

# Day28 - 企業真正的資料來源

前幾天主要在說明資料如何從電子檔案匯入向量資料庫，不過企業最大宗的資料卻都在資料庫，今天就來說說如何從資料庫匯入

## 模擬資料

為了測試方便，資料庫使用嵌入式的 H2，可以省去安裝的時間，專案使用昨天的複製

1. pom.xml 增加以下依賴

```
<dependency>
  <groupId>org.springframework.boot</groupId>
  <artifactId>spring-boot-starter-data-jpa</artifactId>
</dependency>
<dependency>
  <groupId>com.h2database</groupId>
  <artifactId>h2</artifactId>
  <scope>runtime</scope>
</dependency>
```

2. 增加一個 Entity 類別 Issue.java

```
@Entity
@Data
public class Issue {
    @Id
    private Long id;
    private String issue;
    private String solution;
    private String model;
    @Override
    public String toString() {
        return "model: "+model+"\n"
            +"issue: "+issue+"\n"
            +"solution: "+solution;
```

```
}
}
```

### 3. 增加 Repository 介面 IssueRepository.java

```
public interface IssueRepository extends JpaRepository<Issue,
```

### 4. 設定資料庫參數 application.yml

```
#增加以下資料庫設定
h2:
  console:
    enabled: true
datasource:
  url: jdbc:h2:file:./testdb
  driver-class-name: org.h2.Driver
  username: sa
  password:
jpa:
  show-sql: true
  open-in-view: false
  hibernate:
    ddl-auto: update
```

### 5. 啟動專案，讓 JPA 自動建立 ISSUE 表格

```
2024-08-28T00:09:23.304+08:00 INFO 29648 --- [ restartedMain] .s.d.r.c.RepositoryConfigurationDelegate : Finished Spring Data repository
2024-08-28T00:09:24.173+08:00 INFO 29648 --- [ restartedMain] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat initialized with port
2024-08-28T00:09:24.190+08:00 INFO 29648 --- [ restartedMain] o.apache.catalina.core.StandardService : Starting service [Tomcat]
2024-08-28T00:09:24.191+08:00 INFO 29648 --- [ restartedMain] o.apache.catalina.core.StandardEngine : Starting Servlet engine: [
2024-08-28T00:09:24.297+08:00 INFO 29648 --- [ restartedMain] o.a.c.c.C.[Tomcat].[localhost].[/] : Initializing Spring embedded
2024-08-28T00:09:24.297+08:00 INFO 29648 --- [ restartedMain] w.s.c.ServletWebServerApplicationContext : Root WebApplicationContext
2024-08-28T00:09:24.338+08:00 INFO 29648 --- [ restartedMain] com.zaxxer.hikari.HikariDataSource : HikariPool-1 - Starting...
2024-08-28T00:09:24.616+08:00 INFO 29648 --- [ restartedMain] com.zaxxer.hikari.pool.HikariPool : HikariPool-1 - Added connection
2024-08-28T00:09:24.618+08:00 INFO 29648 --- [ restartedMain] com.zaxxer.hikari.HikariDataSource : HikariPool-1 - Start completed
2024-08-28T00:09:24.628+08:00 INFO 29648 --- [ restartedMain] o.s.b.a.h2.H2ConsoleAutoConfiguration : H2 console available at '/
2024-08-28T00:09:24.799+08:00 INFO 29648 --- [ restartedMain] o.hibernate.jpa.internal.util.LogHelper : HHH000204: Processing Persistence
2024-08-28T00:09:24.898+08:00 INFO 29648 --- [ restartedMain] org.hibernate.Version : HHH0000412: Hibernate ORM Core 6.2.10.Final
2024-08-28T00:09:24.956+08:00 INFO 29648 --- [ restartedMain] o.h.c.internal.RegionFactoryInitiator : HHH000026: Second-level cache
2024-08-28T00:09:25.363+08:00 INFO 29648 --- [ restartedMain] o.s.o.j.p.SpringPersistenceUnitInfo : No LoadTimeWeaver setup: using
2024-08-28T00:09:26.597+08:00 INFO 29648 --- [ restartedMain] o.h.e.i.n.j.JtaPlatformInitiator : HHH0000489: No JTA platform
Hibernate create table issue (id bigint not null, issue varchar(255), model varchar(255), solution varchar(255), primary key (id))
2024-08-28T00:09:26.886+08:00 INFO 29648 --- [ restartedMain] j.LocalContainerEntityManagerFactoryBean : Initialized JPA EntityManager
2024-08-28T00:09:26.829+08:00 INFO 29648 --- [ restartedMain] o.neo4j.driver.internal.DriverFactory : Routing driver instance 13
2024-08-28T00:09:27.851+08:00 INFO 29648 --- [ restartedMain] o.s.b.d.a.OptionalLiveReloadServer : LiveReload server is running
2024-08-28T00:09:27.928+08:00 INFO 29648 --- [ restartedMain] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat started on port 8080
2024-08-28T00:09:27.947+08:00 INFO 29648 --- [ restartedMain] com.kt.ai.SpringBootAiDay25Application : Started SpringBootAiDay25Application
```

### 6. 請ChatGPT幫我們產生隨機資料

假設資料庫有 ID, ISSUE, SOLUTION, MODEL 欄位，請隨機幫我生成 20 筆資料  
Model可以放手機型號，資料直接給我 SQL INSERT 語句



以下是包含 `ID`, `ISSUE`, `SOLUTION`, `MODEL` 欄位的 20 筆隨機測試資料的 SQL `INSERT` 語句，問題與解決方案為中文：

```
sql
INSERT INTO issues (ID, ISSUE, SOLUTION, MODEL) VALUES
(1, '螢幕閃爍不定', '重新啟動手機並更新韌體。', 'iPhone 13'),
(2, '電池消耗過快', '關閉背景應用並降低螢幕亮度。', 'Samsung Galaxy S21'),
(3, '相機無法對焦', '清潔鏡頭並重置相機設置。', 'Google Pixel 6'),
(4, '無法連接 Wi-Fi', '忘記該網路，重啟後重新連接。', 'OnePlus 9 Pro'),
(5, '應用程式頻繁崩潰', '清除應用程式快取或重新安裝。', 'iPhone 12'),
(6, '手機過熱', '關閉不必要的應用程式並避免長時間使用。', 'Xiaomi Mi 11'),
(7, '無法接收訊號', '檢查 SIM 卡是否安裝正確或重啟手機。', 'Huawei P40'),
(8, '螢幕觸控失靈', '清潔螢幕並重新校準觸控。', 'Samsung Galaxy Note 20'),
(9, '無法安裝應用程式', '檢查儲存空間並確保網路穩定。', 'Google Pixel 5'),
(10, '通話品質差', '嘗試在訊號較好的區域進行通話。', 'iPhone 11'),
(11, '無法更新系統', '確保電量充足並使用 Wi-Fi 進行更新。', 'OnePlus 8T'),
(12, '藍牙無法配對', '重置藍牙設定並重新嘗試配對。', 'Samsung Galaxy S20'),
(13, '充電速度緩慢', '使用原廠充電器並確保插頭正常。', 'Xiaomi Redmi Note 10'),
(14, '相機閃光燈不工作', '檢查相機設置並重啟手機。', 'Huawei Mate 40'),
(15, '無法同步聯絡人', '確保雲端帳戶設置正確並嘗試重新同步。', 'Google Pixel 4a'),
(16, '螢幕亮度無法調整', '檢查自動亮度設置或重置顯示設置。', 'iPhone SE (2020)'),
(17, 'GPS 定位不準', '檢查 GPS 設置並確保開啟定位服務。', 'Samsung Galaxy A52'),
(18, '手機反應變慢', '清理快取並關閉不必要的背景程序。', 'OnePlus Nord'),
(19, '無法拍攝影片', '檢查儲存空間並重置相機設置。', 'iPhone 12 Pro Max'),
(20, '訊息無法發送', '檢查訊號強度並重啟訊息應用程式。', 'Google Pixel 5a');
```

7. 開啟瀏覽器, 網址輸入 <http://localhost:8080/h2-console>

← → ↻ localhost:8080/h2-console/login.jsp?jsessionid=a 120% A ☆

English Preferences Tools Help

### Login

Saved Settings: Generic H2 (Embedded) ▾

Setting Name: Generic H2 (Embedded) Save Remove

---

Driver Class: org.h2.Driver

JDBC URL: jdbc:h2:file:.\testdb

User Name: sa

Password:

Connect Test Connection

## 8. 登入後新增測試資料

← → ↻ localhost:8080/h2-console/login.do?jsessionid=a 120% A ☆

Auto commit Max rows: 1000 Auto complete Off Auto select On

jdbc:h2:file:.\testdb

Run Run Selected Auto complete Clear SQL statement:

```
(12, '藍牙無法配對', '重置藍牙設定並重新嘗試配對。', 'Samsung Galaxy S20'),
(13, '充電速度緩慢', '使用原廠充電器並確保插頭正常。', 'Xiaomi Redmi Note 10'),
(14, '相機閃光燈不工作', '檢查相機設置並重啟手機。', 'Huawei Mate 40'),
(15, '無法同步聯絡人', '確保雲端帳戶設置正確並嘗試重新同步。', 'Google Pixel 4a'),
(16, '螢幕亮度無法調整', '檢查自動亮度設置或重置顯示設置。', 'iPhone SE (2020)'),
(17, 'GPS 定位不準', '檢查 GPS 設置並確保開啟定位服務。', 'Samsung Galaxy A52'),
(18, '手機反應變慢', '清理快取並關閉不必要的背景程序。', 'OnePlus Nord'),
(19, '無法拍攝影片', '檢查儲存空間並重置相機設置。', 'iPhone 12 Pro Max'),
(20, '訊息無法發送', '檢查訊號強度並重啟訊息應用程式。', 'Google Pixel 5a');
```

INSERT INTO issues (ID, ISSUE, SOLUTION, MODEL) VALUES

```
(1, '螢幕閃爍不定', '重新啟動手機並更新韌體。', 'iPhone 13'),
(2, '電池消耗過快', '關閉背景應用並降低螢幕亮度。', 'Samsung Galaxy S21'),
(3, '相機無法對焦', '清潔鏡頭並重置相機設置。', 'Google Pixel 6'),
(4, '無法連接 Wi-Fi', '忘記該網路，重啟後重新連接。', 'OnePlus 9 Pro'),
(5, '應用程式頻繁崩潰', '清除應用程式快取或重新安裝。', 'iPhone 12'),
(6, '手機過熱', '關閉不必要的應用程序並避免長時間使用。', 'Xiaomi Mi 11')
```

## 9. 接下來增加取得資料以及匯入向量資料庫的 Service 程式 EtlService.java

```

public List<Document> getAllIssue(){
    List<Issue> issues = issueRepository.findAll();
    Map<String, Object> metadata = new HashMap<>();
    metadata.put("type", "issue");
    List<Document> issueDocs = issues.stream().map(
        issue -> new Document(issue.getId().toString(),
                               issue.toString(),
                               metadata)).toList();

    return issueDocs;
}

public void importIssue() {
    vectorStore.write(keywordDocuments(getAllIssue()));
}

```

issueRepository.findAll() 是 JPA 幫我們寫好可以取得所有資料的方法

metadata 使用 map 建立，增加的內容之後可用來篩選  
透過 Lambda 語法將每筆 Issue 轉為 Document 再重組為 List<Document>

寫入向量資料庫前加上 keyword 轉換方法

#### 10. 增加 API 接口 `EtlController.java`

```

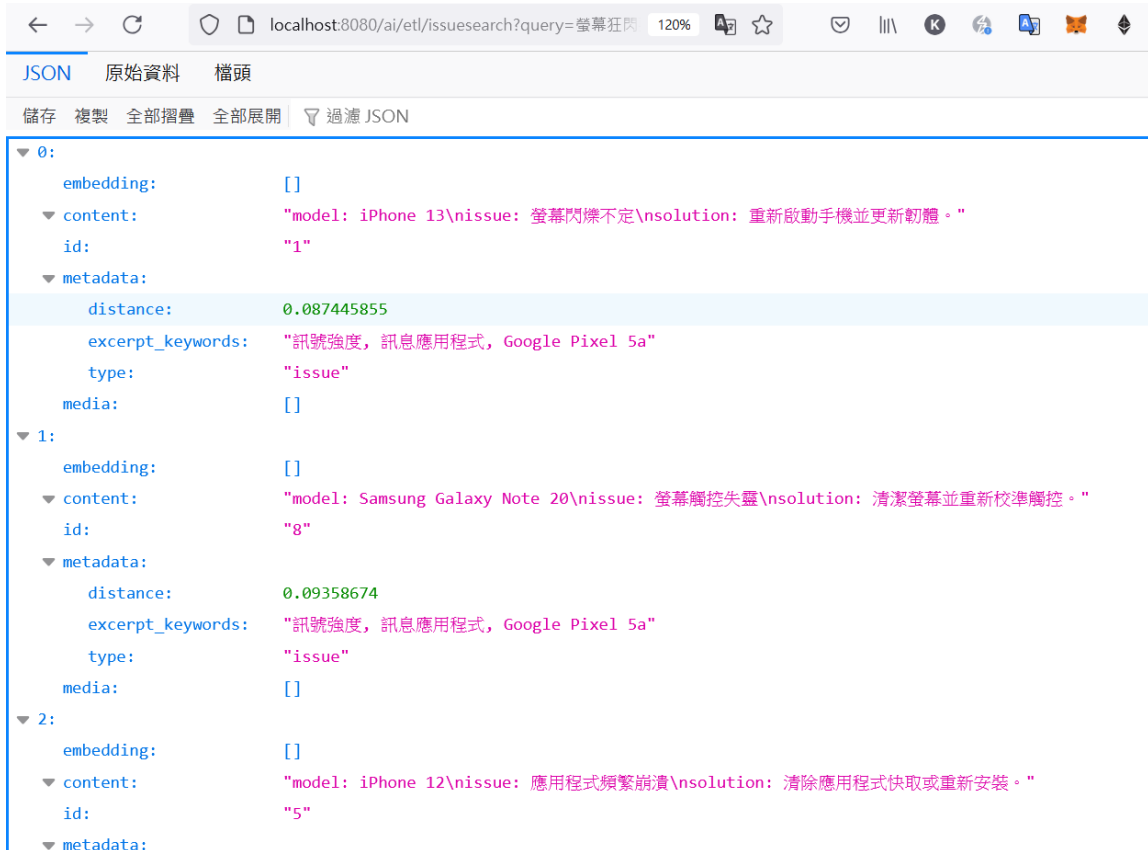
@GetMapping("readissue")
public List<Document> readIssue() throws IOException{
    return etlService.getAllIssue();
}

@GetMapping("importissue")
public String importIssue() throws IOException{
    etlService.importIssue();
    return "OK";
}

@GetMapping("issuesearch")

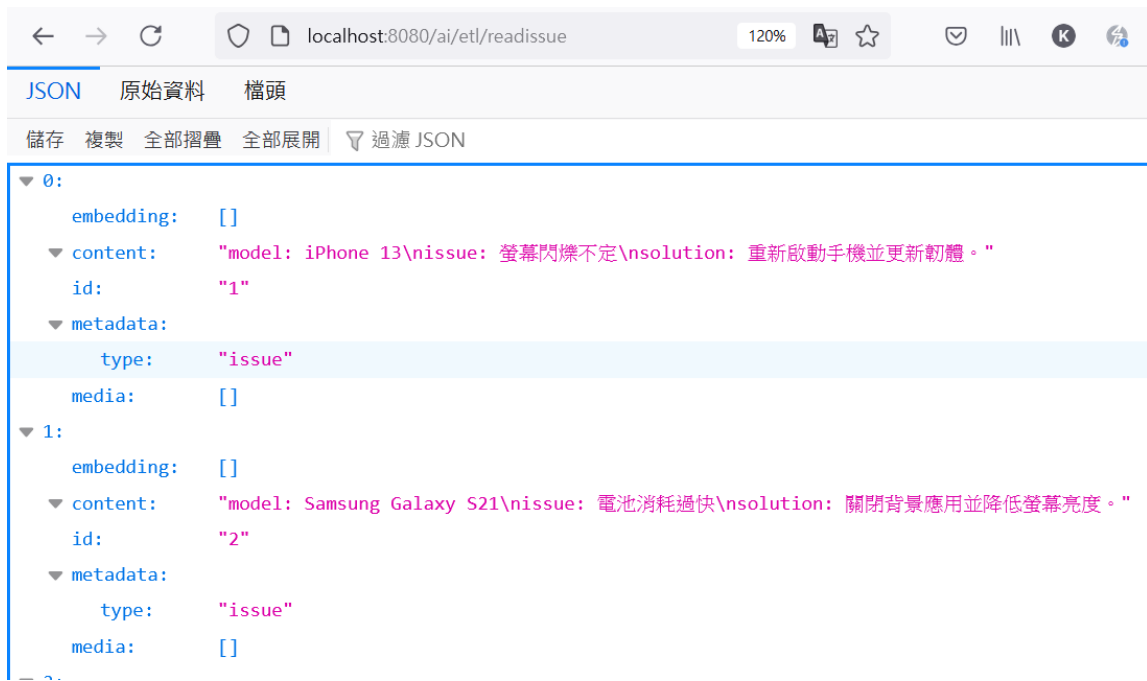
```

```
public List<Document> issueSearch(String query) throws IOException {
    return etlService.issueSearch(query);
}
```

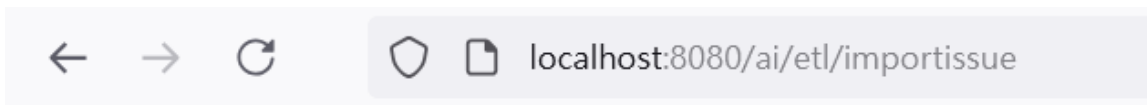


## 驗收程式

網址輸入 <http://localhost:8080/ai/etl/readissue>

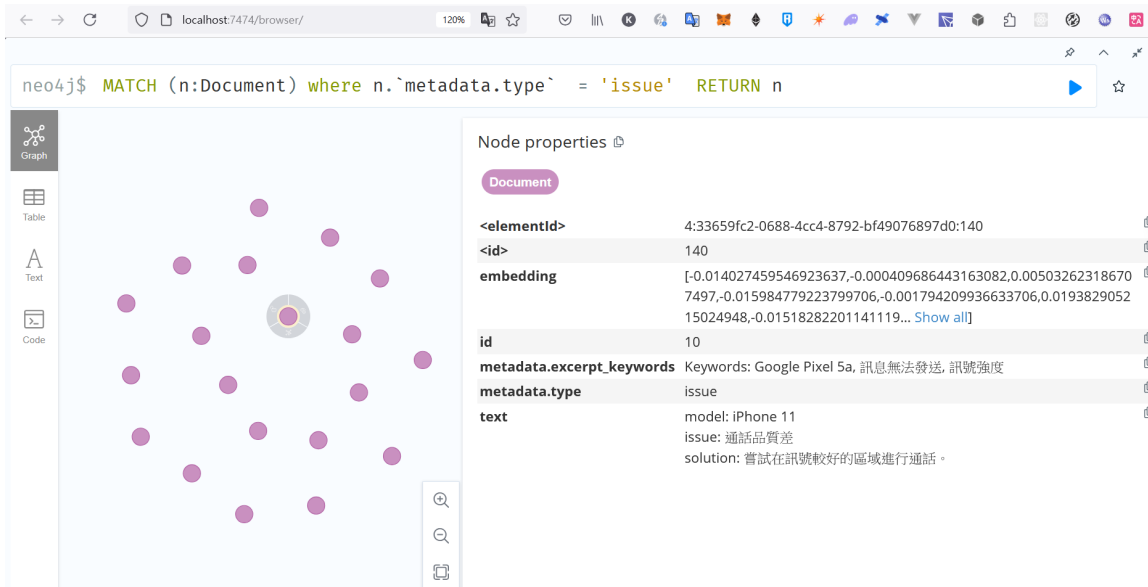


確認資料沒問題後網址輸入 <http://localhost:8080/ai/etl/importissue> 進行匯入工作，出現 OK 表示沒問題了



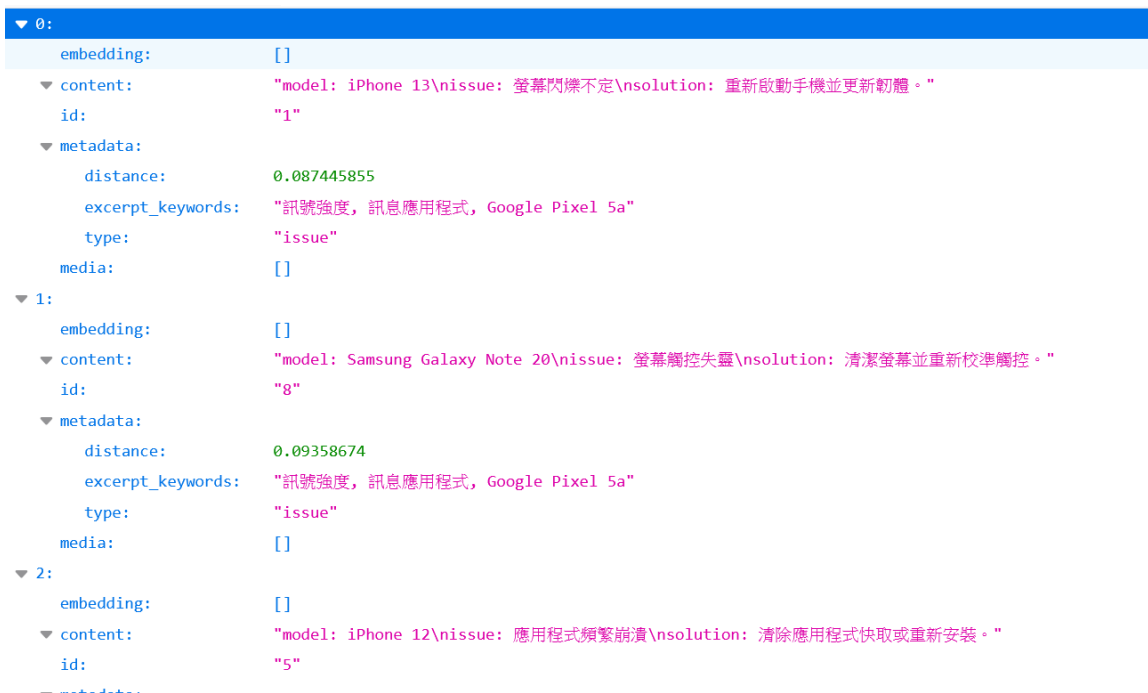
OK

接著進入 Neo4j 確認節點資料 <http://localhost:7474>



除了 text 跟 metadata.excerpt\_keywords，也有建立 Document 時寫入的 metadata ⇒ {"type", "issue"}

接下來網址輸入 <http://localhost:8080/ai/etl/issuesearch?query=螢幕狂閃要如何處理>



第一筆就是