

“HSORT”『超高速ソートマージ プログラム v2.0』 の使い方

[HSORT の概要](#)

HSORT は、C 言語で開発され、自作メモリプールを使用して、超高速で固定長、可変長、CSV ファイルのソート、マージ機能を提供できるコマンドライン型ソフトである。

- ・ **ソート機能**とは、ファイル内のレコードをデータ中の文字または数字を**キー**として、昇順または降順に並べ替える機能です。
- ・ **マージ機能**とは、複数のファイルのデータを一つのファイルに併合する機能です。

[ソート方式](#)

ソート機能の処理には、入力ファイルの総サイズ及び指定されたメモリのサイズにより次の二つの方式があります。

・メモリ上ソート

入力データ量に対して十分なメモリサイズを指定した場合、一時ファイルを使用しないでソートする処理のことです。

・一時ファイルでソート

入力データ量に対して最大メモリサイズが不足した場合、一時ファイルを使用して、入力データを分割しながらソートする処理のことです。

※使用可能なメモリサイズが未指定の場合、入力ファイルの総サイズを基づいて「メモリ上ソート」の必要なメモリサイズを自動計算してメモリを割り当ててみます。メモリは確保できなければ、「一時ファイルでソート」する。

[機能一覧](#)

- ・ ソフト起動の際にメモリプール用メモリを確保し、高速にメモリを割り当て、メモリ断片化を最小化すること
- ・ 最大メモリサイズが指定できること、未指定の場合、自動計算すること
- ・ 固定長、可変長のテキスト、バイナリファイルをサポートすること
- ・ CSV ファイルをサポートすること
 - ※文字列の記述には、ダブルクォーテーション (") で囲むことは可能
- ・ 標準入力、標準出力をサポートすること
- ・ ログ情報を標準エラーに出力すること
- ・ 複数キー又は全レコードの昇順、降順が指定できること
- ・ 等しいソートキーの最初レコードのみ出力すること（ユニック出力機能）
- ・ レコード順序保持機能（安定ソート）
- ・ 不正レコード（最大キー不足）をスキップ、又はエラーファイルに出力すること
- ・ ASCII 順、数字順でソートすること
- ・ 2 G 以上ファイルもサポートすること

パラメータ一覧

パラメータ	意味
/?	ヘルプ
/V	バージョン情報、ライセンス情報
/C	入力データは CSV ファイルです。
/S	安定ソートフラグを指定します。
/U	ユニーク出力フラグを指定します。
/L レコード長	固定長順ファイルを指定します。 入力ファイル編成ごとのレコード長 ※ファイルのレコード長を「1 ～ 640KB」バイトの範囲で指定します。
/K キー開始位置,キー長[n] [a d] ※固定長、可変長の場合	ソートキーの開始位置、長さを指定します。 ※a(昇順、省略可)、d(降順) ※複数指定可、更に各キーの位置は重なることは可能 ※n は数字順でソートします。(未指定すると、ASCII 順となる)
/K キー位置[n][a d] ※CSVの場合	ソートキーの位置を指定する。 ※a(昇順、省略可)、d(降順) ※複数指定可 ※n は数字順でソートします。(未指定すると、ASCII 順となる)
/A	全レコード昇順フラグを指定します。 ※-K と同時に指定すると、-K オプションを省略します。 ※-R と同時に指定すると、後に指定されたオプションを省略します。
/R	全レコード降順フラグを指定します。 ※-K と同時に指定すると、-K オプションを省略します。 ※-A と同時に指定すると、後に指定されたオプションを省略します。
/P 改行コード	CSV ファイルの改行コードを指定します。 ※未指定の場合、CrLf(Window 版)、Lf(Linux 版)を使用します。
/D CSV ファイルの区切り文字	CSV ファイルのデリミタを指定します。(複数文字指定可) ※未指定の場合、カンマを使用します。
/W 最大メモリサイズ	使用できる最大メモリサイズを指定します。 ※未指定の場合、デフォルト値 64MB を使用します。 ※常に最小値16MB で制限されます。 ※メモリ量の単位が、K(KB)、M(MB)が指定できます。
/T ディレクトリ	ソート用一時ファイルのディレクトリを指定します。 ※未指定の場合、環境変数「TEMP」に設定されているディレクトリを使用します。環境変数「TEMP」も設定されていない時、エラーとなります。 ※一時ファイル名は hsort_プロセス id.tmp となります。 ※ソート後、一時ファイルを自動的に削除します。
/O 出力ファイル	出力先ファイルのパスを指定します。 ※出力先ファイルを入力ファイルと同じファイルの指定もできます。 ※未指定の場合、標準出力となること
/E エラーファイル	エラーファイルのパスを指定します。 ※未指定の場合、不正レコードがスキップされ、出力されない。
入力ファイル	入力ファイルのパスを指定します。 ※複数指定可、更に全て引数の最後に指定しなければならない。

使用例

I、CSV ファイルのソート

- ① デリミタ（カンマ）、最大メモリサイズ 100MB、全レコード昇順、標準入力、標準出力

```
hsort /C /W100M < C:¥in.csv
```

- ② 改行コード（Lf）、デリミタ（Tab）、複数キー、昇順、降順、出力ファイル、入力ファイル

```
hsort /C /P¥n /D¥t /K1a /K3d /OC:¥out.csv C:¥in1.csv C:¥in2.csv
```

※エラーファイルが未指定のため、列数が3列不足のレコードがエラーレコードとして出力されない

- ③ ユニーク出力、全レコード降順、一時ファイル、出力ファイル、エラーファイル、入力ファイル

```
hsort /C /U /R /TC:¥ /OC:¥out.csv /EC:¥err.csv C:¥in.csv
```

※エラーレコードがエラーファイル *err.csv* に出力される

- ④ 安定ソート、1 列目 ASCII 昇順、3 列目 数字で降順

```
hsort /C /S /K1 /K3nd /OC:¥out.csv C:¥in.csv
```

II、固定長ファイルのソート

- ① レコード長 64、最大メモリサイズ 100MB、全レコード昇順、標準入力、標準出力

```
hsort /L64 /W100M < C:¥in.dat
```

- ② レコード長 64、複数キー、昇順、数字で降順、出力ファイル、入力ファイル

```
hsort /L64 /K0,5a /K3,8nd /OC:¥out.dat C:¥in1.dat C:¥in2.dat
```

※エラーファイルが未指定のため、列数が3列不足のレコードがエラーレコードとして出力されない

- ③ レコード長 64、ユニーク出力、全レコード降順、一時ファイル、出力ファイル、エラーファイル、入力ファイル

```
hsort /L64 /U /R /TC:¥ /OC:¥out.dat /EC:¥err.dat C:¥in.dat
```

※エラーレコードがエラーファイル *err.dat* に出力される

- ④ レコード長 64、安定ソート、複数キー、降順、昇順、出力ファイル、入力ファイル

```
hsort /L64 /S /K1,3d /K6,9a /OC:¥out.dat C:¥in.dat
```

III、可変長ファイルのソート

- ① 改行コード（CrLf）、最大メモリサイズ 100MB、全レコード昇順、標準入力、標準出力

```
hsort /P¥r¥n /W100M < C:¥in.dat
```

- ② 複数キー、昇順、数字で降順、出力ファイル、入力ファイル

hsort /K5,8a /K0,3nd /OC:¥out.dat C:¥in1.dat C:¥in2.dat

※エラーファイルが未指定のため、列数が3列不足のレコードがエラーレコードとして出力されない

- ③ ユニーク出力、全レコード降順、一時ファイル、出力ファイル、エラーファイル、入力ファイル

hsort /U /R /TC:¥ /OC:¥out.dat /EC:¥err.dat C:¥in.dat

※エラーレコードがエラーファイル *err.dat* に出力される

- ④ 安定ソート、複数キー、昇順、降順、出力ファイル、入力ファイル

hsort /S /K1,3 /K5,9d /OC:¥out.dat C:¥in.dat

注意事項

- 1、入力ファイルのフォーマットが指定されないと、可変長として処理します。
- 2、パラメータは大文字、小文字を区別しません、下記の指定方法は同じとなります。
/C /c -C -c
- 3、可変長、CSV ファイルのデフォルトな改行コードは **LF** となります。
- 4、入力ファイルは最後にならなくてははいけませんが、他の各パラメタ順序は任意です。

連絡方式

不具合を発見したら、下記メール宛に連絡していただいたら、幸いです。

soft@ohji-trade.jp 暁星