

Gufonet 發佈版本通知

版本號碼：1.22

發佈日期：2023/12/29

更新內容：

一、 工具更新

1. Gufonet Manager

- 匯入設定之排序，允許依欄位排序，再手動調整順序。
- 新增網路爬蟲支援 Single-Sign-On 網頁登入功能。
- 修正序號更新時，顯示之序號資訊未更新之問題。
- 修正系統警示訊息設定後，再重新進入時，未能顯示先前設定之問題。
- 修正數個頁面顯示資料時，無法完整顯示之問題；如實體索引、邏輯索引之顯示頁。
- 修正視窗最大化時，標題列被遮蔽之問題。

2. Web API

- 增加紀錄使用者查詢詞及結果篇數之紀錄，並允許在工具使用率查詢頁面中查詢。

3. 安裝程式

- 反安裝時，若系統未對序號進行反註冊，提供提醒並在有網路的情況下，代系統管理者進行反註冊程序。
- 文件安裝，改為可選擇；因在發佈時亦會提供文件。

二、 核心更新

1. 序號資訊增加『索引更新期限』之設定；將此期限與 MA 期限區隔出來。
2. 修正設定系統警示通知後，實際上卻收不到警示 email 之問題。
3. 修正硬碟空間檢驗，僅限於重建索引之後的索引檔，避免更新索引膨脹後索引檔導致檢驗錯誤。

版本號碼：1.21

發佈日期：2023/10/06

更新內容：

三、 工具更新

1. 『Gufonet 來源文件管理』升級為『Gufonet Manager』，不再需要使用 Web UI (取消)，統一系統管理介面。

- 增加系統設定、註冊、序號註冊等功能；介面套用新佈景主題。
- 針對單一主機之安裝，新增『一鍵設定』的功能，一次完成七種

伺服器的設定。

- 針對 ODBC 資料來源，若資料表中有資料時間欄位，新增『SQL 更新匯入』功能，可僅擷取更新的資料進行高效的更新匯入，不必作整個資料表擷取的工作。
- 序號註冊加速：若安裝之機器可直接連線到 Gufonet 序號機，可一次完成註冊，不必進行兩階段序號註冊程序。若安裝之機器無法連線 Gufonet 序號機，則仍舊使用兩階段序號註冊程序。
- 刪除實體索引時，可連帶刪除相關之文件來源設定。
- 匯入設定分頁，提供拖曳排序之功能；拖曳排序後，儀表板分頁的匯入紀錄，將依照相同的排序顯示。

四、 核心更新

1. 增加 NeedTotalTf 檢索參數與 GetTotalTf API，用以回傳檢索詞在所有回傳文件中出現的總次數。
2. 索引機設定增加 AutoIndexingStartTime 和 AllowIndexingTime 參數。
3. 修正多個儲存機時，增加文件會發生 StorageID 錯誤的問題。
4. 修正儲存機文件全刪光後，若進行重建索引，會造成舊文件未移除的問題。
5. 修正欄位排序檢索時，有時會遺漏部份文件的問題。
6. 修正檢索統計時，可能(不定)會多出最多 64 個沒有的文件的問題。
7. 修正檢索統計時，若欄位內容有換行字元，自動更換成空白字元，避免影響輸出解析的正確性。
8. 當字串欄位做為篩選、排序、統計、分組時，自動略除前後的空白、換行等字元；檢索時輸入字串亦自動略除。
9. 修正呼叫檢驗最大文件 ID 時，可能引發當機重啟，或回傳不正確之文件 ID 數。
10. 修正增減 SearchMain 時，必須重新啟動才能繼續的問題。
11. 調整[先行載入索引檔資料所使用的記憶體]內定值為(可用記憶體-3GB)/2，上限 4GB，且只載入基礎資訊，避免載入過久或耗用過量記憶體。
12. 先行載入索引檔資料若發生錯誤，自動忽略，避免導致系統以為有錯而無法啟動。
13. 修正更新索引時，可能會發生 Exception 或讀取資料不足的問題。
14. 加快刪除大量文件的速度，包括指定索引的文件或指定來源的文件，原刪除 500 萬筆約需 10 分鐘，加速後只需約 4 秒。
15. 修正檢索 OR 運算時，有可能卡死的情況。
16. 修正非關鍵詞檢索時，有可能引發 Exception 的問題。
17. 修正索引機傳送索引檔給檢索機時，若發生錯誤，會導致索引機建立索引時卡住的問題。
18. 修正 xml 剖析時，沒有略過<?xml-之類的 tag，並可以容忍無 end tag。
19. 動態摘要最多只取 10 萬筆（約 5 秒），避免處理太久或耗用太多記憶體。

20. 將清除 Cache 資料的工作調往背景處理。
21. 修正刪除文件造成文件群組內容異動必須更新時，呼叫到錯誤函數的問題。
22. 修正清除 Cache 時，多台 SearchMain 間遞迴呼叫，造成滿線的問題。

版本號碼：1.12

發佈日期：2023/06/30

更新內容：

一、 工具更新

1. 『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。
2. 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。

二、 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：

一、 工具更新

- 『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。
- 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。

二、 核心更新

- 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。
- 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。
- 增加支援 7z 壓縮檔之解析。
- 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。
- 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：

一、 工具更新

- 『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。
- 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。

二、 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。

三、 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。

四、 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：

一、 工具更新

1. 『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。

二、 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。

三、 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。

四、 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：

一、 工具新增

1. 新增 Gufonet 索引設定暨工具配置圖形介面。

二、 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 MetaData 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。

三、 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。

四、 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。

五、 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：

一、 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。

二、 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)。
5. 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。
6. 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)。
7. 查詢選項 SearchOption 增加查詢詞是否計入熱門查詢詞選項 (AddHotKeyword)。
8. 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。
9. 修正 RemoveNotIndexDoc 函數，最後一個文件 ID 未處理到的問題。
10. 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)。
11. 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
12. DocMain 啟動時，自動檢驗各儲存機，以校正最大文件 ID 數 (Interface,DocMain,Doc)。
13. 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
14. 修正帶 not 查詢語句時，沒有提示詞與關聯詞的問題。
15. 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。
16. 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
17. 修正字串欄位統計若超出應回傳數量時會發生錯誤。

三、 來源處理模組更新

1. 修正批次新增資料時會遺漏最後一筆。
2. 修正 ODBC 模組 DBFileFields[].RootFolder 指定為 null 時出錯。
3. 修正 ODBC 模組中 TextParsingFields 之 HTML 標籤的處理不正確。

4. 修正 ODBC 模組 DBFileFields 未設定時發生錯誤的問題。
5. gufonet.config 檔的存放位置，改由系統變數%GUFONET%決定。

四、 Web API 更新

1. 以 Window Service 提供獨立 Web Server (不須依賴 IIS) 的方式運作。
2. api/DocDetails 修正 \$parent\$ 欄位取得結果。
3. api/Search QueryOption 增加 TopDocOnly 選項，只回傳最上層父文件。
4. 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
5. api/Search 查詢選項 QueryOption 增加查詢詞是否計入熱門查詢詞 AddHotKeyword 選項
6. api/Search 使用核心的多次統計功能，提升統計的效能。

五、 提供 .Net Core 5 Web UI

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

六、 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\GnDocDIDsExporter.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：

一、 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。

二、 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)

三、 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)