NCHC-自由軟體實驗室

Crawlzilla-輕鬆打造你的專屬搜尋引擎

安裝及使用說明



【安裝前的小提醒】

- **硬體規格**建議:記憶體 1.5G 以上
- 作業系統建議:請參考測試成功的作業系統
- **系統環境**設定:
 - 。 修改主機名稱 (建議不要用 localhost 作為主機名稱)
 - 若系統為 debian 或 ubuntu (底下的 {hostname} 請修改成 您要的主機名稱)

\$ su -

cat {hostname} > /etc/hostname

/etc/init.d/hostanme restart

// debian 則為 /etc/init.d/hostanme.sh restart

// 建議登出再登入,讓修改的主機名稱生效

若系統為 Fedora 或 CentOS (底下的 {hostname} 請修改 成您要的主機名稱)

\$ su -

vim /etc/sysconfig/network

HOSTNAME={hostname}

hostname {hostname}

// 建議登出再登入,讓修改的主機名稱生效

。 確認 /etc/hosts 的 主機名稱 和 IP 位址 也是正確的對應

cat /etc/hosts

127.0.0.1 localhost

140.100.X.X {hostname}

【安裝時需要的套件】

- 系統會自動檢查 openssh, openssh-server, Sun 6 Java 套件是否有安裝, 並試著自動安裝 (ubuntu, debian), 若其他系統無法自動安裝, 使用者需 自行手動安裝
 - 。 由於 Sun 6 Java 是版權軟體,因此某些 linux 版本預設無法安裝
 - Ubuntu 10.04 安裝可參考以下指令

sudo add-apt-repository "deb http://archive.canonical.com/ lucid partner" sudo apt-get update sudo apt-get install sun-java6-jdk sun-java6-plugin sudo update-java-alternatives -s java-6-sun

Fedora 和 CentOS 再安裝 crawlzilla 過程中會自動安裝

【Crawlzilla 單機安裝步驟】

- 安裝 Crawlzilla 在一台電腦上運作,功能與穩定性不會比安裝在多台電 腦上少,兩者差別僅於分析大型網站的效率而已。因此,安裝 Crawlzilla 在 單機上是使用者或體驗者建議的選項。安裝的過程非常簡單,只需彈指的 四步間即可完成(已安裝 sun-java-6 的前提下)。
- 如果你要讓 Crawlzilla 安裝於兩台以上的叢集系統,請看 Crawlzilla 叢集
 安裝步驟

【Step 1. 取得安裝檔】

• 至 <u>http://sourceforge.net/p/crawlzilla/home/</u> 取得 crawlzilla 最新安裝檔

【Step 2. 解壓縮並執行安裝程式】

• 參考指令如下:

tar zxvf Crawlzilla-0.2*.tar.gz

./Crawlzilla_Install/install

ps:此指令會切換成 sudoer ,因此有可能會要您的 sudoer 密碼

【Step 3. 設定密碼及確認網路資訊】

- 此一步驟將會在系統中新建一組 user 帳號-crawler · 系統服務及叢集間的 溝通將會已此一帳號密碼作為執行身份。
- 設定密碼並確認網路狀態資訊後,等候完成安裝即可。
- 畫面如下:

| 檔案(F) 編輯(E) 檢視(V) 終端機(T) 求助(H) | |
|--|---|
| 正在重建相依關係 | |
| 正在讀取狀態資料完成 | |
| 正在讀取延伸狀態檔案 | |
| 初始化套件狀態完成 | |
| 沒有套件將會被安裝、升級或移除。 | |
| 0 個套件升級,0 個新安裝,0 個將移除且 13 個不會升級. | |
| 需要下載 0B 的歸檔檔案. 解裝後將用去 0B. | |
| 止在編輯延伸抹態計局・・・元成 | |
| 止仕讀取長件清里元成 | |
| 正在重建作化物质体 | |
| | |
| 工业最低产价加速调素 | |
| | |
| 系統有 Sun Java 1.6 以上版本 | |
| 系統已有 ssh. | |
| 系統已有 ssh Server (sshd). | |
| 系統已有 dialog. | |
| 歡迎使用Crawlzilla,此安裝程序會為您新建一個crawler帳號並協助您設定密碼 | |
| 請輸入欲設定的crawler密碼: | |
| password: | |
| | |
| 請用職人一次催認密碼: | |
| password: | |
| Mactor细路TD位计为。 1/0 110 132 186 | |
| Master的MAC為, A8,A8,27,99,44,89 | |
| | = |
| | |
| | |

• 待出現"恭喜您完成 Crawlzilla 安裝,按 Enter 鍵離開..."即表示單機環境已 安裝完成!安裝完成後開啟網頁將會顯示畫面如下:

| 檔案 (<u>F</u>) 編輯 (<u>E</u>) 檢視 () | <u>/</u>) 歴史 (<u>S</u>) 書籤 (<u>B</u>) 工具 (<u>T</u>) 說明 (<u>I</u> | <u>+</u>) | | | | |
|---------------------------------------|--|---|--------------------------------|------------|---|-----------|
| 🖕 🗼 🔻 🥰 😣 | (a) http://140.110.138.186:8 | 3080/crawlzilla/cha | ange_PW.jsp | | 🛛 🛛 🚼 🔻 Google | |
| Google | ▼ 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 | \$\$\$ ⊠ ▼ -} ▼ \$ | 👩 🔻 🖸 分享 🔻 | ■ 網頁註解 ▼ 5 | | Shunfa ▼ |
| CrawlZilla | + | | _ | , | , i i i i i i i i i i i i i i i i i i i | |
| CrawlZilla | 網頁管理介面 | | | | | |
| CRAWLZILLA | 首頁 爬網設定 | 索引庫管理 | 系統狀態 | 使用者設定 | | |
| | 這是你第一次登入 安全考量 | , 預設的密碼不 | 下該被使用 | 1 | 搜尋引擎快速 CrawlZilla 搜尋引 | 連結 肇範例 |
| | 原密碼為 | | | | 相關資源 | |
| | 新設定的密碼 | | | | CrawlZilla@Goog | leCode |
| | 確認新設定的密碼 | | | | | |
| | 送出 | 重設 | | | | |
| | сору Тет | right © 2010 Free Softwa Islate provided by: Design | are Lab@NCHC isByDarren.com | | | |
| 完成 | | | | | | 🖺 🔀 |

【註解】

 單機版安裝程序完成後,系統將會自動開啟 tomcat 服務及 hadoop 中的 namenode 及 jobtracker,若要執行網頁 crawl 功能需自行透過系統管理介 面開啟 datanode 及 tasktracker。

【Crawlzilla 叢集安裝說明】

用一台普通 pc (core 2 CPU, 2G Mem) 安裝了 Crawlzilla 的主機·去爬取總 資料量約 200MB 的任務·需要 6 個小時。因此·如果你的搜尋需求更大·即時 性更急迫些·你可以考慮用叢集運作的方式來平行分散工作到多台電腦去運算。

要完成安裝 Crawlzilla 的叢集模式一點都不複雜,只要注意安裝步驟即可。概 念是,安裝的第一台為 Master,此台也可以獨立執行運作全部的服務,方法與 Crawlzilla 單機系統安裝說明內容一模一樣;第二台以上的電腦,則利用 Master 所產生出來的 Client 安裝包來完成安裝;之後就可以用 Crawlzilla 的管理工具輕 鬆的動態[新增 \ 移除]運算節點囉!

你會發現 A 安裝 Master Server 與單機安裝的步驟方法是一模一樣的,因此如 果你已經操作了 Crawlzilla 單機系統安裝說明,之後又想要再加入第二、三.... 多個節點,可以直接跳到【 Step B. 安裝 Slave 節點】分別執行即可

【安裝前的小提醒】

- **硬體規格**建議:記憶體 1.5G 以上
- 作業系統建議:

請至 http://code.google.com/p/crawlzilla/wiki/Support_Distribution 參考測試 成功的作業系統

- 系統環境設定:
 - 。 修改主機名稱 (建議不要用 localhost 作為主機名稱)
 - 若系統為 debian 或 ubuntu (底下的 {hostname} 請修改成 您要的主機名稱)

\$ su # cat {hostname} > /etc/hostname

/etc/init.d/hostanme restart

// debian 則為 /etc/init.d/hostanme.sh restart

// 建議登出再登入,讓修改的主機名稱生效

若系統為 Fedora 或 CentOS (底下的 {hostname} 請修改 成您要的主機名稱)

\$ su -

•

vim /etc/sysconfig/network

HOSTNAME={hostname}

hostname {hostname}

// 建議登出再登入,讓修改的主機名稱生效

。 確認 /etc/hosts 的 主機名稱 和 IP 位址 也是正確的對應

cat /etc/hosts

127.0.0.1 localhost

140.100.X.X {hostname}

【安裝時需要的套件】

- 系統會自動檢查 openssh, openssh-server, Sun 6 Java 套件是否有安裝, 並試著自動安裝 (ubuntu, debian), 若其他系統無法自動安裝, 使用者需 自行手動安裝
 - 。 由於 Sun 6 Java 是版權軟體,因此某些 linux 版本預設無法安裝
 - Ubuntu 10.04 安裝可參考以下指令

sudo add-apt-repository "deb http://archive.canonical.com/ lucid partner"

sudo apt-get update

sudo apt-get install sun-java6-jdk sun-java6-plugin sudo update-java-alternatives -s java-6-sun

Fedora 和 CentOS 再安裝 crawlzilla 過程中會自動安裝

【A 安裝 Master Server 】

此一安裝過程將假設欲安裝 crawlzilla 單機版於 PC1 中

【Step A1. 取得安裝檔】

• 至 http://sourceforge.net/p/crawlzilla/home/取得 crawlzilla 最新安裝檔

【Step A2. 解壓縮並執行安裝程式】

• 參考指令如下:

tar zxvf Crawlzilla-0.2-100813-Shell.tar.gz

./Crawlzilla_Install/install

【Step A3. 設定密碼及確認網路資訊】

- 此一步驟將會在系統中新建一組 user 帳號-crawler · 系統服務及叢集間的 溝通將會已此一帳號密碼作為執行身份。
- 設定密碼並確認網路狀態資訊後,等候完成安裝即可。
- 畫面如下:

| 檔案(F) 編輯(E) 檢視(V) 終端機(T) 求助(H) | |
|--|--|
| 正在重建相依關係 | |
| 正在讀取狀態資料完成 | |
| 正在讀取延伸狀態檔案 | |
| 初始化套件狀態完成 | |
| 沒有套件將會被安裝、升級或移除。 | |
| 0 個套件升級,0 個新安裝,0 個將移除且 13 個不會升級. | |
| 需要下載 0B 的歸檔檔案. 解裝後將用去 0B. | |
| 止在編輯進伸狀態訊息 | |
| 止仕讀取套忤演里元成 工态素决坦体盟派 | |
| 止仕里進相報 開係 | |
| 正在很权权您具料元成 正在德职延伸出能觉安 | |
| 山在最低处理论。金属、 | |
| | |
| 系統有 Sun Java 1.6 以上版本 | |
| 系統已有 ssh. | |
| 系統已有 ssh Server (sshd). | |
| 系統已有 dialog. | |
| 歡迎使用Crawlzilla,此安裝程序會為您新建一個crawler帳號並協助您設定密碼 | |
| 請輸入欲設定的crawler密碼: | |
| password: | |
| | |
| 請冉翰人一次確認密碼: | |
| password: | |
| Mactor 细路TD位 计 在 1/0 110 138 186 | |
| Master的MAC A · A8·A0·27·99·4d·A9 | |
| | |
| | |
| | |

• 待出現"恭喜您完成 Crawlzilla 安裝,按 Enter 鍵離開..."即表示單機環境已 安裝完成!安裝完成後開啟網頁將會顯示畫面如下:

| 檔案 (<u>F</u>) 編輯 (<u>E</u>) 檢視 (<u>V</u> | () 歷史 (<u>S</u>) 書籤 (<u>B</u>) 工具 (<u>T</u>) 說明 (| <u>H</u>) | | | | |
|---|---|--|------------------------------|------------|---------------------|------------------|
| 🖕 🗼 ד 🔁 😣 | http://140.110.138.186: | 8080/crawlzilla/cha | nge_PW.jsp | ☆ 1 | ▼ Google | 0 |
| Google | ▼ 🧏 搜尋 🔻 🖗 | 🎰 M 🛛 🖕 🕷 | 💋 🔻 🔁 分享 🔻 | 🔲 網頁註解 🔻 🕻 | 7 書籤 🔻 » 🔧 🔻 🔵 s | hunfa 🔻 |
| CrawlZilla | + | | | | | ₹ |
| CrawlZilla | 網頁管理介面 | | | | | |
| CRAWEZILLA | 首頁 爬網設定 | 索引庫管理 | 系統狀態 | 使用者設定 | | _ |
| | | | | | 搜尋引擎快速連續 | 結 |
| | 這是你第一次登入 安全考量 | 1,預設的密碼不 | 該被使用 | | CrawlZilla 搜尋引擎範 | (9) |
| | 原密碼為 | | | | 相關資源 | |
| | 新設定的密碼 | | | | CrawlZilla@GoogleCo | ode |
| l | 確認新設定的密碼 | | | | | |
| | 送出 | 重設 | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | right © 2010 Free Softwa plate provided by: Design : | re Lab@NCHC sByDarren.com | | | |
| | | | | | | |
| 完成 | | | | | | <mark>.</mark> ≁ |

【Step B 安裝 Slave 節點】

【Step B1. 透過 PC1 取得安裝提示】

於 client 端執行"ssh PC1", 並執行 "crawlzilla" 指令, 找到"client 安裝步
 驟", 如下圖所示:

| 檔案(F) 編輯(E) 檢 = [Crawlzil] | 視(V) 終蹀機(T) 求助(H) | - |
|-----------------------------------|-------------------|---|
| | [管理功能選項] | |
| | in 迭存: | |
| | < 確定 > < 取消 > | |
| | | |
| | | |
| | | |

• 相關提示字元範例如下:

 $\label{eq:crawler} \ensuremath{\$} scp\ crawler @PC1:/home/crawler/crawlzilla/source/client_deploy.sh\ .$

- \$./client_deploy.sh
 - 由於此一步驟需以 crawler 的身份 ssh 至 PC1 · 因此過程中約需輸入 1~2
 次 crawler 密碼

【Step B2. 於 PC2 執行上述之提示字元】

 取得提示執行後輸入主機之 clawer 密碼兩次並確認網路資訊即可自動完 成安裝

執行畫面如下:

• 確認安裝資訊

| 檔案(F) 編輯(E) 檢視(V) 終端機(T) 求助(H) | | | | | |
|---|--------|------|-----------|-------|---|
| shunfa@ubuntu-187:~\$ ls | | | | | 4 |
| client deploy.sh hosts T | 載圖 | 片 文件 | ⊧ 模板 | | |
| examples.desktop ubuntu10.04-sun-java.sh 🖄 | 共 影 | 片 桌面 | 百樂 | | |
| shunfa@ubuntu-187:~\$./client deploy.sh | | | | | |
| checking ssh found | | | | | |
| checking sshd found | | | | | |
| crawler@140.110.138.186's password: | | | | | |
| client_deploy.sh | 100% | 516 | 0.5KB/s | 00:00 | |
| client_install | 100% | 2882 | 2.8KB/s | 00:00 | |
| client_install_func.sh | 100% | 10KB | 10.5KB/s | 00:00 | |
| client_remove | 100% | 4833 | 4.7KB/s | 00:00 | |
| CrawlzillaForClientOf_140.110.138.186.tar.gz | 100% | 54MB | 26.8MB/s | 00:02 | |
| lang_en_US | 100% | 16KB | 15.5KB/s | 00:00 | |
| lang_zh <u>:</u> TW/2007.00700677 | 100% | 15KB | 15.4KB/s | 00:00 | |
| log.sh | 100% | 996 | 1.0KB/s | 00:00 | |
| README.txt | 100% | 192 | 0.2KB/s | 00:00 | |
| /home/shunfa/crawlzilla_client_install/client | _insta | all | | | |
| [sudo] password for shunfa: | | | | | = |
| Sorry, try again. | | | | | |
| [sudo] password for shunfa: | | | | | |
| <pre>/home/shunfa/crawlzilla_client_install/client_</pre> | _insta | ill | | | |
| 身份是 root | | | | | |
| Master 的 IP位址: 140.110.138.186 | | | | | |
| 資料是否正確 (yes/no): yes | | | | | |

• 輸入密碼

| 檔案(F) 編輯(E) 檢視(V) 終端機(T) 求助(H) |
|---|
| 正在讀取狀態資料完成 |
| 正在讀取延伸狀態檔案 |
| 初始化套件狀態完成 |
| 正在編輯延伸狀態訊息完成 |
| 沒有套件將會被安裝、升級或移除。 |
| 0 個套件升級,0 個新安裝,0 個將移除且 30 個不會升級. |
| 需要下載 0B 的歸檔檔案. 解裝後將用去 0B. |
| 正在編輯延伸狀態訊息完成 |
| |
| 正在重建相依關係 |
| |
| 正在讀取延伸狀態檔案 |
| 初始化套件狀態 完成 |
| 达到"client安装步器",如下窗所示: |
| CNECK_SUNJAVA |
| Crawlzilla 需要 Sun Java JDK 1.0 以上的成本 変体素 Sun Java 1.6 NL L版本 |
| 系統有 Sun Java I.O 以上版本 |
| 永航口有 SSII. 系统日本 sch Server (schd) |
| 示航已有 SSII SEIVEI (SSIIU). 系统日右 dialog |
| |
| |
| 請輸λ Master ト crawler 使田老的廖碼・ |
| 請爾納久 Huster 上 erancer 使用自時出版 1 |
| |

• 安裝完成



【Step B3. 驗證是否安裝成功】

 於 PC1 執行指令-"crawlzilla",出現以下畫面後選擇 "檢查 Cluster 狀態", 畫面如下:

| 檔案(F) 編輯(E) 檢視(V) 終端機(T) 求助(H) | |
|--|---|
| [管理功能選項] 請選擇: | 1 |
| cluster_status cluster_setup server_setup tomcat_switch | |
| <mark>< 確定 ></mark> < 取消 > | |
| | |

• 若出現2個運算節點表示安裝成功!

| 檔案(F) 編輯(E) 檢視(V) 編 = [Crawlzilla 管 | ^{终端機(T)} 求助(H) 理介面]~by N(| CHC = | | A |
|--|---------------------------------------|---------------------------|--------------------------|----------|
| [IP] | [Hostname] | [Cluster 狀態] [Network] | [Dtatnode & Tasktracker] | 7 |
| 140.110.138.186 140.110.138.187 | ubuntu-186 ubuntu-187 | online online | running shutdown | I |
| | | < 🛍 📙 > | 100% | |

【註解】

- 叢集版安裝完成後,需回 PC1 執行系統管理介面開啟運算服務後才可加入 crawl 運算分派資源中。
- 第三個運算節點以上的安裝方式,則是重複步驟 B 即可

【Crawlzilla 網頁執行介面】

管理介面預設網址為: <u>http://localhost:8080</u> 或 <u>http://ServerIP:8080</u>

• 登入後首頁如下:

| 檔案 (<u>F</u>) 編輯 (<u>E</u>) 檢視 (⊻) 歷史 (<u>S</u>) 書籤 (<u>B</u>) 工具 (<u>T</u>) 說明 (<u>H</u>) | |
|--|---|
| 🔶 🗼 🔻 🥃 🏠 💿 http://140.110.138.186:8080/crawlzilla/index.jsp 😭 | 🕽 🔻 🖁 🔻 Google 🔍 |
| Coogle 🛛 🔻 提尋 🔻 🗄 📉 🔻 🚭 🗸 🖉 分享 🛛 | 🗸 » 🔌 🔻 🔵 shunfa 🔻 |
| | v |
| CrawlZilla 網頁管理介面 CTAWLZ 首頁 應網設定 索引庫管理 系統狀態 使用者設定 | 登出系統 |
| 管理介面功能介紹 ·抓取網頁設定:透過網頁介面建立搜尋引擎察引 ·察引庫管理:查詢、删除已爬取的察引庫 ·系統狀態:瀏覽目前系統狀態 ·管理者設定:包含個人化與語系設定 ·修改密碼 | 搜尋引擎快速連結 CrawiZilla 搜尋引筆範例 系統功能 修改管理者表碼 相關資源 CrawiZilla@GoogleCode |
| 完成 | X • |

【設定網頁管理者密碼】

首次進入網頁介面時,必須先重設管理者密碼(預設密碼為:crawler),
 設定密碼點選送出並重新登入後就可執行系統。

| 檔案 (<u>F</u>) 編輯 (<u>E</u>) 檢視 (<u>)</u> | /) 歷史 (<u>S</u>) 書籤 (<u>B</u>) 工具 (T) 說明 (<u>F</u> | <u>i</u>) | | |
|---|---|--|--------------------|-------------------------------|
| 🔶 🗼 ד 😅 😣 | http://140.110.138.186:8 | 080/crawlzilla/change_PW.jsp | ☆ <mark>1</mark> ▼ | 🛚 🛃 🔻 Google 🔍 |
| Google | ▼ 🧏 搜尋 ▼ 🔶 | 🎰 M 🔻 🐈 🔻 🧔 分型 | 🔋 🔍 🦏 網頁註解 🔻 😭 | 書籤 🔻 » 🔌 🔻 🔵 shunfa 🔻 |
| CrawlZilla | + | | | ₹ |
| CrawlZilla | 網頁管理介面 | | | |
| CRAWLZILLA | 首頁 爬網設定 | 索引庫管理 系統狀態 | 使用者設定 | |
| | 這是你第一次登入 安全考量 | ,預設的密碼不該被使用 | | 搜尋引擎快速連結 CrawlZilla 搜尋引擎範例 |
| | 原密碼為 | ••••• | | 相關資源 |
| | 新設定的密碼 | | | CrawlZilla@GoogleCode |
| | 確認新設定的密碼 | | | |
| | 送出 | 重設 | | |
| | сору Тетр | ight © 2010 Free Software Lab@NCHC late provided by: DesignsByDarren.com | | |
| 完成 | | | | PB |
| | | | | |

 當然之後若要在更改密碼,也可以透過右方的連結 "更改網頁密碼" 以進 行更改

【1. 建立第一個搜尋引擎】

【Step 1.1.開啟所有運算服務】

由於執行 Crawl 必須透過 Hadoop 運算,因此執行 Crawl 前請先依序確認以下服務是否已開啟,若為關閉狀態,請依序開啟這些服務。

- Namenode and Jobtracker
- Datanode and Tasktracker(需開啟全部的運算節點)

若不熟悉開啟步驟,請參考系統管理介面操作說明?

【Step 1.2.至 Crawl 網頁中設定爬取項目】

依序填入:索引庫名稱·欲抓取的網址(可多行·如圖所示)及設定爬取
 深度即可送出

| 檔案(E) 編輯(E) 檢視(| V) 歴史 (S) 書簡 (B) 工具 (I) 説明 (H) | | | |
|------------------|--|-----------------------|--------------------|----------|
| 👙 🗼 ד 😅 🖸 |) 🏫 间 http://140.110.138.186:8080/craw | lzilla/crawl.jsp | 🏫 🚺 🔻 🔡 🔻 🛛 Google | ۹, |
| Google | ■ 312日 ▼→ 金 M ▼ | - 👍 🔻 🦽 🛪 📴 分享 🔻 🔍 網頁 | 註解 🛪 🏠 書籤 🛪 🔌 🛪 🍯 | shunfa 🔻 |
| CrawlZilla | ★ ■家高速網路與計算中心 Intr | 2 6 回研究回想中心 | x + | * |
| CrawlZilla | 網百管理介面 | | | / |
| Cruwizina Tal | | | | |
| CR BWI 2111 B | | | | |
| CITAMERIC | 首頁 爬網設定 索引庫管理 | 系統狀態 使用者設定 | 豐出系統 | _ |
| 爬網設定 | | | 搜尋引擎快速速的 | 4 |
| | | | CrawlZilla 搜尋引筆氣 | |
| AH I-J DC/13 | | | 系统功能 | _ !! |
| - 索引庫名: | 稛 | | | _ |
| | 索引庫名稱 [tracCloud_and_nch | ncTW_3 | | _ |
| | | | 相關貿惠 | _ |
| ■■輸入抓取 | 的網址 | | CrawlZilla@GoogleC | ode |
| | http://trac.nchc.org.tw/cloud/ | | | |
| | nccp://www.ncnc.org.cw/cw/ | | | |
| | | | | |
| | | | | |
| 輸入抓取的網絡 | 址 | | | |
| 447 100 44 100 | | | | |
| | 度設定 | | | |
| | 通律抓取的深度 3 1 | | | |
| | 送出 重19 | | | |
| 完成 | | | | |

• 送出後如圖所示,等候時間需視視每台主機的運算速度而定。



【Step 1.3.瀏覽網頁爬取進度】

• 透過系統狀態頁面,可即時了解網頁爬取進度

| 檔案 (<u>F</u>) 編輯 (<u>E</u>) 檢視 (<u>V</u>) 歷史 (<u>S</u>) (| 書籤 (<u>B</u>) 工具 | (<u>T</u>) 說明 (<u>F</u> | <u>1</u>) | | |
|---|------------------------|----------------------------|---------------------------------------|------------|-----------------------|
| 븢 🗼 v 🥰 🛽 🏠 🗻 | http://140.110 | .138.186:8 | 080/crawlzilla/sysinfo.jsp | ু 🚺 🖌 | Google 🔍 |
| Google | ▼ 😽 | 搜尋 ▼↔ | 🎰 M 🔻 📲 🛛 🧔 🖉 分享 🔍 💭 網頁語 | 註解 🔻 🖒 | 書籤 🔻 🔌 🔻 🔵 shunfa 🔻 |
| CrawlZilla | 🔪 國家高速 | 網路與計算中 | 中心_Intr 🗱 🕑 國研院國網中心 | * | ÷ • |
| CrawlZilla 網頁管 CMWL/IIIA 普頁 | 理介面 _{爬網設定} |] 索引 | 章管理 <u>系統狀態</u> 使用者設定 | 登出系統 | |
| 77 R/G11/1 X8X | | 승규 그는 아프 아 | ۵. Mis | | CrawlZilla 提尋引筆範例 |
| tracC | 索引庫名 _loud_and_ | 繁5]庫新 稱 nchcTW_ | ★ 1 新取狀態 删除狀態 3 crawling Delete | | 系統功能 |
| Jol | otracker ⊥ f | 乍排程器制 | 犬態 (New Window) | | 相關資源 |
| Running Jobs | | | Quick | Links 🔺 | CrawiZilia@GoogleCode |
| completed jobs | | | | | |
| Jobid | Priority | User | Name | Mai Con | |
| job_201008181050_0001 | NORMAL | crawler | inject tracCloud_and_nchcTW_3/urls | 100 | |
| 完成 | | | | | 50 |

• 待出現"Finish"表示索引庫已建立,並可將此一訊息刪除



• 完成此一步驟,第一個搜尋引擎已建置,右側快速連結中的 "tracCloud_and_nchcTW_3"即為此次所建立的搜尋引擎。

【Step 1.4.測試搜尋引擎功能】

點選右側快速連結中的"tracCloud_and_nchcTW_3"進入搜尋引擎後,輸入
 一組關鍵字測試搜尋結果,下圖為輸入"nchc"為例:

| 檔案(E) 編輯(E) 检視(⊻) |) 歴史 (S) 書籍 (B) 工具 (T) 説明 (H) | |
|-------------------------|--|------------|
| 🝁 🌩 🔻 😅 😒 | | ۹, |
| Google | 🗾 📲 隆尋 平 🕂 🏫 闷 平 🎲 平 🧭 平 🐼 平 😵 🔧 🔍 平 | 🔵 shunfa 🔻 |
| 😤 Crawlzilla Search Eng | ine 🖶 | Ŧ |
| annaziun | Crawtzilla管理介面 | |
| | | |
| | NCHC HER ME | |
| | | |
| | ca de en es fi fi bu if jn ms ni pi p5 sh st st fi zh cn | |
| | Refer Ling BR意志思想(320)合計師中心 Eveter LPC Better Ling Eveter LPC Better Ling | |
| | | |
| | | |
| | | |
| 完成 | | M • |

• 搜尋結果:



【2.索引庫管理】

索引庫管理頁面中將會顯示目前已建立的所有索引庫·管理者可於此頁面
 進行瀏覽·刪除及提供網頁嵌入語法·如下圖所示:

| 檔案 (<u>F</u>) 編 | _{扁輯} (<u>E</u>) 檢視 (⊻) | 歷史 (<u>S</u>) 書籤 (<u>B</u>) 工具 (<u>T</u> |) 說明 (<u>H</u>) | | | | | |
|-------------------|-----------------------------------|---|-------------------|----------------|--------------|----------------|---------------------|---------|
| 🔶 🗼 | v C 🛛 | http://140.110.1 | 38.186:8080/0 | crawlzilla/nut | ch_DB.jsp | ☆ ! ▼ [| ∃ ▼ Google | 0 |
| Google | | v 24 | 韓 ▲ ∲ 🔍 🖡 | 🖂 v 👍 v . | 🐔 🔻 🔽 分享 🔻 🔋 | ■網頁註解 ▼ 5 | 🟠 » 🔍 🔻 🔵s | hunfa 🔻 |
| CrawIZ | /illa | | 1 113 | | | | ~~~~~ | v |
| C ciunz | | - | | | | | | |
| Cra | awlZilla 🕯 | 周頁管理介面 | | | | | | |
| | | | | | | | | |
| CRA | WLZIĻLA 🛛 📸 | 頁爬網設定 | 象引庫管理 | 系統狀態 | 使用者設定 | 登出系統 | | |
| 索引庫 | 管理 | | | | | 扬 | 椱尋引擎快速連結 | |
| | キョナクダ | 28 n== 88 | | 預覽 | 嵌入搜尋引擎 | c | rawlZilla 搜尋引擎範例 | |
| | <u>家</u> 51 庫 名 碑 | 建立时间 | 惻际 察 51 厚 | 統計資料 | 到網頁的語法 | n | chc-en_3 | |
| | nchc-en_3 | 2010-08-24_16:16:14 | Delete | Preview | embed code | n | chc-tw_3 | |
| | nchc-tw_3 | 2010-08-24_15:22:48 | Delete | Preview | embed code | 务 | 系統功能 | |
| | | 資 | 炓 總覽 | | | * | 改管理者密碼 | |
| | 起始URL | | | | | 相 | 目關資源 | |
| 本 | 機索引路徑 | | | | | | rawlZilla@GoogleCod | e |
| 縜 | 課文字數 | 0 | 文件檔[| 數量 | 0 | | | |
| 索 | 引庫更新日 期 | | 使用者 | 名稱 | | | | |
| | | 被搜尋分 | ·析到的網址 null | : | | | | ~ |
| 完成 | | | | | | | | ו |

【2.1 索引庫瀏覽】

進入索引庫管理頁面後·在欲瀏覽的索引庫欄位點選"preview"即可瀏覽此一索引 庫的資訊·目前提供瀏覽的資訊包括:

- 爬取網址
- 爬取文字數
- 爬取文件數
- 相關索引排名

如下圖所示:

| 檔案 (] | E) 編 | 輯 (<u>E</u>) 檢視 (⊻ |) 歴史 (<u>S</u>) 書籤 (<u>B</u>) 〔 | 工具 (<u>T</u>) 割 | 說明 (<u>H</u>) | | | | | | |
|-------|---------------|-----------------------|-------------------------------------|-------------------|-----------------|---------------------|---|-------------|--------|-----------------------|-------|
| 4 | \Rightarrow | • C 🗵 | 💧 间 http://140 | .110.138. | 186:80 | 80/crawlzilla | a/Statistics.do?fileN | ame=n 🏠 | 1 . | Soogle | |
| Goog | le 🗍 | | ~ | 2 捜尋 | ▼ . | a M 🔻 ф | 🔻 🧔 🔻 🔂 分享 🔻 | 🔲 網頁語 | 主解 ▼ 7 | 🟠 » 🔌 🔻 🔵 shun | ifa ▼ |
| Cr | awlZi | lla | 4 | | | | | | | | v |
| 索 | 引庫 | 管理 | | | | | | | | 搜尋引擎快速連結 | 4 |
| | | | | | | 25 8 | | 88 4 | | CrawlZilla 搜尋引擎範例 | |
| | | 索引庫名和 | 爯 建立時間 | ₩ | 除索引 | 庫 祝計 | 日本 1 | 去 | | nchc-en_3 | |
| | | nchc-en_ | 3 2010-08-24_16: | 16:14 | Delete | Previ | ew embed code | | | nchc-tw_3 | |
| | | nchc-tw_3 | 3 2010-08-24_15: | 22:48 | Delete | Previ | ew embed code | | | 系統功能 | |
| | | | | 資料 | 息覽 | | 4 | | | | = |
| 6 | | | | | | | |] | | FC 目 坦 首 広 偽 | |
| | ŧ | 記始URL 豊志 리 欧 御 | /h a ma lar | http:/ | www. | nchc.org.t | w/tw/ | | | 相關資源 | |
| - | 华节 | 展系 1 始 (空) 井 立 字 動 | 37095 | awiencia | awizina ★4 | a/archieve/ 上燈數号 | 1036 | | | CrawlZilla@GoogleCode | |
| | 索引 | 引庫更新日 | Tue Aug 24 15: | 22:46 | | | 1030 | | | | |
| | | 期 | CŠT 2010 | | (伊月 | 百名稱 | crawier | | | | |
| 6 | HF | | 被打 | 夏尋分析 | 到的創 | 肖址: | | | | | |
| | 排 序 | | 內容 | 51用次 數 | 序 | | 內容 | 51用次 數 | | | |
| | 0 | site:www | w.nchc.org.tw | 336 | 1 | site:pcclu | ster.nchc.org.tw | 87 | | | |
| | 2 | site:bioin | fo.nchc.org.tw | 66 | 3 | site:wv | ww.narl.org.tw | 57 | | | |
| | 4 | site:edu | u.nchc.org.tw | 53 | 5 | site:serv | ice.nchc.org.tw | 35 | | | |
| | 6 | site:acc | ta.nchc.org.tw | 28 | 7 | site:coli | fe.nchc.org.tw | 14 | | | |
| | 8 | site:wlan | rc.nchc.org.tw | 13 | 9 | site:eli | b.nchc.org.tw | 13 | | | |
| | 10 | site:www. | medicalgrid.org | 13 | 11 | site:volun | teer.ncnc.org.tw | 9 | | | |
| | 14 | site:econ | rid nchc org tw | 6 | 15 | site no | w sina gov tw | 3 | | | |
| | 16 | site:asp. | 104ehr.com.tw | 3 | 17 | site:vin | nl.nchc.org.tw | 3 | | | |
| | 18 | site:ww | w.ym.edu.tw | 2 | 19 | site:wv | w.tnu.edu.tw | 2 | | | |
| | 20 | site:ww | w.usc.edu.tw | 2 | 21 | site:www | .ssvs.tp.edu.tw | 2 | | | |
| | 22 | site:www.sr | nelearning.org.tw | 2 | 23 | site:ecoc | am.nchc.org.tw | 2 | | | |
| 승규 | 1 | | | - | INC | | | | | 1991 🔽 | |

由於有加入中文分詞功能·因此可以明顯看出索引庫的建立是以"中文字詞"作為 基本單位

| 檔案 | 檔案 (E) 編輯 (E) 检視 (⊻) 歴史 (S) 書籤 (B) 工具 (T) 説明 (出) | | | | | | | | | | | | | |
|-----|--|----------------------|--------------|------------|------|-------------------------|----------------|-------|------|---|----|-----|-------------|---|
| | 🝦 🔻 😴 🛞 🔞 http://140.110.138.186:8080/crawlzilla/Statistics.do?fileName=n 👔 🛛 🚼 🛛 Google 🔍 | | | | | | | | | | | | | |
| Goo | gle | | v | #显 ▼ | r di | 🏔 M v 📥 v 🙈 v 🕅 5 | ↑夏▼ ■ 網百言 | *解, | • • | » | er | ~ (| shunfa | |
| | | 2 | • . . | | T | | and the second | 2.000 | ~~~~ | | | | | _ |
| | Idwiziii | d | 分析的 | 文件 | 刑 | 能: | | | | | | | | × |
| | 排序 | 内容 | 引用次數 | 排序 | | 内容 | 引用次數 | | | | | | | |
| | 0 | type:text/html | 989 | 1 | ī | type:html | 989 | | | | | | | |
| | 2 | type:text | 989 | 3 | Ī | type:application | 47 | | | | | | | |
| | 4 | type:application/pdf | 34 | 5 | | type:pdf | 34 | | | | | | | |
| | 6 | type:xml | 10 | 7 | | type:application/xml | 10 | | | | | | | |
| | 8 | type:msword | 3 | 9 | 1 | type:application/msword | 3 | | | | | | | |
| | | | | | | | | | | | | | | |
| | | | 出現次數前 | Ξ + | 名 | 的字符: | | | | | | | | |
| | 排序 | 内容 | 引用次數 | 非 | 序 | 内容 | 引用次數 | | | | | | | |
| | 0 | content:網 | 805 | | 1 | content:路 | 777 | | | | | | | |
| | 2 | content: | 758 | 3 | | content:中心 | 750 | | | | | | | |
| | 4 | content:計 | 744 | 5 | | content:資 | 742 | | | | | | | |
| | 6 | content:與 | 740 | | 7 | content:訊 | 734 | | | | | | | |
| | 8 | content:頁 | 712 | ! | 9 | content: | 705 | | | | | | | |
| | 10 | content:學 | 698 | 1 | 1 | content:算 | 696 | | | | | | | |
| | 12 | content:家 | 692 | 1 | 3 | content:的 | 684 | | | | | 3 | | |
| | 14 | content:關 | 676 | 1 | .5 | content:議 | 674 | | | | | | | |
| | 16 | content:統 | 666 | 1 | .7 | content:1024 | 665 | | | | | | | |
| | 18 | content:768 | 664 | 1 | 9 | content:系 | 662 | | | | | | | |
| | 20 | content:高速 | 648 | 2 | 1 | content:- | 643 | | | | | | | |
| | 22 | content:號 | 636 | 2 | 3 | content: 🖪 | 635 | | | | | | | |
| | 24 | content:站 | 633 | 2 | 5 | content:導 | 632 | | | | | | | |
| | 26 | content:解析 | 628 | 2 | 27 | content:建 | 627 | | | | | | | |
| | 28 | content:會 | 627 | 2 | 9 | content:解析度 | 624 | | | | | | | |
| | 30 | content:務 | 622 | 3 | 1 | content:覽 | 618 | | | | | | | |
| | 32 | content:請 | 614 | 3 | 3 | content:發 | 614 | | | | | | | |
| - | 34 | content:體 | 608 | 3 | 5 | content:上 | 607 | | | | | | | V |
| 元成 | | | | | | | | | | | | | SC F | |

【2.2 索引庫刪除】

• 在欲刪除的索引庫中點選刪除,確認後即完成刪除索引庫



【3.在網頁中嵌入搜尋引擎】

若企業內部有需要將 Search Bar 整合於企業首頁中,則可使用此一功能,方式如下:

【Step 3.1. 開啟索引庫管理頁面】

| 檔案(<u>F</u>) 編輯(<u>E</u>) 檢視(<u>V</u>) |) 歴史(<u>S</u>) 書籤(<u>B</u>) 工具(<u>T</u>) 訪 | 明 (<u>H</u>) | | | |
|--|--|-----------------------|-----------------|------------------|--------------|
| 🖕 🗼 🔻 ピ 😣 | http://140.110.138.1 | L86:8080/crawlzilla/n | utch_DB.jsp | 🟫 🚺 🔻 🔀 🕷 Google | e 🔍 |
| Google | ▼ 🤮 搜尋 | V 🕂 🎰 M V 🕂 V | 🧔 🔻 🔁 分享 🔍 💭 | 網頁註解 🔻 😭 » 🛛 🖏 | 🔻 🔵 shunfa 🔻 |
| CrawlZilla | + | | | | ∇ |
| | 網頁管理介面 | 引庫管理 系統 | 伏懸 使用者設定 | 豐出系統 | |
| 索引庫管理 | | | | 搜尋引擎快: | 速連結 |
| 索引庫名 | 稱建立時間間 | 除索引庫 預覧 | 嵌入搜尋引擎 | CrawlZilla 提尋 | 引擎範例 |
| nchc-en | 3 2010-08-24_16:16:14 | Delete Previe | w embed code | nchc-en_3 | |
| nchc-tw | 3 2010-08-24_15:22:48 | Delete Previe | embed code | 系統功能 | |
| | 資料 | 總覽 | | 修改管理者密码 | • |
| 起始URL | http | ://www.nchc.org.t | wltwl | 相關資源 | |
| 本機索引路徑 | /home/crawler/c | rawlzilla/archieve/ | nchc-tw_3/index | CrawlZilla@Go | ogleCode |
| 總共文字數 | 37095 | 文件檔數量 | 1036 | Crawizina@00 | logiecode |
| 索引庫更新日期 | Tue Aug 24 15:22:46 CST 2010 | 使用者名稱 | crawler | | |
| | 被搜尋分析 | f到的網址: | | | |
| 推 | 內容 引用次 | 推 | 內容 弓 | 用次 | |

【Step 3.2.點選 embed code】



<form name="search" action="http://140.110.138.186:8080/nchc-en_3/search.jsp" method="get"><input name="query" size=15></form>



【Step 3.3.複製後貼在欲整合 Search Bar 的頁面】

• embed code 範例:

| |
|--|
| <pre><form <="" action="http://140.110.138.186:8080/nchc-en_3/search.jsp" name="search" pre=""></form></pre> |
| method="get"> |
| <input name="query" size="15"/> |
| |

【4.系統操作】

將現在的工作、叢集運算與儲存系統整合於一頁

| Crav | /IZilla 糹 | 周貢管理介面 | | | | |
|------------------------------|-----------|---|----------|--------------------|-----------------------------|-----------------|
| СПЛШ | ZILLA | | វញ្ញ រ | に 網設定 索引庫管理 | 系統狀態使用者設定 | 登出系統 |
| 系統狀態 | 1 | | | | | |
| | | | 索引 | <u> </u> | 間删除狀態 | |
| | | | Jobtra | cker 工作排程器狀態 (N | lew Window) | |
| Version Compil Identif | ter Su | 745977 b 20 00:16:34 UTC 200 0141911 nmary | 09 by nd | aley | | |
| Maps | Reduces | Total Submissions | Nodes | Map Task Capacity | Reduce Task Capacity | Avg. Tasks/Node |
| 0 | 0 | 0 | 1 | 2 | 2 | 4.00 |
| Sche | duling | Information | Name | node 交間管理員狀能(N | lew Window) | |

【5.管理者設定】

可以設定 搜尋引擎的名稱 · 管理者 Email · 以及欲使用的語言 (中/英)

| CrawlZilla Management Pa | ge | | | | | |
|--------------------------|------|--|---|--------|-------|--------|
| CRAWLZILLA | Home | Crawl | Manage | System | Setup | Logout |
| Setup | | Engin waue-crawl Adm waue@ncho Choose Eng | ne Name zilla in Email c.org.tw Language glish v | | | |

【系統管理介面操作說明】

- 此一說明頁面操作環境為兩台已安裝好之叢集環境
- 單機版操作方式皆同。

【系統運算架構】

- 已熟悉 Hadoop 架構之使用者可略過此一段落
- 由於底層的運算是交由 Hadoop 作運算·相關的運算原理可參考 Hadoop 官方網頁說明
- 若無法了解上述運算架構及原理,僅需記住以下啟動順序即可。

Step1. 啟動 Namnode & Jobtracker

Step2. 啟動 Datanode & Tasktracker

 系統安裝完成時,運算節點預設為關閉,需透過系統管理介面開啟服務, 執行網頁爬取前,請先確認 Hadoop 相關運算服務已開啟,否則將無法順 利執行。

【系統管理介面功能】

在 PC1 終端機中輸入指令-"crawlzilla"即可進入系統管理介面如下圖:

| 檔案(F) 編輯(E) 檢視(V) 終端機(T) 对 [請 選 擇: | [*] 助(H) 管理功能選項] | • |
|---|---|--|
| luster_status cluster_setup server_setup tomcat_switch tomcat_port lang_switch client_install exit | 檢查 Cluster 狀態 設定 datanode & tasktracker 設定 namenode & jobtracker 啟動 /停止 /重新 啟動 Tomcat 更改 Tomcat port 更換語言 Client 安裝步驟 結束 | |
| < 疆定 | - 取消 > | (((((((((((((((((((|

此一管理介面提供數個系統管理功能,於下列子段落一一說明。

【1. 檢查 Cluster 狀態】

此一功能主要顯示叢集中的電腦狀態,主要顯示欄位分別為:

| 【 IP 】 【 Hostname 】 【 Network 連線狀態 】 【 Datanode & Tasktracker 狀 |
|---|
|---|

如下圖:

| 檔案(F) 編輯(E) 檢視(V) 編 | 冬端機(T) 求助(H) | | | A |
|------------------------------------|--------------------------|------------------|-----------------------|--------|
| | | [C]uston 收除] | | |
| [IP] | [Hostname] | [Network] | [Dtatnode & Tasktrack | (er] |
| 140.110.138.186 140.110.138.187 | ubuntu-186 ubuntu-187 | online online | running shutdown | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | < 離開 > | 1 | .00% - |
| | | | Ŕ | |

• 註:Namenode & Jobtracker 狀態需進入其管理頁面查詢之

【2. 啟動 Datanode & Tasktracker】

此一功能主要管理叢集中的電腦狀態,並可分為管理全部節點或部份節點

- 主要的功能為【啟動】·【停止】及【重新啟動】:
- 選擇管理全部或部份節點:

| 檔案(F) 編輯(E) # | 檢視(V) 終端機(T) 求助(H) | |
|---------------|-------------------------------|----------|
| = [Crawlzi] | lla 管理介面]~by NCHC = | ^ |
| | | |
| | | |
| | [全部或部份 nodes] | |
| | | |
| | All 全部 nodes Part 部份 nodes | |
| | Main 回主選單 | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | < <mark>薩定 ></mark> < 取消 > | |
| | | |
| | | |
| | | |
| | | ▼ |

• 進入後·選擇欲操作的功能:

| 檔案(F) # = [Cr | 編輯(E) 檢視(V) 終端機(T) 求助(H) awlzilla 管理介面]~by NCHC = | - |
|------------------|---|---|
| | [Datanode & Tasktracker 設定] 請選擇: | |
| | start啟動 Datanodes & Tasktrackersshutdown停止 Datanodes & Tasktrackersrestart重新啟動 Datanodes & Tasktrackersmain回主選單 | |
| | | |
| | < 確定 > < 取消 > | |
| | | ■ |

【3. 啟動 Namenode 及 Jobtracker】

此一功能主要管理叢集中的主要電腦狀態(此為 PC1)·主要的功能為啟動·停止及重新啟動·操作方式如下:

[Step 3.1]

• 進入時,程式將先檢查目前的服務狀態:

| Comon Wiff 1 | |
|--|--|
| [Server 10.08] Namenode & Jobtracker 正在運作 | |
| < 薩定 > | |
| | |

[Step 3.1]

• 進入後,選擇欲操作的功能:

| 檔案(F) 編輯(E) 檢視(V) = [Crawlzilla 管 | 終端機(T) 求助(H) [理介面] ~by NCHC = | |
|--------------------------------------|--|-----|
| 請選擇: | [Namenode & Jobtracker 設定] | |
| | tart 啟動 Namenode & Jobtracker shutdown 停止 Namenode & Jobtracker restart 重新啟動 Namenode & Jobtracker main 回主選單 | |
| | | |
| | < 薩定 > | |
| | | H N |

【4. 啟動/停止/重新啟動 Tomcat 服務】

此一功能主要管理叢集中的主要電腦狀態(此為 PC1)·主要的功能為啟動·停止及重新啟動網頁伺服器·操作方式如下:

[Step 4.1]

• 進入時,程式將先檢查目前的服務狀態:

| 【Tomcat 狀態】 Tomcat 正在執行 !!! | |
|--------------------------------|---|
| < 確定 > | - |
| | |
| | |
| | |

[Step 4.2]

• 進入後,選擇欲操作的功能:

| _ | [Tomcat 服務選項] | 1 |
|-----|--|---|
| 請選擇 | : | |
| | tart 啟動 Tomcat shutdown 停止 Tomcat restart 重新啟動 Tomcat main 回主選單 | |
| | | |
| | <mark>< 雛定 ></mark> < 取消 > | |
| | | |

【5. 更改 Tomcat Port】

此一功能主要為更改網頁伺服器的 Port 號·若預設的 Port 號已被其他程序佔用· 即可透過此功能進行更改·操作方式如下:

• 進入後,直接輸入欲更改的 Port 號即可。

| 檔案(F) 編輯(E) = [Crawlzi | 檢視(V) 終端機(T) 求助(H) lla 管理介面] ~by NCHC = | |
|---------------------------|---|---|
| | | |
| | [Tomcat Port 更改] | |
| | 請輸入要更改的 Tomcat port 號碼 (預設值為 8080): | |
| | | |
| | | |
| | | ₽ |
| | | |
| | | = |
| | | = |

【6. 更換語言】

提供中英文操作語言更換,進入後直接選擇操作語言即可。

| 檔案(F) 編輯(E) # | 桧視(V)終端機(T)求助(H) | |
|---------------|--|-----|
| - [Claw(21) | | - 1 |
| | 「建物签字」 | |
| | 請選擇: | |
| | ■ <mark>■■_US</mark> 英文 zh_TW 中文 main 回主選單 | |
| | | |
| | <mark>< 齏定 ></mark> < 取消 > | |
| | | |
| | | |

【移除 Crawlzilla】

單機移除

打開終端機,用 root 帳號或有 sudoer*權限的帳號輸入

crawlzilla_remove

sudoer 帳號需要輸入該帳號的密碼,等程式跑完就移除完囉!

叢集移除



- Crawlzilla 的叢集是主(master)-從(slave)架構的服務模式,因此移除
 的時候,先將所有 slave 移除之後,最後再移除 master。
- master 與 slave 的移除都是輸入指令

crawlzilla_remove

Slave 移除方法

移除方法同 單機移除

Master 移除方法

請先確認所有的 Slave 都已經移除 crawlzilla 完畢後,再執行移除程式

crawlzilla_remove

移除方法同 單機移除