

## Gufonet 發佈版本通知

版本號碼：1.12

發佈日期：2023/06/30

更新內容：

### 工具更新

- 『Gufonet來源文件管理』
  - 修正對於未新增ODBC資料來源而造成UI損毀的問題。
  - 修正unhandledException之呈現方式。
  - Gufonet.config設定檔異動：新增crawlerMemory limit模式，用以限制爬蟲索引使用過多記憶體。
  - Gufonet.config設定檔異動：AttachedColumnErrorByPass（File Path in DB的檔案不存在時忽略該檔案）預設為True；也就是預設的情況下，會忽略檔案不存在的錯誤。
- 文件匯入程式IndexUpdate.exe
  - 修正部分Report Exception捕捉方式。
  - 修正匯入檔案檔名消失之問題。

### 核心更新

1. 修正主題查詢會有錯誤的問題。
2. 修正更新索引時會有錯誤的問題。
3. 修正檢索時，若欄位內容只取詞頻或分數，會造成Exception錯誤的問題。
4. 修正精確檢索、無動態摘要時，使用非關鍵詞的詞頻(取得或排序)，其值會都是0的問題。
5. 修正第二階排序取值錯誤的問題，會導致多台檢索機檢索結果合併時，排序順序錯誤。
6. 修正原形詞AND運算錯誤的問題。
7. 修正取\$Parent\$、\$TopParent\$欄位內容時，回傳錯誤的值。
8. 修正模糊查詢／主題查詢時，篩選條件會產生部份錯誤文件的問題。
9. 修正GetDocTerms回傳的詞未按詞長詞頻排序、錯誤詞頻、以及重複字串的問題。
10. 增加檢索回傳結果參數ExactlyTotalDocNoParam，以控制不回傳精確文件總數時的額外處理工作。
11. 增加API函數IsTotalDocNoExactly以檢測回傳文件總數是否精確。
12. 增加API函數GetDocImportantTerms以取得文件重要詞彙。
13. 關聯文件改採用GetDocImportantTerms API取得文件重要詞彙進行主題檢索。
14. 增加GetSameMeaningTerm取得指定詞彙的同義詞組API。
15. 增加AddSystemTerm增加系統詞庫API，以便在不動到現有詞庫前提下，額外加入詞彙。
16. 改良並減少文件下載Cache檔的佔用空間（採用NLZ4跨文件整體壓縮，原20.1GB下降至7.2GB，約省下60%空間）。
17. 儲存機背景進行文件內容最佳化壓縮，以減少索引機、檢索機的文件內容索引檔大小（與舊版不相容，需一起更換）。500萬份文件壓縮時間約8日，文件內容索引檔原19.2GB下降至14.8GB(-23%)，文件下載Cache檔原7.2GB下降至5.6GB(-22%)
18. 回收之可用已刪文件ID，由儲存機改到DocMain，避免已刪文件ID愈

- 放愈多（例如#1做暫存再移轉到他台的情況），同時加入遺失文件ID回收機制（與舊版不相容，需一起更換）。
19. 文件儲存機改採用批次同步，以大幅加快資料同步的速度（500萬份文件約需20分鐘即可完成同步）。
  20. 篩選條件And運算進行多範圍數值域合併最佳化。
  21. 改良模糊查詢/主題查詢/關聯文件的動態摘要取得方式，儘量集中於長詞出現處，避免過於零散。
  22. 增加和\$tf\$相同功用的\$Score\$欄位，在意義上為用來取得相似度分數、使用相似度排序等，以便和"詞頻"的意義做區分(精確查詢時，相似度分數就是詞頻)。
  23. 模糊查詢/主題查詢允許第二階排序（第一階固定為按分數由大到小排序）
  24. 檢索機增加檢索檔事先載入機制（SetSearcherPreloadOption API），將部份索引檔先行載入到記憶體中，以加快存取速度。
  25. 索引檔分檔大小由4GB調降至2GB，以方便事先載入。
  26. 實作檢索結果Cache機制（SetSearchCacheTime API），相同之檢索（包括關聯文件）不必重複運算，直接回傳。
  27. 文件中心API呼叫時，批次新增文件功能獨立成一個執行緒，以免較長的呼叫時間影響其他功能呼叫的運作。
  28. 容錯字標準嚴格化，必須是極形似或完全同音的常見字，不再加入音似字或形近字，避免出現奇怪的檢索結果，同時也可加快容錯檢索的速度。
  29. 調校檢索邏輯運算、模糊查詢、主題查詢、關聯文件、動態摘要、二階排序、詞頻排序、統計運算、分組運算、容錯查詢、回傳最上層文件的檢索效能。
  30. 調整模糊查詢、主題查詢、關聯文件的取詞方式，以加快運算速度。
  31. 文件主控的狀態字串中，加入顯示文件總數，以便容易判斷是否已快超出授權文件總數。
  32. 變更索引檔詞文件資訊結構（版本4），以加快檢索速度。
  33. 調校原形詞檢索、非關鍵詞檢索、鄰近檢索的檢索效能。

版本號碼：1.10

發佈日期：2023/03/31

更新內容：

#### 工具更新

1. 『Gufonet來源文件管理』
  - 修正設定檔匯入紀錄功能，在資料庫檔案不存在時，發生錯誤。
2. 文件匯入程式IndexUpdate.exe
  - 對於ODBC文件來源匯入過程中DocName欄位必須包含Database主鍵，此限制已解除。
  - 修正ODBC索引中的檔案路徑發生「檔案不存在」時，會直接整個索引失敗之處理邏輯：新增彈性選項(告知錯誤但仍可建立)，建立後可以在『Gufonet來源文件管理』的儀錶板查看警告文件數量。
  - 修正DLL或EXE被誤判為RAR檔時，可能會造成當機的問題。

#### 核心更新

1. 增加『鄰近檢索』功能：查詢詞中萬用字元\*後面可接括號()，裡面的

數字表示前後文中間限制的單字數(token)，也就是\*最多可填入的單字數，內定為80。

2. 檢索查詢詞增加可使用萬用字元\*和?，但不做為英數單字的萬用字元，英文只能使用\*做後切截查詢。
3. 增加支援7z壓縮檔之解析。
4. 修正在1分鐘內連續更新索引，若索引檔大小未變但內容有變時，Searcher不會異動該索引檔的問題。
5. 修正不取欄位內容，只取動態摘要會當機的問題。

版本號碼：1.06b

發佈日期：2022/12/30

更新內容：

#### 工具更新

1. 『Gufonet來源文件管理』
  - 改為使用 .Net Core 6 執行。
  - 增加支援「資料夾檔名索引」資料來源。
  - 新增：儀錶板增加顯示子文件數目。
  - 修正：AddMode=false時子文件未正確處理之問題。
  - 修正：ODBC來源匯入，在設定DocName欄位為非資料庫主鍵欄位時，匯入失敗的問題。
2. CsvSetIndexField
  - 配合新增之『禁止取值』的索引欄位設定，修正匯入匯出之檔案內容。

#### 核心更新

1. 修正：字首提示詞在最後一個字為英文時，會回傳重覆結果的問題。
2. 修正：增加各伺服器的狀態訊息，例如文件儲存機會顯示剩餘待同步文件數，方便了解同步進度。
3. 修正索引機取得詞庫等資訊時，均以IndexMain裡的資料為準，而不以最新的為準，以便和IndexMain的資料同步。
4. 修正：輸入同義詞庫時，自動濾除輸入字串前後的空白字元。
5. 增加內定同義詞庫。
6. 增加禁止取值的欄位索引定義。
7. 內文欄位內容不再限制讀取。
8. 允許設定內定欄位的索引定義。

#### WebUI

1. 新增：序號資訊更新頁面的「序號註冊網站」文字，加入超連結。
2. 要掛載在IIS上的模組，分開在不同目錄安裝，以解決衝突問題。

#### Web API

1. 要掛載在IIS上的模組，分開在不同目錄安裝，以解決衝突問題。

版本號碼：1.05b

發佈日期：2022/09/30

更新內容：

#### 工具更新

- 『Gufonet索引設定暨工具配置圖形介面』擴充為『Gufonet來源文件管理』：
  - 新增來源文件管理『儀表板』功能。
  - 新增開啟WebUI/WebAPI頁面的連動功能。
  - 新增『設定檔匯入』、『設定檔匯出』，以及『檢視紀錄』等功能。
  - 設定檔儲存機制變更：原由使用者點擊『儲存鈕』觸發儲存，修改為『確定變更即儲存』。
  - 新增『執行匯入』按鈕及其對應功能。
  - 修正『複製執行命令到剪貼簿』產生之執行字串的跳脫字元。
  - ODBC附檔欄位UI設定，允許選擇網路芳鄰磁碟目錄。
  - 對於『資料匣檔案』的來源類型，新增『檔名建索引』的選項。

### 核心更新

1. 調整索引檔格式，加快建索引速度。
2. 調整建索引程序成為多執行緒處理，以加快建索引速度。
3. 修正重建索引時，Host欄位資訊沒有複製（消失）的問題。
4. 修正非英文的拉丁單字誤處理，導致部份拉丁單字之一字原形詞未建立索引。
5. 修正當上層文件被禁建索引時，其下層文件仍會建索引，導致最上層文件資訊錯誤的問題。亦即禁建索引文件會包括其所有的下層文件。
6. 修正文件中心資料移轉後，若其異動時間比受覆蓋系統的資料異動時間早時，移轉無效的問題。
7. 調校檢索速度（讀取欄位內容、一階排序、關鍵詞精確檢索）。
8. 修正設定文件群組（非動態）時，同一層Searcher備援機不是全部更新到（只更新一台）。
9. 修正英文字根處理(Stemming)程式非多重執行緒可呼叫的問題，當多人查詢時，可能會造成英文字根處理不正確，導致檢索錯誤，甚至破壞堆疊記憶體，引發莫名的問題。
10. 修正模糊查詢、主題查詢之動態摘要，部份應標記的字未標記的問題。
11. 修正取關聯文件時，採用的1字詞原則，舊版為同詞頻時任意取，新版為同詞頻時，依次為英文、拉丁文、中文、數字、非符號、符號。
12. 大量回傳之欄位內容與動態摘要字串，改成自行管理記憶體，以加速處理時間。
13. 增加序號功能資料（如註冊／反註冊）的額外CRC檢驗，避免誤判錯誤輸入為正確。
14. 修正部份程式誤寫，會導致文件來源資訊實際上並未載入的問題。
15. 修正當文件無內容時，會導致建索引出現『讀取資料不足』的錯誤。

### 來源處理模組更新

1. 以DataSourceSetting.exe更新附檔欄位設定，允許選擇網路芳鄰磁碟目錄。
2. 修正ODBC資料來源執行DBSQLSelectCount參數的邏輯。
3. 修正ODBC附檔欄位執行更新時，檔案不存在造成的錯誤。
4. 修正ODBC附檔欄位路徑組合的邏輯。
5. 修正Dir來源處理模組處理網路芳鄰時，產生錯誤的資料來源名稱。
6. 修正處理非只增不減模式(AddMode=false)時，刪除檔案的判斷。



## Web API更新

1. 加強錯誤紀錄機制。
2. 新增『Gufonet WebAPI工具使用率』功能。
3. 變更檔案解析時，沒有權限的回報訊息。

版本號碼：1.04e

發佈日期：2022/07/01

更新內容：

### 工具新增

1. 新增Gufonet索引設定暨工具配置圖形介面。

### 核心更新

1. 修正MoveStorageDoc在有待回收已刪除文件時，會固定選擇特定Meta Data儲存機之問題。
2. 修正從1.02版升級、建過索引、其文件更新時間為空時，會造成索引失敗的問題。

### 來源處理模組更新

1. 修正遞迴處理壓縮檔時，深層檔案解析錯誤。
2. 修正ODBC資料模組，在進行新舊資料比對時之錯誤。
3. 修正ODBC資料模組，若Mappings.DBFileField[].ExpandAttachment = false, 且為多檔欄位時，只寫入最後一筆檔案內容的問題。
4. 修正目錄檔案模組，子目錄文件之資料來源名稱錯誤之問題。

### Web API更新

1. 提供啟用swagger之bat檔。
2. 修正api/DocDetails功能，欄位名稱大小寫不一致會取不到值的錯誤。
3. 修正在序號未具備相關權限時，無法正常運作之問題。
4. 當QueryOption與ResultOption設定不合權限時，改為回傳警告而非無法執行。
5. 修正appsettings.json參數CORS:AllowOrigin，填寫全域"\*"時會出錯的問題。

### Web UI更新

1. 修正WebUI以非本地網址(localhost, 127.0.0.1)登入時，會跳回登入畫面的問題。
2. 調整WebUI的序號反註冊功能，加上確認的流程。
3. 修正在序號未具備相關權限時，無法正常運作之問題。

版本號碼：1.04c

發佈日期：2022/04/01

更新內容：

### 安裝方式異動

以安裝程式方式進行安裝，詳情請見使用手冊。

### 核心更新

1. 檢索統計功能擴增為可一次統計多個。
2. 擴增AbstractMaxCharNo上限到50000。
3. 增加AbstractMode回傳結果選項，可控制動態摘要的輸出原則。
4. 增加父文件、最上層父文件系統欄位\$Parent\$、\$TopParent\$，可篩選

與取值 (索引檔版本2, 請依序更新Doc、Search、Index、SearchMain伺服器)。

- 查詢選項SearchOption增加TopmostDocOnly選項, 只回傳最上層父文件, 但動態摘要含有查到的各子文件之資訊與動態摘要。
- 規格變更: 資料庫多重建值文件名稱之分隔字元, 由ASCII 02h改成08h字元, 避免與動態摘要之分隔字元衝突(須修訂資料庫來源處理程式)。
- 查詢選項SearchOption增加查詢詞是否計入熱門查詢詞選項(AddHotKeyword)。
- 增加強制立即重建索引功能函數RebuildIndex, 並可選擇是否先刪除舊索引, 以便空出硬碟空間。
- 修正RemoveNotIndexDoc函數, 最後一個文件ID未處理到的問題。
- 修正字串欄位排序資料寫入錯誤格式, 導致檢索出現錯誤的問題(須重建索引)。
- 修正字串欄位設定為排序等屬性後, 建索引會出現Exception的問題
- DocMain啟動時, 自動檢驗各儲存機, 以校正最大文件ID數(Interface, DocMain, Doc)。
- 修改當文件中心父子文件關係有問題時, 在建索引時自動視為無父文件
- 修正帶not查詢語句時, 沒有提示詞與關聯詞的問題。
- 修訂動態摘要輸出原則, 原本以行為原則, 每行最長160字元(前80後80), 改依(AbstractMaxCharNo/所得詞數)而定, 以突出AbstractMaxCharNo與AbstractMaxTermNo參數的成效。
- 修訂在模糊查詢或主題查詢時, 詞頻欄位\$tf\$改成回傳相似度分數
- 修正字串欄位統計若超出應回傳數量時會發生錯誤。

#### 來源處理模組更新

- 修正批次新增資料時會遺漏最後一筆。
- 修正ODBC模組DBFileFields[].RootFolder指定為null時出錯。
- 修正ODBC模組中TextParsingFields之HTML標籤的處理不正確。
- 修正ODBC模組DBFileFields未設定時發生錯誤的問題。
- gufonet.config檔的存放位置, 改由系統變數%GUFONET%決定。

#### Web API更新

- 以Window Service提供獨立Web Server (不須依賴IIS) 的方式運作。
- api/DocDetails修正\$parent\$欄位取得結果。
- api/Search QueryOption增加TopDocOnly選項, 只回傳最上層父文件。
- api/Search 規格變更: 資料庫多重建值文件名稱之分隔字元, 由ASCII 02h改成08h字元, 避免與動態摘要之分隔字元衝突(須配合資料庫來源處理程式的更新); 資料庫來源處理程式設定檔新增Mappings.PrimaryKeySeperator參數
- api/Search 查詢選項QueryOption增加查詢詞是否計入熱門查詢詞AddHotKeyword選項
- api/Search 使用核心的多次統計功能, 提升統計的效能。

#### 提供.Net Core 5 Web UI

- 以Window Service提供獨立Web Server (不須依賴IIS) 的方式運作。

#### 新增開發輔助工具

1. 檢索結果did匯出工具：匯出檢索結果文件id。工具位置Tools\GnDocDI DsExporter.exe
2. 禁索引文件工具：工具位置 Tools\GnNotIndexDoc.exe。

版本號碼：1.02a

發佈日期：2022/01/07

更新內容：

#### 核心更新

1. 增加熱門查詢詞的功能（API異動，必須更新COM元件）。
2. 增加：設定索引機重建索引的條件（異動文件數、更新次數、允許時間），以及索引更新間隔時間。
3. 修改建索引時，採用批次方式取得索引文件內容；加速文件匯入速度。
4. 檢索功能獨立分流，不與其他功能共用分流管控，改善負載平衡問題。
5. 修正依詞頻(\$TF\$)排序時會有錯誤的問題。
6. 修正同義詞檢索時無效的問題（需重建索引）。
7. 修正索引更新流程的bug，造成不必要的索引重建。
8. 修正特定查詢語句無法查到資料的問題。
9. 修正字串欄位統計，若起出應回傳數量時出錯的問題。
10. 修正建索引時，在特定情況下會出現檔案無法刪除而錯誤的問題。
11. 增加一個索引設定參數，當硬碟剩餘空間少於指定數量時，即重建索引。
12. 修正檢索機因磁碟存取造成的排序效率降低問題，並予以優化。

#### 來源處理模組更新

1. 網站模組加上續傳機制。
2. 網站模組加上HTML XPath資訊過濾機制。
3. 改善網站模組的暫存檔使用空間。
4. 修正ODBC模組在匯入時，對於\$Catalog\$映對不正確的問題
5. 改為使用.Net 5開發(需使用.Net 5執行環境)

#### Web API更新

1. 修正跨網域使用時CORS錯誤的問題。
2. 加入服務請求的排隊機制，回應伺服器過忙以避免過載。
3. 改為使用.Net 5開發(需使用.Net 5執行環境)