Cross-Referencing for Bilingual Electronic Legal Documents

The Experience of HKLII

Dr Kevin K H Pun
Associate Professor of Computer Science & Law
HKLII team
The University of Hong Kong
HONG KONG
Cross-referencing of Legal Information in HKLII

Statutory Law – provided by Department of Justice, HKSAR Government

UK Acts – kept on BAILII

Cross-reference links provided by HKLII

Judicial decisions - provided by Judiciary, HKSAR Government

Foreign Judicial decisions - stored in WorldLII
Cross-Referencing for Bilingual documents in HKLII

- Main problem:
  - Correct identification of references
  - Insert at the proper places within the document

- Cross-referencing:
  - Between statutes
  - Between judicial decisions and statutes
  - Between judicial decisions
Creating references to statutes

- Hong Kong statutes (English)
  - Identified by:
    - `<<statute name>>` (e.g. “Trade Marks Ordinance”)
      - or
    - `<<Chapter name>>` (e.g. “Chapter 559”)
      - or
    - `<<statute name>>` + `<<Chapter name>>` (e.g. “Trade Marks Ordinance (Cap. 559)”)

```
Creating references to statutes

- Hong Kong statutes (English)
  - URL constructed by:
    - where
    - xxx = Chapter number
    - nn = section number
Creating references to statutes

- Problem – Incorrect URLs:
  - a) Difficulty in differentiating current from past versions of a statute
  - b) Change of chapter number (e.g. Copyright Ordinance changed from Cap. 39 to Cap. 528 in 1997)
  - c) Change of statute name (e.g. “Multi-Storey Buildings Ordinance (Cap. 344)” re-titled as “Building Management Ordinance (Cap. 344)” in 1993)

- Solution adopted:
  - a) use the latest version of the statute
  - b) same as above
  - c) not a pressing issue, as the number of such statutes still small
Creating references to statutes

- Hong Kong statutes (Chinese)
  - Problem - Segmentation
Segmentation problem in Chinese

- No word delimiters in Chinese, the computer system will have to “find a way” to segment the document into individual words that reflect the correct meaning of the document

- Two approaches to the problem:
  - Dictionary-based approach
  - Statistics-based approach
Segmentation problem in Chinese

Chinese language does not use space or any other delimiter as word boundaries

Eg

我們要發展中國家用電器
We Want To develop China’s Home electrical appliances

我們要發展中國家用電器
We Want Developing country To use Electrical appliances
Segmentation problem in Chinese

Dictionary-based approach

- Use a dictionary
  - Use “Longest Match” to identify all words in the sentence, always preferring the longest matching word
  - Relies on comprehensiveness of the dictionary
  - Longest matching does not always guarantee correct segmentation

Statistics-based approach

- Use a built-in table
  - Compute the score for potential segmentation of a sentence. Select the segmentation with highest score.
  - Relies on the table generated from training data
  - Performance depends on accuracy and scope of training
Creating references to statutes

- Hong Kong statutes (Chinese)
  - Identified by:
    - Angle brackets
      - and
    - Chinese keyword “條例” (“Ordinance”)
  - E.g. << 商標條例 >> (<< Trade Marks Ordinance >>)
Creating references to statutes

- Hong Kong statutes (Chinese)
  - URL constructed by:
    - Table look up of the string enclosed within each pair of brackets with statute name stored in the table

- Problem:
  - Time required for the lookup

- Solution:
  - Making use of a hash table
Objective:
There are total 950 HTML files and a list of 780 ordinances. For each HTML file, we want to find out all the occurrences of the listed ordinances.
Approach 1
(Regular expression with Perl)

- Using **Perl regular expression** to search for ordinances in a html file.
Approach 2
(Regular expression with Java)

- Using **Java regular expression** to search for ordinances in a html file.
Using a hash table to search for ordinances in a html file.
Example of Using **Regular Expression** (Approach 1 & Approach 2)

- Traverse the list of 780 ordinances one by one.
- Use regular expression to search for a particular ordinance in a HTML file.
Example of using **Hash Table**
(Approach 3)

Step 1: Build a hash table (key = value)
- For building a hash table of ordinances, the two characters in front of the phrase '條例' are used to form the key.
- E.g 1: 裁判官條例, **Key** = 判官,
- E.g 2: 刑事罪行條例, **Key** = 罪行
- **Hash Table:**

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>判官</td>
<td>裁判官條例</td>
</tr>
<tr>
<td></td>
<td>死因裁判官條例</td>
</tr>
<tr>
<td>罪行</td>
<td>刑事罪行條例</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>
Example of using **Hash Table** (Approach 3)

**Step 2: Search for the key** *(e.g: 判官)*

For each HTML file:
- Search for ‘條例’
- Search for the key *(e.g: 判官)*
- Use the key “判官” to retrieve all the values from the hash table

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>判官</td>
<td>裁判官條例</td>
</tr>
<tr>
<td></td>
<td>死因裁判官條例</td>
</tr>
<tr>
<td>罪行</td>
<td>有組織及嚴重罪行條例</td>
</tr>
<tr>
<td></td>
<td>刑事罪行條例</td>
</tr>
<tr>
<td></td>
<td>殺人罪行條例</td>
</tr>
<tr>
<td></td>
<td>複雜商業罪行條例</td>
</tr>
</tbody>
</table>

- **Result:** 裁判官條例 and 死因裁判官條例
Example of using **Hash Table** (Approach 3)

**Step 3: Select the longest one**

Eg: Both ‘死因裁判官條例’ and ‘裁判官條例’ are found in the hash table, ’死因裁判官條例’ will be selected as its length is the longest.

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>判官</td>
<td>裁判官條例</td>
</tr>
<tr>
<td>死因裁判官條例</td>
<td></td>
</tr>
</tbody>
</table>
### Result

<table>
<thead>
<tr>
<th>Approach</th>
<th>Time used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Regular expression with Perl)</td>
<td>~ 3 minutes</td>
</tr>
<tr>
<td>2 (Regular expression with Java)</td>
<td>~ 2.5 minutes</td>
</tr>
<tr>
<td>3 (Hash Table with Java)</td>
<td>~ 3 seconds</td>
</tr>
</tbody>
</table>
Creating references to judicial decisions

Problems:
- Difficulty in constructing the correct URLs
  - Different format between judicial decision title and its filename in HKLII
  - E.g.
    - Title format is `<plaintiff’s name> + <defendant’s name> + <citation>`
    - File name is `<case number>`
  - Therefore, translation from title to file name not straightforward
- More than one citation (e.g. in different law reports) for a single judicial decision
Creating references to judicial decisions

- Current available law reports include:
  - The Hong Kong Court of Final Appeal Reports (HKCFAR)
  - Hong Kong Law Reports & Digest (HKLRD)
  - Hong Kong Cases (HKC)
  - Hong Kong Unreported Judicial decisions (HKCU)
  - Hong Kong Electronic Cases (HKEC)
Creating references to judicial decisions

Solution:

- Use of a parallel citation table – constructed by the Lui Che Woo Law Library of the University of Hong Kong

  - Provides all citations of each judicial decision

  - any citation of a judicial decision is mapped onto the URL of the decision via a primary key stored in the parallel citation table
Creating references to judicial decisions

<table>
<thead>
<tr>
<th></th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year</strong></td>
<td>1997</td>
</tr>
<tr>
<td><strong>Case name</strong></td>
<td>CHAN SZE TING and Another v. HONG KONG SPECIAL ADMINISTRATIVE REGION</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>FAMC000004/1997</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>18 December 1997</td>
</tr>
<tr>
<td><strong>HKLII citation</strong></td>
<td>[1997] HKCFA 7</td>
</tr>
<tr>
<td><strong>HKCFAR</strong></td>
<td>[1997-98] 1 HKCFAR 46</td>
</tr>
<tr>
<td><strong>HKLRD</strong></td>
<td>[1998] 1 HKLRD 45</td>
</tr>
<tr>
<td><strong>HKC</strong></td>
<td></td>
</tr>
<tr>
<td><strong>HKCU</strong></td>
<td>[1997] 978 HKCU 1; [1997] 388 HKCU 1</td>
</tr>
</tbody>
</table>
Way forward

“Noteup”
- A pre-defined query embedded in the web page which searches for all documents referring to the statute

Available in English version only based on Sino search engine

Planned to develop for Chinese version using the mnoGoSearch search engine
Way Forward

- More accurate identification of keywords
- Improved time for constructing URLs for Chinese statutes via the use of a hash table
- Automatic generation of Parallel Citation Table