編輯(E) 格式(O) 說明(H)

```
<?xml version="1.0" encoding="Big5"?>
<news>
2002/05/10 創刊號
<title>「92 、93年中程綱要計畫書」開始初審作業</title>
                                                    計畫辦公室秘書組/顧秋芬
<summary>
本國家型計畫「計畫辦公室」對所有參與計畫進行「92、93年中程綱要計畫書」的初審作業

--------------------------------------------------------------------------------
全文
    <who>本國家型計畫「計畫辦公室」</who>
對所有參與計畫
<what>進行「92、93年中程綱要計畫書」的初審作業</what>
，
<when>5月20日前</when>
需完成。
<how>請各計畫於5月15日前提供對審查意見之書面答辯說明書卅份予計畫辦公室秘書組，以

    本次審查的重點在於：計畫內容是否符合原規劃項目、費用，以及技術可行性。同時需

    <why>此項初審作業係受國科會人文處委託辦理</why>
，將分兩階段進行：4月25日至5月10日爲書面審查，5月18日爲會議審查（請參見議程）。巨

<table>時間
議程內容
8:30—9:00
報 到
9:00—12:30
機 構 計 畫 部 分

第
一
```

檢索新聞呈現

閱讀文章 - Microsoft Internet Explorer

Next

2002/05/10 創刊號

「92 、93年中程綱要計畫書」開始初審作業

計畫辦公室秘書組/顧秋芬

本國家型計畫「計畫辦公室」對所有參與計畫進行「92、93年中程綱要計畫書」的初審作業，5月20日前需完成。請各計畫於5月15日前提供對審查意見之書面答辯說明卅份予計畫辦公室秘書組，以利5月18日會議審查的順利召開

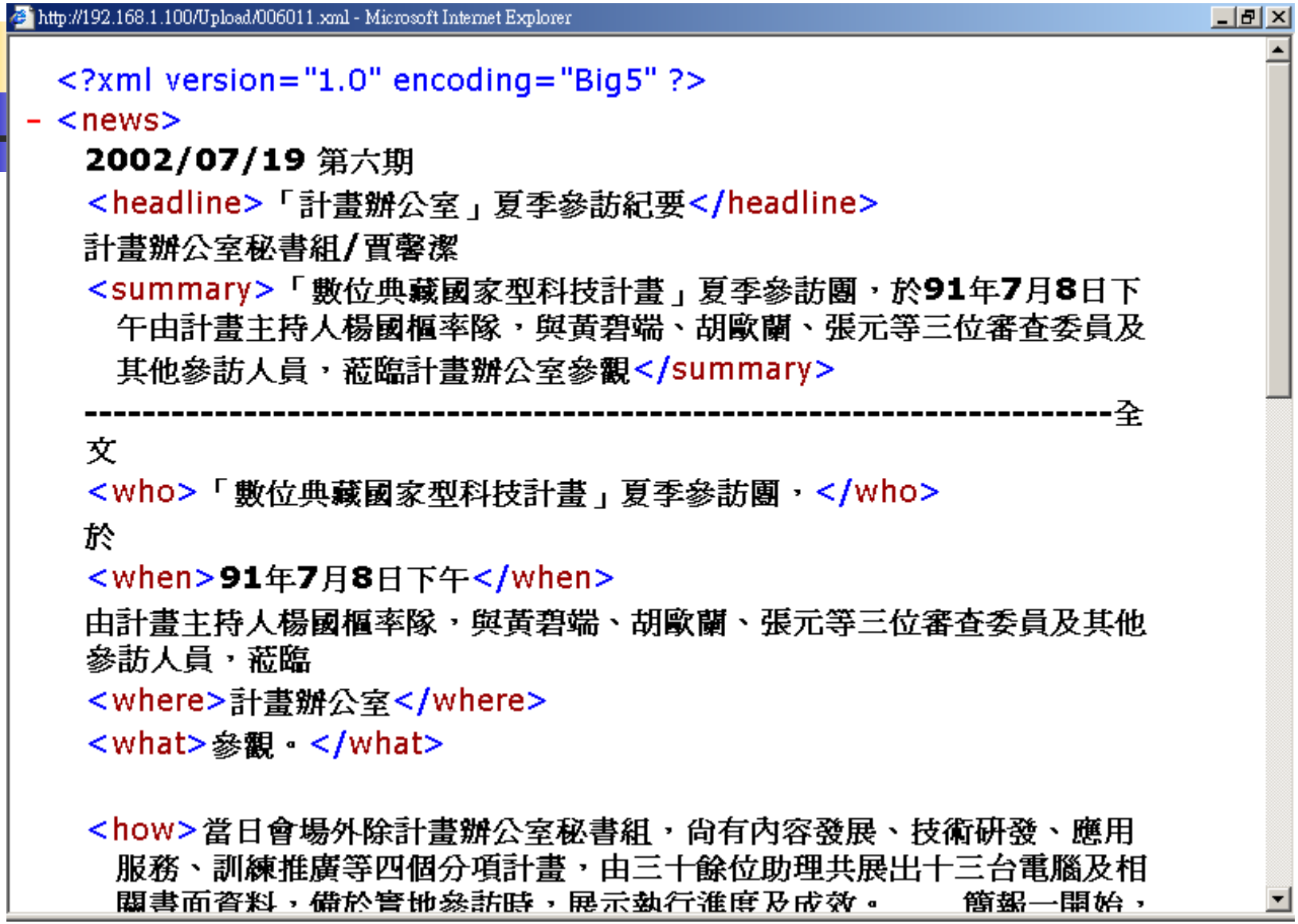
全文

本國家型計畫「計畫辦公室」
對所有參與計畫
進行「92、93年中程綱要計畫書」的初審作業
，
5月20日前
需完成。
請各計畫於5月15日前提供對審查意見之書面答辯說明書卅份予計畫辦公室秘書組，以利5月18日會議審查的順利召開。

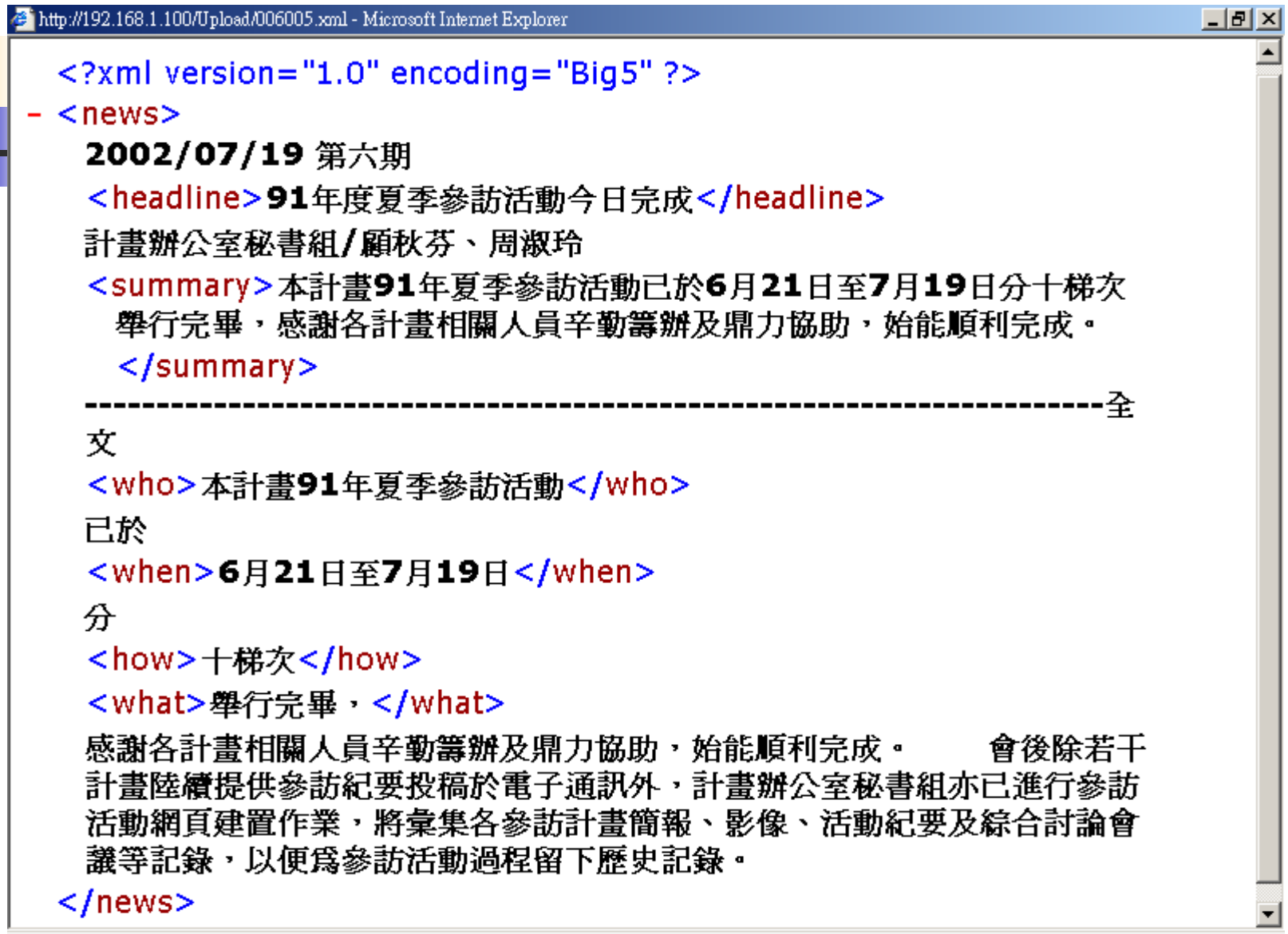
本次審查的重點在於：計畫內容是否符合原規劃項目、費用，以及技術可行性。同時需將整體計畫最後的15%概算做優先排序。待計畫辦公室完成全部初審作業，將由國科會進行複審作業。

此項初審作業係受國科會人文處委託辦理
，將分兩階段進行：4月25日至5月10日爲書面審查，5月18日爲會議審查（請參見議程）。目前計畫

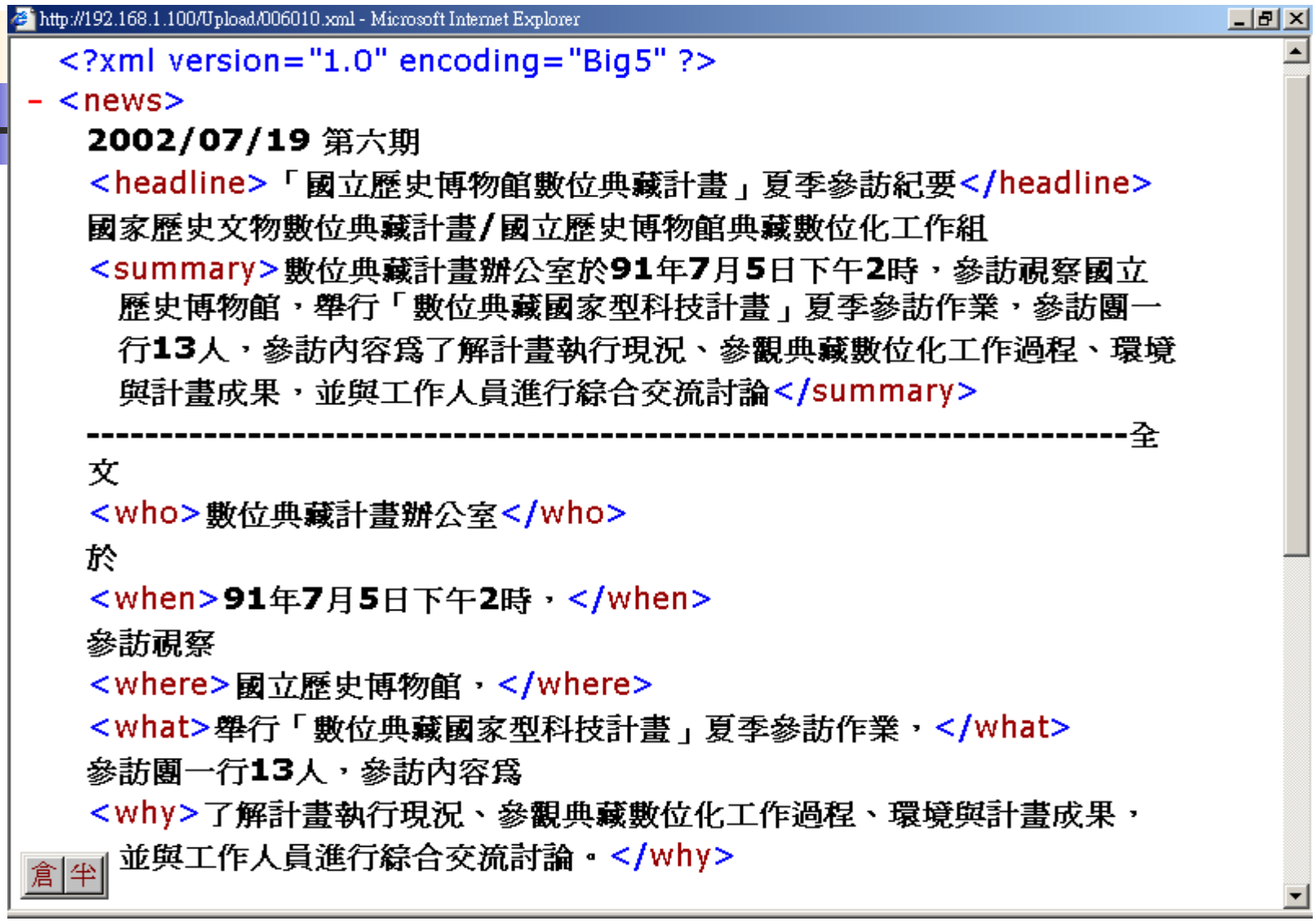
瀏覽Markup後的原始文章(XML)



瀏覽Markup後的原始文章(XML)



瀏覽Markup後的原始文章(XML)



CONCLUSION

- This is only a small, simple project, trying to initiate researches on Chinese news content markup including newspapers, televisions and so on.
- To help readers easily access news database, to help educators, students, reporters, editors in news writing are the future goal of this study.