

## 第二章 相關研究

本章主要是相關文獻的整理與探討。第一節介紹語音合成技術，此技術可將文字檔案轉換成多媒體聲音檔案(Wav)，是本研究中語音元件所必備的技術之一。第二節介紹語音壓縮技術，此技術可將多媒體聲音檔案(Wav)壓縮成 MP3 音樂格式，是本研究中語音元件所必備的技術之二。第三節介紹知識樹概念，此概念用在本研究以整合不同網站的網頁資源，提供使用者一個更豐富、更有組織的領域知識庫。最後，第四節介紹無障礙網頁的可及性規範，本研究所提出的雙模式知識網站將依循此規範以符合 Universal Web Access 的設計概念。

### 2.1 語音合成技術

語音合成 (Speech Synthesizer) 又名文句轉語音(Text-To-Speech, TTS)，是指將輸入的文字或文件，轉換成清晰、流暢、自然的語音輸出技術。語音合成技術大致可分成四個模組【23】，包括文句分析、韻律產生器、合成單元產生器、語音合成等，分述如下：

1. 文句分析：主要在分析文句的語法與語意，並將分析結果轉換成語言特徵參數，讓電腦知道本文中哪些是詞，哪些是句子，發什麼音，怎麼發音，發音時到哪應該停頓，停頓多長等等。
2. 韻律產生器：將語言特徵參數送入韻律產生器，產生文句的音節對應韻律訊息，包括有基頻軌跡、音量及音長等，可將說話的聲調、語氣、停頓方式及發音長短轉換成韻律參數。
3. 合成單元產生器：根據語音資料庫中的單音節音素語音波形樣本輸出合成單元。
4. 語音合成：根據需要發的音從聲音資料庫中選擇出合適的聲學參數，然後根據在韻律模型中得到的韻律參數，透過語音合成演算法產生語音。

而一個語音合成器技術的好壞通常取決於以下三點，包括：

1. 發音的自然度。

2. 破音字的處理。
3. 即時處理的能力。

目前的語音合成技術發展日趨成熟【1】【9】【19】，許多語音合成器的產品也已經大量地商品化。因此，藉由運用已成熟的語音合成技術，對於我們軟體系統的開發，可以大大減少時間的消耗，而能專注於語音解說模式的研究。

本研究所採用的語音合成技術為 IQTTS 文字轉語音引擎，由網際智慧公司 IQ China 公司所開發，其特點為能將中英文文字資訊，轉換為聲音輸出。並且能對中文破音字、阿拉伯字、前後文語意及特殊念法作分析判斷，朗讀最可能的讀音，目前該技術廣泛應用於個人電腦系統、電話系統、行動裝置、宣告系統及家電玩具等方面。

### 2.1.1 IQTTS系統架構

IQTTS的語音合成技術架構，如圖2-1所示，主要包含有語意處理模組、語音處理模組、聲音庫三大部分，搭配應用系統程式使用。

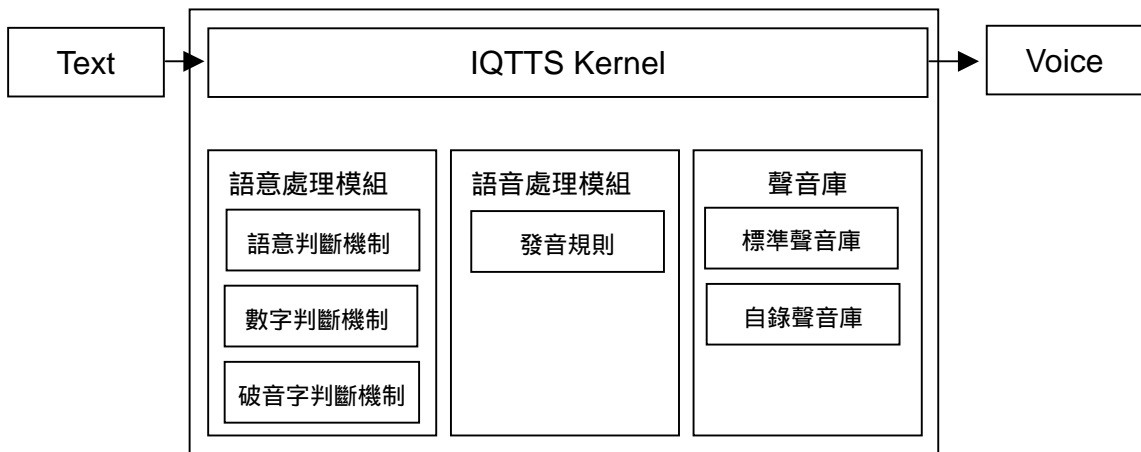


圖 2-1 IQTTS 元件圖【10】

文字檔案輸入後，語意處理模組將文章解譯成文句，先經由語意判斷機制、數字判斷機制、破音字判斷機制對文章語句的語意、數字、破音字作處理，再經由語音處理模組套用個別的發音規則，最後利用聲音庫合成語音後輸出，其聲音庫並允許自錄語音。

## 2.1.2 IQTTS 功能規格

以下將介紹 IQTTS 的各種功能規格，包括發音範圍、語音輸出、語音介面、作業系統支援以及聲音庫等，分述如下。

### 1. 發音範圍

- 中文字、數字（全型、半型）、英文字母（全型、半形、大寫、小寫）
- 英文以單一字母為單位發音，可以中英文夾雜方式發音。

### 2. 語音輸出

- 支援.wav 格式檔案輸出，格式為 11KHz，8bits，PCM，Mono。
- 支援音效卡輸出。

### 3. 語音介面

- 提供介面可以調整發音的頻率、音長、間隔、音量、回音、聲音種類、停止發音等功能。

### 4. 作業系統支援

- Windows 98、Windows Me、Windows NT、Windows 2000、Windows XP。

### 5. 聲音庫

- 標準國語成年男子聲音，單音長度 0.3~0.45 秒間，平均 0.4 秒。
- 標準國語成年女子聲音，單音長度 0.3~0.45 秒間，平均 0.4 秒。

## 2.1.3 IQ-TTS API

IQTTS 文字轉語音引擎為.DLL 程式元件，可以利用 Visual C/C++、Visual Basic、Borland C/C++、PowerBuilder 等各種程式語言呼叫。以下為 IQ-TTS API：

1. void setPath(LPSTR path) – 聲音庫路徑設定
2. int TextOutsc (LPSTR szTxt) – 聲音輸出到Speaker

3. int TextOutWave(LPSTR szTxt, LPSTR szWav) – 聲音輸出到.wav
4. void ShowLastError(int errCode) – 發音錯誤訊息顯示
5. void Break (void) – 發音強制中斷
6. BOOL IsStop (void) – 發音結束否？
7. void setParam(int rest, int freq, int echo) – 發音參數設定
8. void setVolume(int level) – 音量設定
9. void setWavLen(int percent) – 單音長度設定
10. void setPronceNumber2 (int method) – 數字“2”的發音 (僅wav版提供)
11. void setVoiceType(char type) – 聲音種類
12. int text2ph(LPSTR szTxt, LPSTR szPh) – 中文文字轉注音(不處理英文、符號，而且僅wav版提供)
13. void SetUseSAPI(int UseSapi) – 搭配SAPI 5.1朗讀英文

## 2.2 語音壓縮技術

Wave 檔的產生方式，是把音波切割為無數個小段，再把各小段的聲波儲存為小型的聲音數位樣本，這種過程，即稱為取樣。當聲音錄製成 Wave 檔時，每秒鐘取樣的數量越多，電腦所能呈現的波型就越精確越完整；同時，播放時也就越能達到原音重現的水準。每秒鐘抽選的樣本數量，即為所謂的「取樣率」(Sampling rate)。如果每秒鐘抽選十次樣本，在播放該段聲波時，是無法辨別出原來的聲音的，你只能聽到一連串的調音。若希望能呈現較為精確的放音效果，音效卡每秒鐘必須抽選數千個樣本才能如願，但是所佔的檔案空間也變大。

用來存放聲音樣本的位元數量，是另外一個影響音質的要素。音效卡測量一段音波的「振幅」(強度)時，測定過程的精準程度，完全仰賴可用數值的多寡而定。4 位元的音效卡使用 4 個位元來存放一個樣本，而 4 位元只能提供 16 個不

同的數值，所以數值量明顯不足。因此，原始音波的細節部分可能會無法保存，因為這些數值僅足以表現音波大略的形狀。

目前,高品質的聲音以 CD 音質為主。一般 CD 音質的取樣頻率為 48kHz 或 44.1 kHz，每一個取樣點經由 16 位元的脈衝碼調變 (PCM, Pulse Code Modulation) 而成。其所需的數據量相當的龐大，每一單聲道聲音在取樣頻率為 48 kHz 時每秒可達到 768 kbps (Kilobits per second)；在取樣頻率為 44.1 kHz 時則為 705.6 kbps；若是雙聲道則分別為 1.54 Mbps (Megabits per second) 及 1.41 Mbps【18】。

雖然 CD 音質為一般人所認可的高品質聲音，但在網路風行的數位資訊時代中，卻因為其高容量或高傳輸率需求使其實用性大打折扣，無論在網路傳輸或資料儲存等方面，均造成使用者的負擔。數位音訊壓縮便是為了要使寬頻聲音訊號（高取樣頻率）經由壓縮後，能夠保持高品質的聲音（近 CD 音質）以及較小容量的儲存空間，以方便人們使用、保存和傳輸。其主要目的在於將聲音訊號處理後以最少的位元 (Bit) 數來儲存，而訊號重建時希望能達到最小聽覺差異。簡言之，音樂壓縮希望達到即使受過專業訓練的試聽員也聽不出原先聲音以及壓縮處理過後的聲音差別【3】。

### 2.2.1 壓縮比(Compression ratios)、位元率(bitrate)與音訊品質

通常，音訊檔案經壓縮再解壓後，並不等於原來的檔案，因原始檔案在壓縮過程中已經遭到破壞，雖然如此，他們聽起來通常是相同的，而其品質決定於音訊壓縮的方式。

一般而言，較低的壓縮比 Compression ratios 會有較好的音訊品質，反之，較高的壓縮比其音訊品質將較差，另外一個常用來評量音訊品質的項目為 bitrate，bitrate 表示壓縮音訊資料每秒所佔用的平均位元數，通常以 kbps 為單位， $\text{kbits}/\text{s}$ ，或  $1000 \text{ bits}/\text{s}$ ，較高的 bitrate 會有較好的音訊品質，反之，則表示音訊品質較差，表 2-1 為常用的 bitrate 與音訊品質對應表。

Bitrate	Bandwidth	Quality comparable to or better than
16 kbps	4.5 kHz	shortwave radio

32 kbps	7.5 kHz	AM radio
96 kbps	11 kHz	FM radio
128 kbps	16 kHz	near CD
160-180 kbps (variable bitrate)	20 kHz	perceptual transparency
256 kbps	22 kHz	Studio

表 2-1 Bitrate 與音效品質對照表【20】

## 2.2.2 MPEG - 1 音頻壓縮標準

MPEG - 1 音頻壓縮標準是第一個可高度保留音訊品質的資料壓縮標準【3】。MPEG - 1 音頻標準提供三個獨立的壓縮層次，讓用戶可在複雜性和壓縮品質之間權衡選擇。

1. 層次 1 (Layer 1)最為簡單，使用的 bitrate 384kb / s，主要用於數位小卡座 DCC(Digital Compact Cassette)。
2. 層次 2 (Layer 2)的複雜度屬於中等，使用的 bitrate 192kb / s 左右。其應用包括：數位廣播(Digital Audio Broadcasting)的音頻編碼，CD - ROM 上的音頻信號以及互動式 CD - I(CD - interactive)。
3. 層次 3 (Layer 3)最為複雜，但音質最佳，使用的 bitrate 為 64kb / s，尤其適用於綜合業務資料網(ISDN)上的音頻傳輸。

MP3 是最近幾年來非常受到大家注意的一種音效檔案格式，具有「檔案小」、「音訊佳」的特點，其原理是利用 MPEG - 1 Layer 3 的技術將 WAV 聲音檔案以 1:10 甚至 1:12 的壓縮率壓縮而來，以 44KHz 加上 16BIT 立體聲播放出來的音色將可媲美 CD 音樂音質。

### 2.2.3 LAME Encoder

LAME (LAME Ain't an Mp3 Encoder)是免費的MP3編碼引擎【20】，可將音訊檔案如Wav檔案以MPEG - 1 Layer 3的技術壓縮成為MP3音樂檔案。LAME編碼出來的MP3音色純厚、空間寬廣、低音清晰、細節表現良好，音質幾乎可以媲美CD音頻，但文件體積卻非常小。

Lame支援了以下三種編碼模式：Constant bitrate(CBR)、Variable bitrate(VBR)和Average bitrate(ABR)

1. Constant bitrate：固定的bitrate，指文件從頭到尾都是同一種bitrate，這也是最常用的壓縮方式。
2. Variable bitrate：可變的bitrate，也就是沒有固定的bitrate。在這樣的設定下，壓縮軟體在壓縮時根據音頻數據即時決定使用什麼bitrate。這是音質與文件大小都兼顧的方式，推薦選用。
3. Average bitrate：平均的bitrate，是VBR的一種插值參數演算法。它在壓縮低頻信號時使用較低的bitrate，而在壓縮聽覺敏感的高頻時使用較高bitrate。

IQTTS與Lame構成本研究的語音元件，文字檔案經由IQTTS轉換成為Wav音訊檔案後，再交由Lame將Wav檔案壓縮成較小的MP3檔案，以利網路傳輸。

## 2.3 Ontology

Ontology是指某領域下之共有的認知，目的是為了以明確且正規的定義來表達共同概念化之事物，並且定義出的概念必須要能清楚的傳達在人與電腦之間【14】【23】。「概念化」是有關真實世界某事件概念的抽象模型，「共同」的涵義為Ontology所傳達的認知為眾人所認可，並非個人化觀點，「明確」是指各類概念的使用與限制皆能被清楚地定義，「正規」是要確保Ontology格式為電腦可解讀。

為了讓來自不同來源的資訊能夠整合，Ontology提供共通的字彙來支援知識的分享和重用【7】。Hendler (2001)明確地指出Ontology的意涵，「Ontology

為知識名詞的集合，包含詞彙（vocabulary），語意關聯（semantic interconnections），和一些簡單的推論法則（rules of inference）和某議題的邏輯（logic for some particular topic）。」

因此，Ontology所表達出的知識具備有「物件導向」、「特殊化/普遍化」與「可推論性」等三項特性【6】，分述如下：

1. 物件導向式(Object-oriented)：這是『Ontology』的『概念』中最常具備的屬性，尤其是應用在資訊處理的領域中。
2. 特殊化/一般化(Generalization/Specialization)：這是當『Ontology』被建立出來的時候，知識表達要表現出來的屬性。所謂的特殊化與一般化指的是『Ontology』領域中的明確性與模糊性。特殊化的意思是更為明確，更為細分的領域。相反的，一般化指的是更為模糊，更為放大的領域。
3. 可推演性(Reasoning)：可以在『Ontology』的知識結構中，靠著推演的方式取得更多的有用資訊。

藉由Ontology描述的知識結構，代表某個領域具有一般且可以共同使用的知識，讓人們與電腦應用程式系統之間可以作溝通【8】，而不需要顧慮到異質電腦環境所帶來的困擾。同時，字彙的一致性能夠提昇資料搜尋的準確性；另一方面，使用者將可使用階層（和屬性限制）透過類別和子類別屬性或其他語意關聯來找到相關的資訊。

### 2.3.1 特殊知識領域(Domain Ontology)

Ontology 可以分為非常多種不一樣的形式，那麼一個特定領域的Ontology是如何建構出來的？對於定義該領域的專家來說，必須要作以下的分析【15】：

1. 檢查作為描述實體的字彙
2. 對於字彙裡面的名詞，建立起正規的敘述（將它們轉化成概念、關係）
3. 描繪出名詞間在概念上的關係為何

我們可以圖 2-2 來表示一個領域所具有的知識架構：



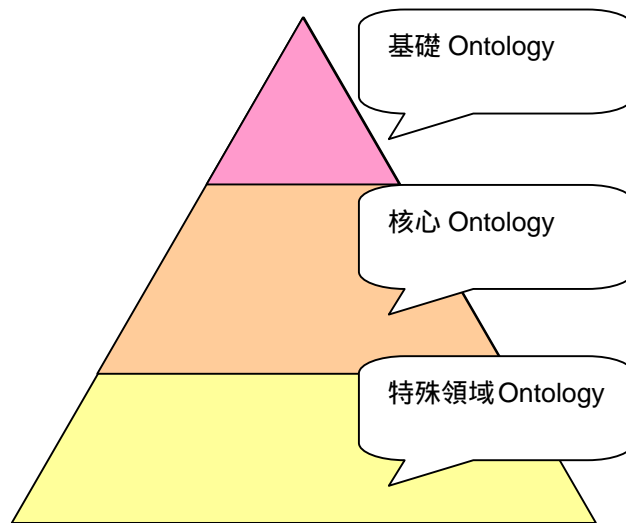


圖 2-2 Ontology 階層分類圖

在這三層之中，基礎Ontology 是由哲學家與人工智慧專家所定義的，而這個Ontology 主要陳述的觀念，與哪一個特定領域是沒有關係的，而它的作用最主要是能夠在不同的領域中，可以確保有再利用性的機能。至於核心Ontology，則是由知識工程師依據該領域的知識所架構出來的，所以通常這裡面會有數百個知識觀念。至於特殊領域的Ontology，則是依據不同開發專案的需求而定義的特別觀念、關係，而這樣所必須付出的代價就是知識觀念有可能與他人不一致，並且在使用上會有所限制。

### 2.3.2 Ontology的應用

由於Ontology可用來描述領域內部的知識結構，所以在使用上大多擷取其觀念當作知識庫來使用，但Ontology的應用不止於此，且其應用領域有愈來愈廣泛的趨勢，表2.2說明Ontology的應用範圍：

幫助溝通Communication
在實作的運算系統間
在人與人間
在人和實作的運算間

幫助運算推論 Computational inference
內部表示和操作規劃的資訊
使用理論和概念的辭彙來分析實作系統的內部架構、演算法、輸入和輸出
幫助知識的重用和組織 Reuse and Organization
組織或結構化領域的辭彙和知識

表2.2 Ontology的應用範圍【5】

## 2.4 無障礙網頁設計規範

全球資訊網上的資訊包羅萬象，世界各類網站內容日新月異，已經有許多民眾在生活中和學習上，皆以全球資訊網為主要資訊獲取來源。但是這些網頁內容在建置之時並無考慮到身心障礙者在使用這些資訊時，所可能碰到的各種問題，例如視覺障礙者因為眼睛看不到，在上網之時碰到網頁內的圖片的時候，即使使用點字輸出機或者先進的語音系統，仍然無法理解圖片所表示的訊息。

近年來國外有些組織開始注意到這個問題，提倡 Universal Web Access 的概念【21】，希望不論一般人或身心障礙者，都能夠享受全球資訊網的服務。其主要概念是透過一些網頁的設計規範和法規條例，希望身心障礙者在瀏覽各種全球資訊網的網頁資訊時沒有任何障礙。

### 2.4.1 無障礙網頁內容可及性規範

目前主導國際全球資訊網技術的標準機構「全球資訊網協會」(W3C)所形成的「資訊網可及性推動組織」(Web Accessibility Initiative, 簡稱WAI)【22】已經在1999年五月五日訂定出一份「無障礙網頁內容可及性規範」(Web Content Accessibility Guidelines; 簡稱WCAG)的國際標準，其中明確建議網頁在使用HTML語言時，應該遵守那些可及性原則，其根據網頁內各種不同的特性，共分為十四條規範(Guideline)，六十五項檢測點(Checkpoint)。這六十五項檢測點再

根據達到可及性要求的程度，分別隸屬於三種優先等級(Priority)。屬於第一優先權(Priority One)的項目，為滿足可及性的基本要求，是所有的網站和網頁都應該要達到的最低標準。屬於第二優先權(Priority Two)的項目，為滿足可及性要求的建議項目，是建議所有的網站和網頁最好能夠達到的標準。而屬於第三優先權(Priority Three)的項目，則是滿足可及性要求的最高目標。能夠完美的通過第三優先權中各項檢驗標準的網站和網頁，是真正的無障礙網頁。自從WCAG標準公佈之後，經過各國身心障礙團體和專家學者的推動，各國政府(包括美國、英國、澳洲、加拿大、歐盟各國、日本、韓國等)也逐漸重視全球資訊網在無障礙網路空間的推動。

#### 2.4.2 無障礙網頁開發規範

在國內，研考會於九十一年六月「無障礙網頁開發標準暨標章核發作業」委外案，參考 W3C 協會的 WAI 組織的無障礙網頁內容標準相關規範，並參照各國在制訂無障礙網頁相關政策和推廣策略，與國內近年來在身心障礙者保護政策等相關措施，訂定我國的「無障礙網頁開發規範」。規範內容包括無障礙網頁可及性設計四項原則、三個優先等級、六個檢測等級、十四條規範、和九十個標準檢測碼。

無障礙網頁可及性設計四項原則是告訴網站設計者在規劃一個公眾網站時，應該考慮四個重要層面的可及性因素，包括內容(content)、結構(structure)、技術(technique)、和瀏覽(navigation)等。在內容因素上，應該考慮網頁文字資訊和多媒體資訊內容的可及性設計；在結構因素上，應該考慮網頁內版面規劃和內容結構上的可及性設計；在技術因素上，應該考慮處理網頁內容、文件語言技術、程式語言技術、媒體技術、和輸出入裝置技術等的可及性設計；在瀏覽因素上，應該考慮網站內各網頁間瀏覽結構的可及性設計。

而無障礙網頁規範的三個優先等級定義如下所示：

1. 第一優先權：網頁內容開發者在開發網頁時必須滿足此優先權的檢測碼。否

則，某些使用者或團體將發現無法使用此文件資訊。滿足此優先權的檢測碼對使用網頁文件的某些使用者或團體來說，是一種基本的需求。

2. 第二優先權：網頁內容開發者在開發網頁時應該滿足此優先權的檢測碼。否則，某些使用者或團體將發現難以使用此文件資訊。滿足此優先權的檢測碼，將可減少網頁文件使用上的障礙。
3. 第三優先權：網頁內容開發者在開發網頁時可以納入此優先權的檢測碼要求。否則，某些使用者或團體將發現有可能難以使用此文件資訊。滿足此優先權的檢測碼，將可改善網頁文件的可及性。

「無障礙網頁開發規範」在訂定之初就被賦予訂定國內官方無障礙網頁標準和發展網頁檢測工具等雙重任務，因此在設計上主要考慮三個因素，包括有優先等級、機器檢測還是人工檢測及標準檢測碼的狀態。其中優先等級的考量可以反映網頁設計的可及性程度；機器檢測或人工檢測的考量可以反映檢測者的身份；標準檢測碼的狀態的考量可以反映檢測處理方式的認定。在此機器檢測/人工檢測雙軌並行制度下衍生的認證等級也隨之形成六種認證等級：機器檢測(單星級、雙星級、參星級)、人工檢測(單星級、雙星級、參星級)。

#### 1. 機器檢測

- 單星級：通過第一優先權、檢測狀態為 0 的檢測碼
- 雙星級：通過第一、二優先權、檢測狀態為 0 的檢測碼
- 參星級：通過第一、二、三優先權、檢測狀態為 0 的檢測碼

#### 2. 人工檢測

- 單星級：通過第一優先權、檢測狀態為 1 和 2 的檢測碼
- 雙星級：通過第一、二優先權、檢測狀態為 1 和 2 的檢測碼
- 參星級：通過第一、二、三優先權、檢測狀態為 1 和 2 的檢測

此外，為了讓網頁開發者能夠對網頁開發在可及性設計的考慮上有明確的規範條文，研考會特參考 WAI 組織在相關無障礙網頁標準的設計，以十四條規範

來引導網頁開發者設計可以讓所有人士都可以使用的無障礙網頁。分述如下：

1. 對於聽覺及視覺的內容要提供相等的替代文字內容。
2. 不要單獨靠色彩來提供特殊資訊。
3. 適當地使用標記語言和樣式表單。
4. 闡明自然語言的使用。
5. 建立編排良好的表格。
6. 確保網頁能在新科技下良好地呈現。
7. 確保使用者能處理時間敏感內容的改變。
8. 確保嵌入式使用者介面具有直接可及性。
9. 設計裝置獨立網頁。
10. 使用過渡的解決方案。
11. 使用國際與國內官方訂定的技術和規範。
12. 提供內容導引資訊。
13. 提供清楚的瀏覽網站機制。
14. 確保簡單清楚的網頁內容。

最後，九十個標準檢測碼是「無障礙網頁開發規範」的核心，也是網頁在實施無障礙檢測時的主要依據，檢測碼編碼格式以 H101000 為例，其訊息為「圖片需要加上替代文字說明」，編碼中 H 代表 HTML 語言；接著 1 代表第一優先等級；接著 01 代表規範一；接著 0 代表檢測狀態 0 機器辨識/機器檢測；接著 00 為規範一內的流水號。檢測碼的檢測狀態共計有 0(機器辨識/機器檢測)、1(機器辨識/人工檢測)、2(人工辨識/人工檢測)三種狀態，狀態值反應了檢測方式和檢測碼的生命週期現象。