

日本語対応 Febrl

バージョン 0.4.2

日本語簡易マニュアル

小児慢性特定疾病情報センター

1. はじめに.....	2
2. 動作環境.....	2
3. 免責事項・著作権.....	3
4. お問い合わせ先.....	3
5. インストール・アンインストール.....	4
6. 起動・終了.....	4
6-1. 起動.....	4
6-2. 終了.....	4
7. DEDUPLICATION	4
7-1. 入力ファイルの選択.....	4
7-2. データの確認.....	5
7-3. INDEX の生成.....	6
7-4. COMPARE.....	7
7-5. CLASSIFY	8
7-6. OUTPUT/RUN	9
7-7. EVALUATE.....	9
8. LINKAGE	10
8-1. 入力ファイルの選択.....	10
8-2. データの確認.....	10
8-3. INDEX の生成.....	12
8-4. COMPARE の規定.....	13
8-5. CLASSIFY の規定.....	14
8-6. OUTPUT/RUN の規定.....	14
8-7. EVALUATE.....	15

1. はじめに

「日本語対応 Febrl ver 0.4.2」(以下, 日本語対応 Febrl とします) は, 複数のデータベースを用いた研究において, 個々人のデータの正確な連結を行うためのレコードリンケージ (record linkage) のソフトウェアです. Febrl は, GUI (Graphical User Interface) によるシステムを採用しており, Microsoft Excel のような表計算ソフトと同じような直感的な操作で, 基礎的なデータのクリーニング・標準化から, 最新の高度な連結アルゴリズムまでを利用することができます.

Febrl は, Australian National University のコンピュータ科学部門が開発したフリーソフトウェアであり, 本ホームページでは, 日本語対応した Febrl を公開しています.

2. 動作環境

OS : Windows 7, Windows8, 8.1

メモリ : 1GB 以上推奨

その他 : マニュアルを参照するには Adobe® Reader™が必要です.

Adobe® Reader™はアドビシステムズ社ホームページからダウンロードできます.

アドビシステムズ社ホームページ : <http://www.adobe.com/>

3. 免責事項・著作権

* 日本語対応 Febrl は、Australian National University が開発した無償のオープンソースソフトウェア「Febrl ver 0.4.2」を母体としています。

Febrl : <http://datamining.anu.edu.au/projects/linkage.html>

* 日本語対応 Febrl は、PSF ライセンスで公開された Python 言語を利用しています。

Python : <https://www.python.org/>

* 日本語対応 Febrl は、MIT ライセンスで公開された WinPython ライブラリを利用しています。

WinPython : <http://winpython.sourceforge.net/>

* 日本語対応 Febrl は、LGPL ライセンスで公開された pygtk ライブラリを使用しています。

pygtk : <http://www.pygtk.org/>

* 日本語対応 Febrl は、GPL ライセンスに準じた修正 BSD ライセンスで公開された libsvm ライブラリを利用しています。

libsvm : <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

* 日本語対応 Febrl を使用して起きた障害またはその影響について、小児慢性特定疾病情報センターは責任を負いません。

4. お問い合わせ先

日本語対応 Febrl に関してご不明の点などあれば、下記お問い合わせ先までご連絡ください。本マニュアルに記載している URL および内容は公共データベースの WEB サイトであり、変更されている可能性があります。また、本マニュアルに記載している内容は予告なしに変更される場合があります。

【お問い合わせ先】

大学共同利用機関法人 情報・システム研究機構

統計数理研究所 野間久史

E-mail: noma@ism.ac.jp

HP: <http://www.ism.ac.jp/~noma/>

5. インストール・アンインストール

同ホームページに公開されている、インストールガイドをご参照ください。

6. 起動・終了

6-1. 起動

プログラムメニューの「Febri」フォルダ内の Febri をクリックすると図 6-1 の起動ウィンドウが表示されます。図 6-1 の通り、機能選択のラジオボタン、データ種別選択のラジオボタン、ファイル選択のファイル参照ボタン、処理結果表示のタブが配置されています。

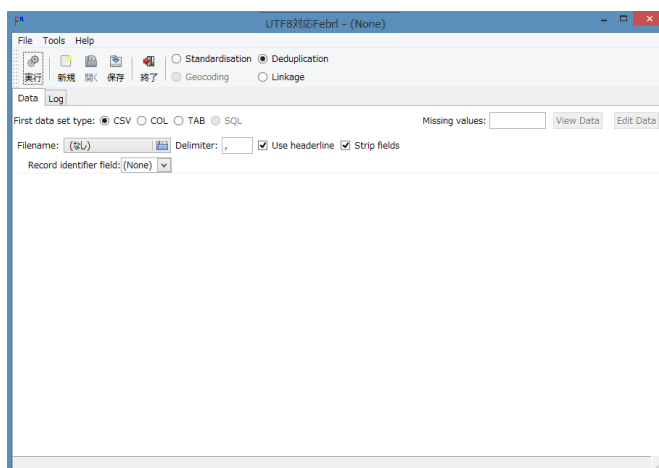


図 6-1. 起動ウィンドウ

6-2. 終了

図 6-1 の File メニューから「終了」をクリックすることで行います。

7. Deduplication

読み込んだデータの重複除外を行う機能です。図 1 で Deduplication のラジオボタンを On にして以下の操作を進めます。

7-1. 入力ファイルの選択

図 1 の Filename のファイル選択ボタンをクリックし、表示されるファイルツリーから入力ファイルを選択すると、図 7-1 のように表示されます。

レコードID	氏名	役職名	住所1	住所2	郵便番号	都道府県	生年月日	年齢	電話番号	社会保険ID	ブロック番号
261	鈴木一郎	主任	足立区神明	南1丁目16-1	121-0051	東京都	19500131	34	03-3605-4711	3905518	3
3321	田中三郎	課長	大田区池上	1-33-8	140-0082	東京都	19930704	37	03-3753-9381	2748326	8
1621	田中博		国分寺市泉町	2丁目7-2	185-0024	東京都	19370724	27	042-322-0211	8731381	7
3311	佐藤明	課長	品川区小山台	1丁目18-24	142-0061	東京都	19662025	37	03-3712-7587	4464331	4
611	佐藤真一	次長	世田谷区池尻	2丁目4-10	154-0041	東京都		33	03-3413-5561	7491079	9
1021	佐藤信一	部長	立川市高松町	1丁目12-35	190-0011	東京都	19080518	35	042-523-5238	7383697	7
4351	山本太郎		千代田区北の丸公園	2-1	102-0091	東京都	19240811	28	03-3212-2440	6678692	1
4651	山本太郎	技師	八王子市元本郷町	3-24-1	192-0051	東京都	19476010		042-558-1111	4750952	4
351	鈴木三郎	副部長	港区東新橋	1丁目5-3	105-0021	東京都			03-3572-1872	2484605	1
721	鈴木史郎	取締役	相模原市緑区長竹	1469	252-0154	神奈川県	19110305	19	042-784-0630	1942604	7
4221	伊藤和美	課長	藤沢市遠藤	641-3	252-0816	神奈川県	19680329	21	0466-87-3251	1887989	5
4621	鈴木美和	部長	三浦市南下浦町	1710	238-0104	神奈川県	19130816		04-881-1042	2793374	1
4881	田中幸代	専務	横浜市長南区野庭町	630	234-0056	神奈川県	19650910	22	045-841-6666	9245372	3
3151	佐藤明子	主任	横浜市長北区大倉山	7丁目2-1	222-0037	神奈川県	19761118	31	045-542-2849	6719149	8
2451	佐藤智代	次長	柏市東柏	2丁目2-15	277-0717	千葉県	19857009	32	04-7167-1298	4334669	3
2641	加藤由美		袖ヶ浦市永吉	419	299-0204	千葉県	19761118	31	0438-60-5511	8189194	6

図 7-1. 入力ファイルの読み込み結果の例

入力ファイルの読み込み終了後に図 1 の「実行」をクリックし、初期計算を行います。

7-2. データの確認

初期計算後新たに追加される「Explore」タブをクリックすると、図 7-2-1 のウィンドウ表示となります。「use sample」にデータ使用率（最大 100）を設定し、実行すると図 7-2-2 のウィンドウ表示となります。「実行」をクリックするとデータの各列ごとに文字列長、重複チェックを行い、結果をウィンドウ表示します。

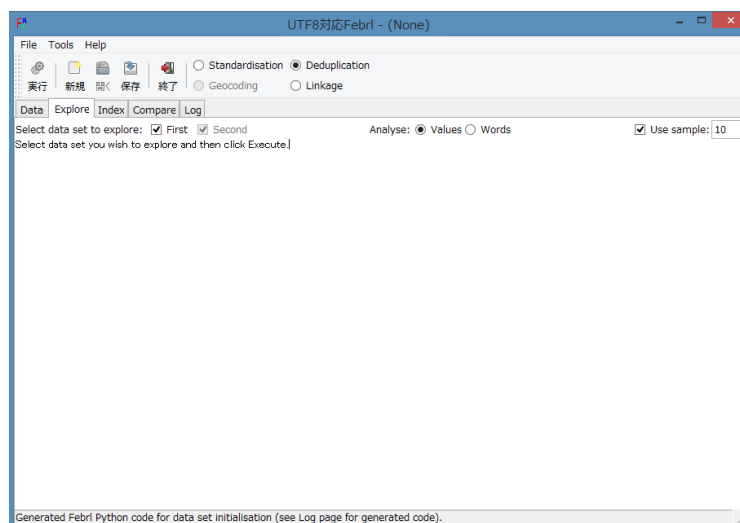


図 7-2-1. Explorer タブ表示内容（実行前）

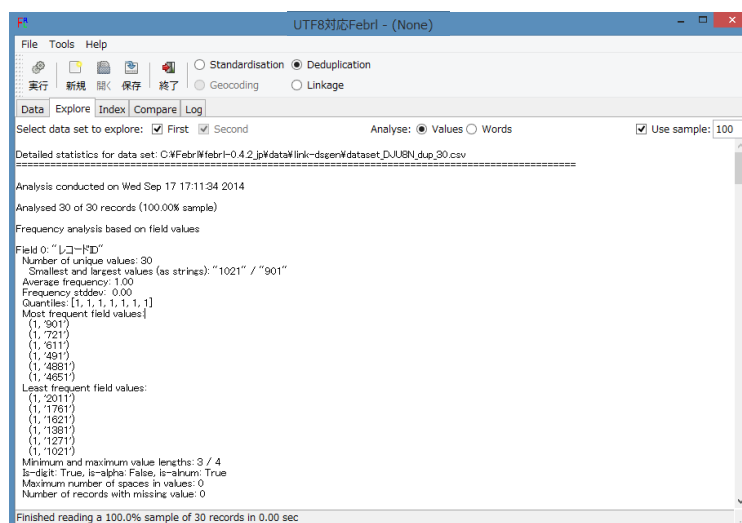


図 7-2-2. Explorer タブ表示内容（実行後）

7-3. Index の生成

初期計算後新たに追加される「Index」タブをクリックすると、図 7-3-1 のウィンドウ表示となります。「Add new index」をクリックすることで、図 7-3-2 のように生成する Index を規定でき、「実行」をクリックすることで Index を生成します。

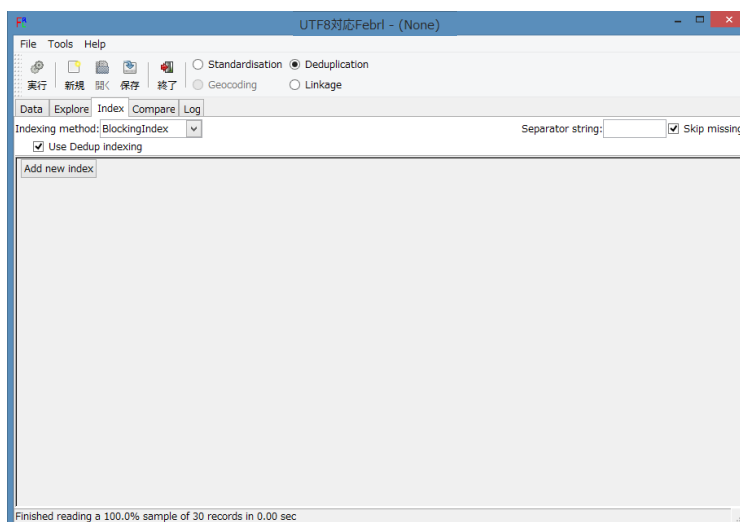


図 7-3-2. Index タブ表示内容（「Add new index」実行前）

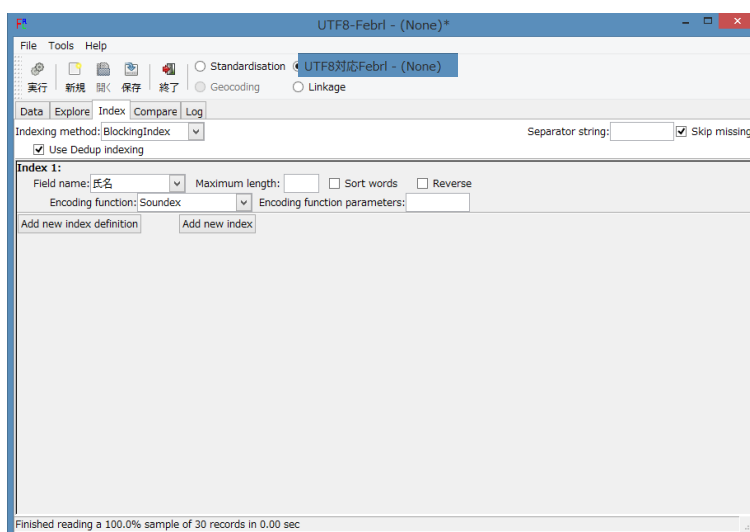


図 7-3-2. Index タブ表示内容（「Add new index」実行後）

7-4. Compare

初期計算後新たに追加される「Compare」タブをクリックすると、図 7-4-1 のウィンドウ表示となります。「Add new comparison function」をクリックすることで、図 7-4-2 のように Compare の規定を行うことができ、「実行」をクリックすることで規定した Compare を実行します。

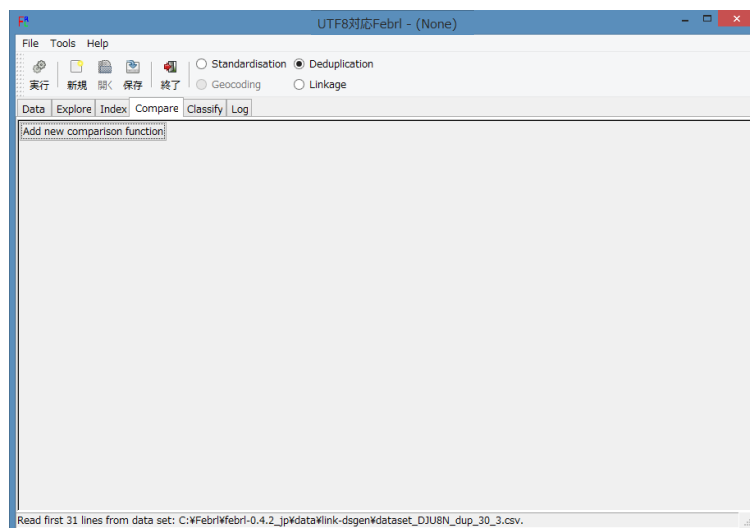


図 7-4-1. Compare タブ表示内容（「Add new comparison function」実行前）

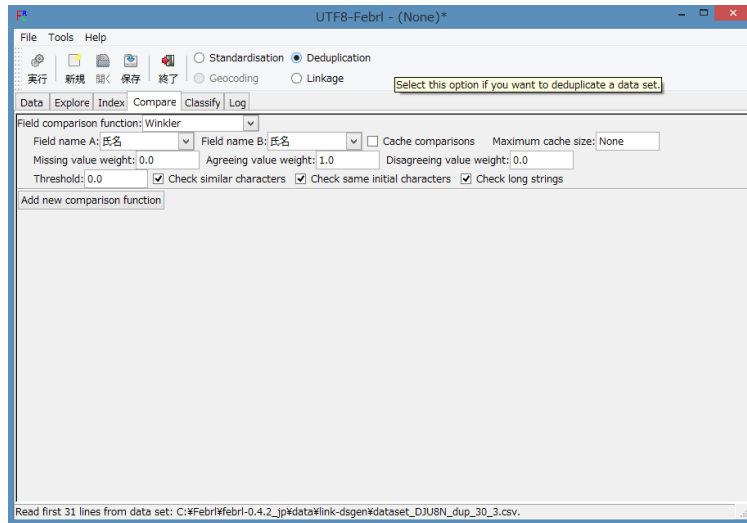


図 7-4-2. Compare タブ表示内容（「Add new comparison function」実行後）

7-5. Classify

Compare タブの実行後新たに追加される「Classify」タブをクリックすると、図 7-5-1 のウィンドウ表示となります。Weight vector Classification method の選択と maximum iteration count の設定を行い「実行」をクリックします。

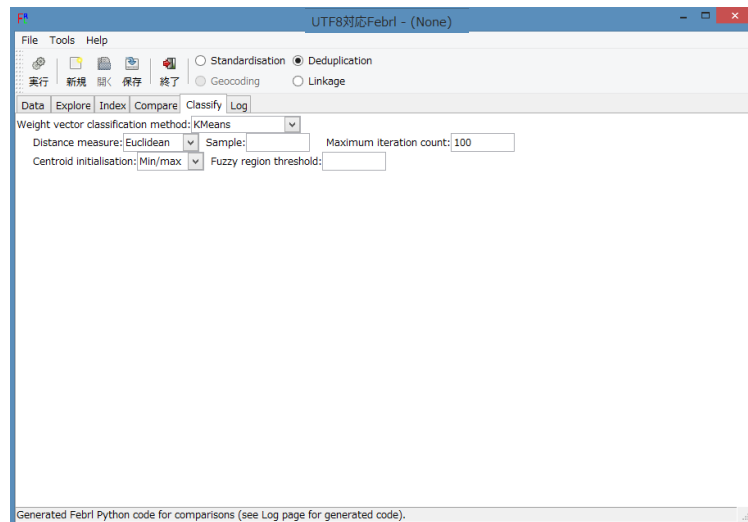


図 7-5-1 Classify タブ表示内容

7-6. Output/Run

Classify タブの実行後に追加される Output/Run タブをクリックすると図 7-6-1 のウィンドウ表示となります。Output/Run タブでは出力ファイルの設定を行い「実行」をクリックします。

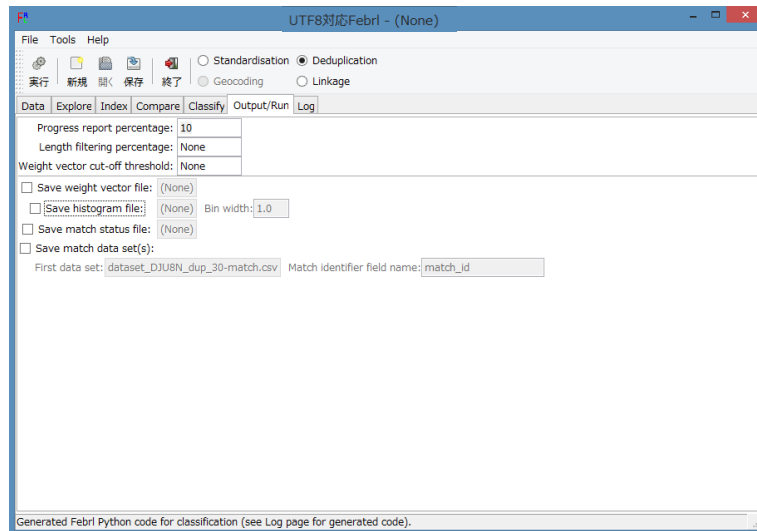


図 7-6-1 Output/Run タブ表示内容

7-7. Evaluate

Output/Run タブの実行後に追加される Evaluate タブをクリックする図 7-7-1 のウィンドウ表示となります。ウィンドウ内に表示されるヒストグラムを確認することで、入力ファイルの品質、重複状況の把握が行えます。

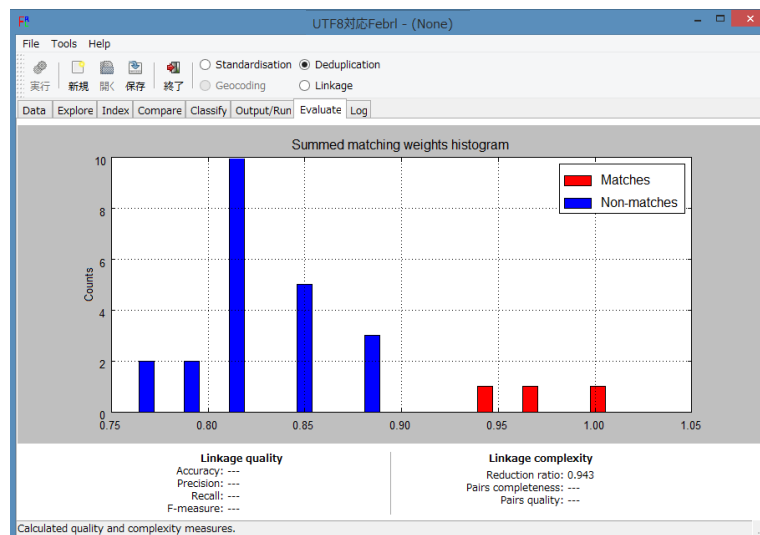


図 7-7-1 Evaluate タブ表示内容

8. Linkage

2つのデータの Linkage を行う機能です。図 1 で Linkage のラジオボタンを On にして以下の操作を進めます。

8-1. 入力ファイルの選択

図 1 の Filename のファイル選択ボタンをクリックし、表示されるファイルツリーから入力ファイルを選択すると、図 8-1 のように表示されます。

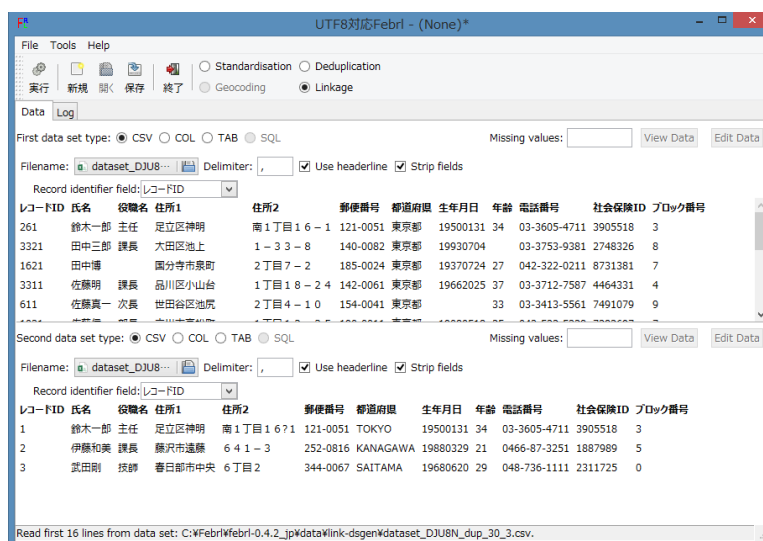


図 8-1 Data タブ表示内容

入力ファイルの読み込み終了後に図 1 の「実行」をクリックし、初期計算を行います。

8-2. データの確認

初期計算後新たに追加される「Explore」タブをクリックすると、図 7-2-1 のウィンドウ表示となります。「use sample」にデータ使用率（最大 100）を設定し、実行すると図 7-2-2 のウィンドウ表示となります。「実行」をクリックするとデータの各列ごとに文字列長、重複チェックを行い、結果をウィンドウ表示します。

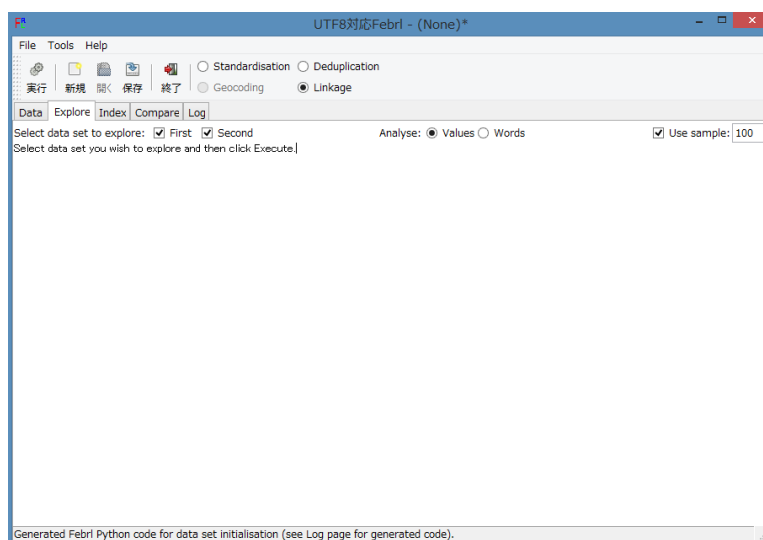


図 8-2-1. Explorer タブ表示内容 (実行前)

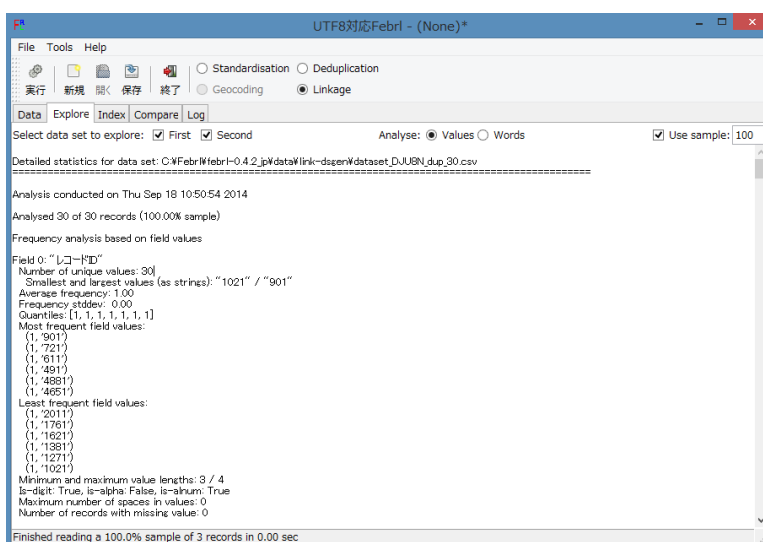


図 8-2-2. Explorer タブ表示内容 (実行後)

8-3. Index の生成

初期計算後新たに追加される「Index」タブをクリックすると、図 8-3-1 のウィンドウ表示となります。「Add new index」をクリックすることで、図 8-3-2 のように生成する Index を規定でき、「実行」をクリックすることで Index を生成します。

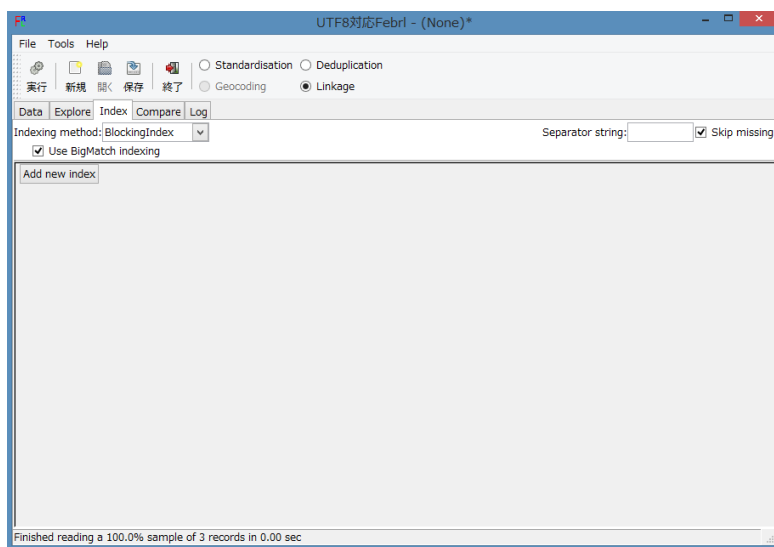


図 8-3-1. Index タブ表示内容（「Add new index」実行前）

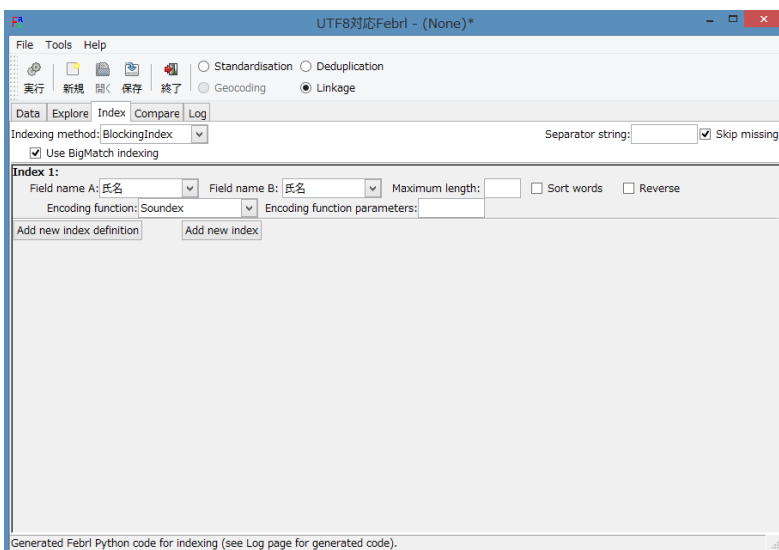


図 8-3-2. Index タブ表示内容（「Add new index」実行後）

8-4. Compare

初期計算後新たに追加される「Compare」タブをクリックすると、図 8-4-1 のウィンドウ表示となります。「Add new comparison function」をクリックすることで、図 8-4-2 のように Compare の規定を行うことができ、「実行」をクリックすることで規定した Compare を実行します。

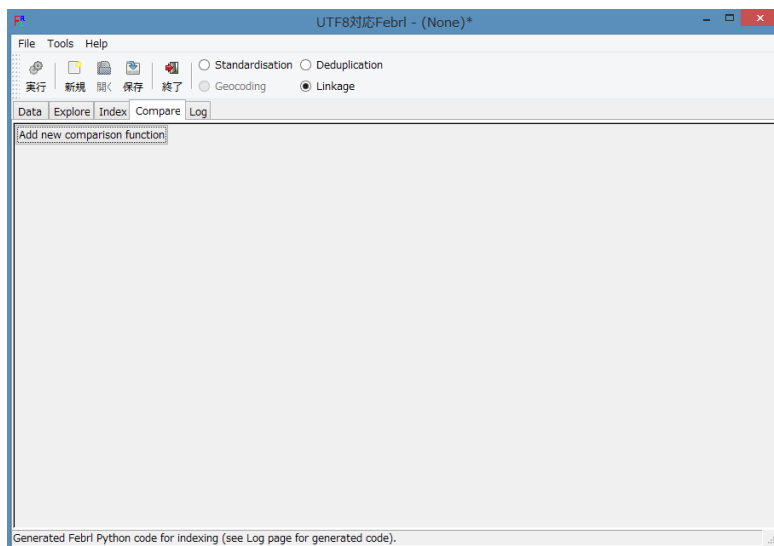


図 8-4-1. Compare タブ表示内容（「Add new comparison function」実行前）

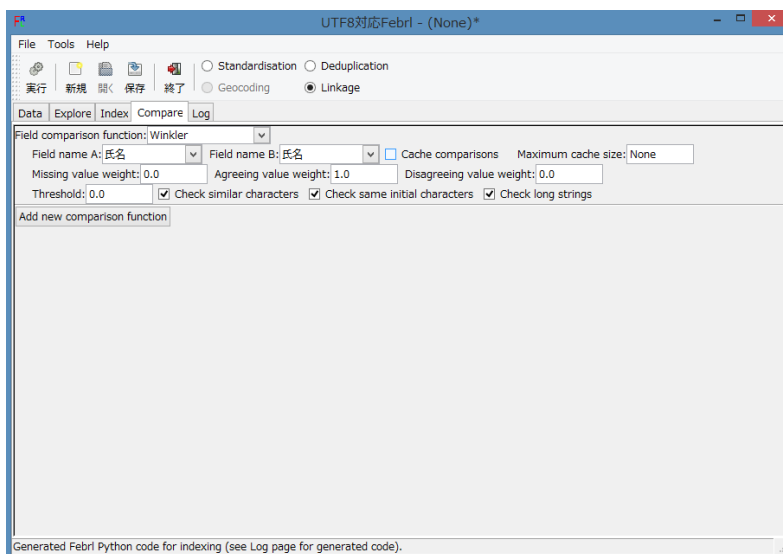


図 8-4-2. Compare タブ表示内容（「Add new comparison function」実行後）

8-5. Classify

Comapre タブの実行後新たに追加される「Classify」タブをクリックすると、図 8-5-1 のウィンドウ表示となります。Weight vector Classification method の選択と maximum iteration count の設定を行い「実行」をクリックします。

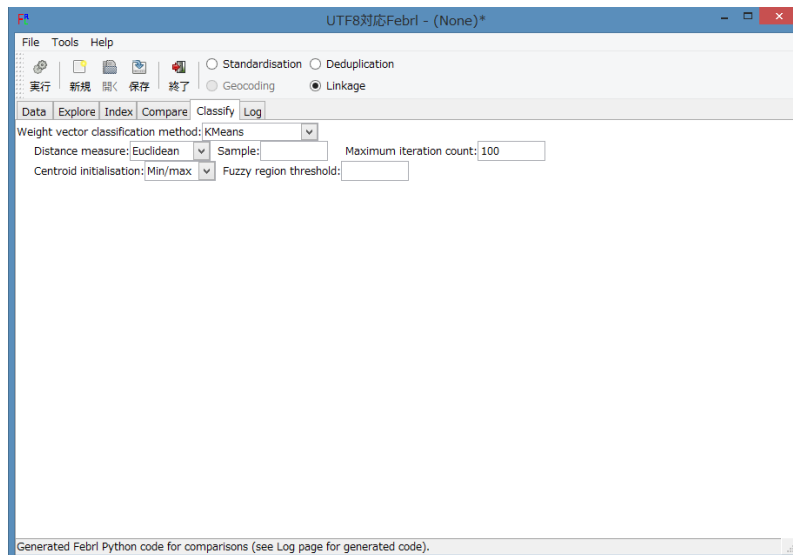


図 8-5-1 Classify タブ表示内容

8-6. Output/Run

Classify タブの実行後に追加される Output/Run タブをクリックすると図 8-6-1 のウィンドウ表示となります。Output/Run タブでは出力ファイルの設定を行い「実行」をクリックします。

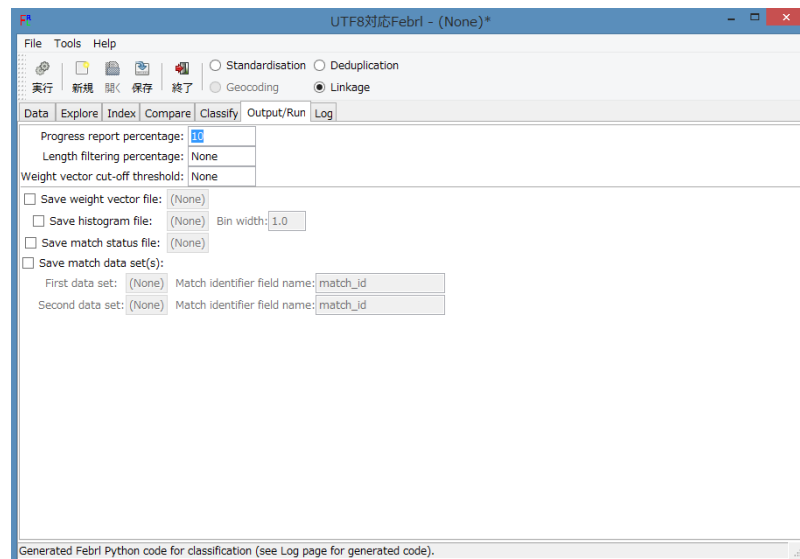


図 8-6-1 Output/Run タブ表示内容

8-7. Evaluate

Output/Run タブの実行後に追加される Evaluate タブをクリックする図 8-7-1 のウィンドウ表示となります。ウィンドウ内に表示されるヒストグラムを確認することで、Linkag 状況の把握が行えます。

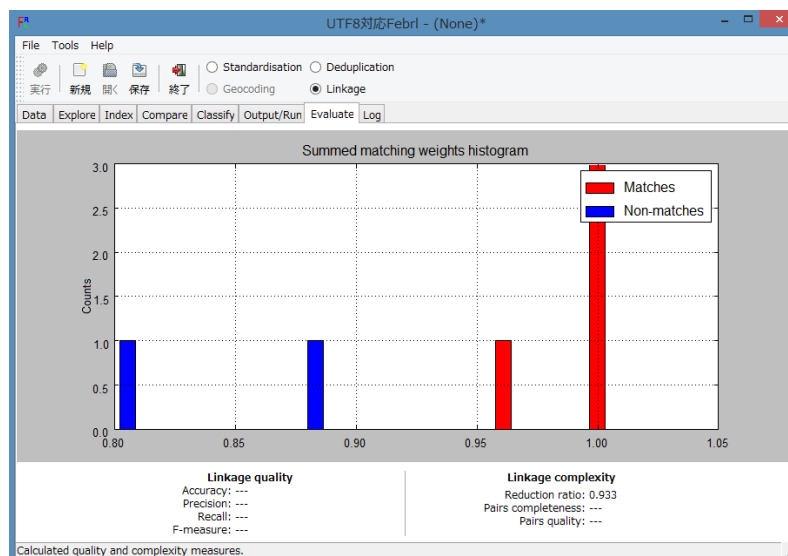


図 8-7-1 Evaluate タブ表示内容