

# (注意喚起) Web データベースの CSV 取り込みで大量のレコードを登録した場合に長時間キーワード検索の対象にならない

2025 年 3 月  
ディサークル株式会社

Web データベースの CSV 取り込みで大量のレコードを登録した場合、キーワード検索用のインデックスが作成されるまでに時間がかかり、キーワード検索の対象になるまでに時間がかかる場合があります。以下の詳細をご確認のうえ、運用変更または設定変更をご検討いただきますようお願い申し上げます。

## 1. 対象製品

POWER EGG3.0 Ver3.3c 以降

## 2. 事象の詳細について

Web データベースのキーワード検索では、レコード登録時に専用の検索サーバ(以後、Solr)にインデックスを作成し高速検索を実現しております。ただし、Solr のインデックス作成処理はシステム負荷が高いため、レコード登録と同時に作成するのではなく、10 秒間隔で最大 10 件までを対象とすることで負荷を抑制しております。また、インデックス作成は全ての Web データベース共通でレコード登録順に実行しております。そのため、CSV 取り込みなどで大量のレコードを登録した場合、大量レコードのインデックス作成が終わるまで、新たに登録・更新された別のレコードはインデックスが作成されず、キーワード検索の対象になるまでに時間がかかります。

たとえば、CSV 取り込みで 1,000 件のレコードを登録した場合、10 秒ごとに 10 件ずつ処理されるため、1,000 秒（16 分 40 秒）以上かかります。この間、処理されていないレコードに関しては、キーワード検索の対象になりません。また、CSV 取り込みを行った Web データベースだけではなく他の Web データベースにおいても、この間に登録されたレコードは 1,000 件のインデックス作成が終わるまでインデックスが作成されず、キーワード検索の対象になりません。

## 3. 対応方法について

上記のように Web データベースの大量レコード登録によりキーワード検索の対象になるまでに時間がかかる場合、以下の 2 つの方法をご検討ください。

### a) CSV 取り込みするレコード数を調整する

CSV 取り込みの際に既存レコードを全削除して全レコードを取り込んでいる場合、インデックス作成に時間がかかります。変更があるレコードのみの CSV データを作成し、既存レコー

ドの全削除を行わずに取り込むようにしてください。

また、運用時間帯は CSV 取り込みを控え、終業時間後や営業日以外に取り込むなど、運用時間帯に大量のインデックス作成が行われないよう実行時間を調整してください。

## b) インデックス作成のパラメタを変更する

以下のファイルのパラメタを変更することで、1回のインデックス作成処理で対象となる最大レコード数を調整することができます。

### ■ ファイル格納先

```
C:¥PE4J¥PESEARCH¥bin¥config¥peindexservice.cfg
```

下線部分は、キーワード検索エンジンインストール先により異なります。

### ■ 変更箇所

```
"Journal.Queue.Size":10,
```

下線部分の数値を変更します。

※ パラメタ変更を反映するには、サービス「POWER EGG SEARCH Index WebDB」の再起動が必要になります。

ただし、変更することでサーバの負荷が上昇します。多くの場合、検索サーバはデータベースサーバと同一サーバで稼働しているため、負荷が上昇すると POWER EGG のレスポンスに影響する可能性があります。リソース状況(主に CPU 使用率)を確認し、POWER EGG のレスポンスなどに影響が出ない範囲で調整してください。

また、10 秒間隔の 1 回の処理でどの程度の時間がかかっているのかをログから確認することができます。

### ■ ログファイル

```
C:¥PE4J¥PESEARCH¥log¥PESearchIndex_webdb.log
```

下線部分は、キーワード検索エンジンインストール先により異なります。

### ■ 出力例

```
2025-02-13 06:01:03,911 INFO Run
2025-02-13 06:01:05,958 INFO [WebDB] BASE:10(1016ms), BASEFILE:0(0ms), COMMENT:0(0ms)
2025-02-13 06:01:06,036 INFO [AppDB]
2025-02-13 06:01:06,724 INFO [CRMDB]
2025-02-13 06:01:13,911 INFO Run
```

ログに「INFO Run」が出力されてから「INFO [CRMDB] ~」が出力されるまでの時間が 1 回の処理時間になります。この時間が 10 秒に近づくと連続して実行することになり、負荷が高くなります。処理数を 10 ずつ増やし、サーバの負荷状況、1 回の処理時間を勘案し、処理数を調整してください。

以上