

Dynamic Enterprise AI Solution

<AIRead White Paper>
AIRead
個別読取結果変換の設定例

2025年7月
(AWP-006-10)

 **Arise Innovation, Inc.**

AIとローコード開発を活用し
世の中をより良い世界に

1. はじめに
2. 個別読取結果変換とは
3. 正規表現とは
4. 使用例

1. はじめに

- 本書はAIReadの帳票定義における「正規表現」を使った個別読取結果変換について記載した文書です。
- 正規表現とはAIReadの中だけで使用されるものではなく、一般的に利用されている表現方法です。
- 帳票上の記載や誤読の状況によっては、そのまま利用することのできないパターンもございますので、WEB上の正規表現チェッカー等を利用し、正規表現の調整を行ってください。

2. 個別読取結果変換とは

- 個別読取結果変換とは、読み取った帳票の特定の文字を任意の文字に変換するための処理です。AIReadの帳票定義における後処理設定として AIRead ControlPanelで設定します。
- 設定手順：
 1. AIRead ControlPanelの後処理設定タブを開く（帳票設定・メタデータ設定画面の設定ボタンからも開けます）
 2. 変換リスト名を入力し+ボタンを押す
 3. 「変換対象」に変換したい文字を入力（※ここで正規表現を使用します）
 4. 「変換後の文字」に変換対象の文字から変換する文字を入力
 5. 帳票定義の「個別読取結果変換」で、プルダウンから反映したい変換リストを選択する



3. 正規表現とは (1/7)

- 正規表現とは、文字列の検索や置換をする際に文字や文字列の組み合わせを一つの文字列で表すことができる表現方法です。
- 正規表現には「メタ文字」という特別な意味を持った文字があります。メタ文字は全て半角の記号です。

| メタ文字 | 意味 |
|------|-----------------------------------|
| . | いずれかの1文字を表現します。 |
| * | 直前の文字の0回以上の繰り返しを表します。 |
| + | 直前の文字の1回以上の繰り返しを表します。 |
| ? | 直前の文字の0回、または1回の出現を表します。 |
| ^ | 正規表現の先頭を表します。 |
| \$ | 正規表現の終端を表します。 |
| [] | 中に含まれている文字のいずれかの1文字を表します。 |
| {n} | 直前の文字のn回の繰り返しを表します。※nには半角数字が入ります。 |
| () | 中に含まれている文字をグループ化します。 |
| | 複数のパターンのうちどれか1つと一致することを表します。 |
| ¥ | メタ文字のエスケープを表します。 |

3. 正規表現とは (2/7)

■ . : いずれかの1文字

「私は が好きです。」に挟まる1文字が対象です。

| | |
|---------------|--|
| 正規表現 | 私は. が 好きです。 |
| 検索できる文字列のパターン | 私は 海 が好きです。 私は 山 が好きです。 |

■ * : 直前の文字の0回以上の繰り返し

「私は 好きです。」に何文字挟まれていても、1文字も挟まれていなくても対象になります。

| | |
|---------------|--|
| 正規表現 | 私は.*好きです。 |
| 検索できる文字列のパターン | 私は うどん が好きです。 私は そば が好きです。 私は好きです。 |

■ + : 直前の文字の1回以上の繰り返し

「私は が好きです。」に挟まれた、1文字以上が対象になります。

| | |
|---------------|---|
| 正規表現 | 私は.+が好きです。 |
| 検索できる文字列のパターン | 私は うどん が好きです。 私は そば が好きです。 私は 麵 が好きです。 |

3. 正規表現とは (3/7)

■ ? : 直前の文字の0回、または1回の出現

httpsの?の直前のsが入っていてもいなくても対象となります。

| | |
|---------------|---------------|
| 正規表現 | https? |
| 検索できる文字列のパターン | http https |

■ ^ : 正規表現の先頭

「こんにちは」から始まる文字列が対象です。

| | |
|---------------|-----------------|
| 正規表現 | ^こんにちは.* |
| 検索できる文字列のパターン | こんにちは、良いお天気ですね。 |

■ \$: 正規表現の終端

「さようなら。」で終わる文字列が対象です。

| | |
|---------------|-----------------|
| 正規表現 | .*さようなら。\$ |
| 検索できる文字列のパターン | また会いましょう、さようなら。 |

3. 正規表現とは (4/7)

■ [] : 中に含まれているいずれかの1文字

「私は 歳です。」に挟まれる1文字以上の数字が対象です。

| | |
|---------------|---------------------------------------|
| 正規表現 | 私は [0-9]+ 歳です。 |
| 検索できる文字列のパターン | 私は 40 歳です。 私は 3 歳です。 |

■ {n} : 直前の文字のn回の繰り返し

「私の誕生日は、です」に挟まれる数字4桁年、数字1~2桁月、数字1~2桁日 が対象です。

| | |
|---------------|--|
| 正規表現 | 私の誕生日は、 [0-9]{4}年[0-9]{1,2}月[0-9]{1,2}日 です。 |
| 検索できる文字列のパターン | 私の誕生日は、 2000年3月14日 です。 私の誕生日は、 1950年12月2日 です。 |

■ () : 中に含まれている文字列のグループ化

「ha」がグループ化の対象です。

| | |
|---------------|--------------------------------|
| 正規表現 | a (ha)+ |
| 検索できる文字列のパターン | a ha a hahaha |

3. 正規表現とは (5/7)

■ | : 複数のパターンのうちどれか1つ

「私は が好きです。」に挟まれる犬か猫か鳥のいずれか1つが対象です。

| | |
|---------------|--|
| 正規表現 | 私は (犬 猫 鳥) が好きです。 |
| 検索できる文字列のパターン | 私は 犬 が好きです。 私は 猫 が好きです。 私は 鳥 が好きです。 |

■ ¥ : メタ文字のエスケープ

「¥」でエスケープしたメタ文字が対象です。

| | |
|---------------|-----------|
| 正規表現 | ¥[|
| 検索できる文字列のパターン | [|

| | |
|---------------|-----------|
| 正規表現 | ¥¥ |
| 検索できる文字列のパターン | ¥ |

| | |
|---------------|-----------|
| 正規表現 | ¥(|
| 検索できる文字列のパターン | (|

3. 正規表現とは (6/7)

■ 数字のみで構成される文字列を対象とした正規表現

```
^[0-9]+$
```

^は正規表現の先頭、\$は終端を表します。

[]は、中に入っている文字のいずれか1文字に一致しているという表現になります。

0-9は、0～9の数字という表現の方法です。

+は、[0-9]で一致した0～9までのいずれかの文字を1回以上繰り返すという表現です。

上記を組み合わせることで、先頭から終端まで、0～9の数字が1文字以上の文字列にのみ一致する表現となります。

例：123、1357、5、99999…など

■ 数字と記号で構成される文字列を対象とした正規表現

```
^[0-9,¥.¥-]+$
```

数字以外に対象としたい文字がある場合は、[]に対象とした記号を含めた表現にします。

.(ドット)や-(ハイフン)等のメタ文字を対象とする場合は、ただの文字にするために文字の前に¥を付与します。

よって上記は、0～9の数字・,(カンマ)・.(ドット)・-(ハイフン)のいずれかの文字が1文字以上の文字列に一致する表現となります。

例：123,456、-9876、3、12345、3.14159…など

3. 正規表現とは (7/7)

■ 文字数を指定した正規表現

```
^[0-9]{3}$
```

[]の後に、{}と数字を入力すると、{}の中の数字の文字数で繰り返しの回数が固定されます。

よって上記は、3桁の数字で構成されている文字列に一致する表現となります。

例：123、000、080、246 など

```
^[0-9]{3,5}$
```

上記の様に{}の中の数字を,(カンマ)区切りで2つ記載すると、繰り返しの回数を範囲指定することができます。

よって上記は、3～5桁の数字で構成されている文字列に一致する表現となります。

例：123、2467、00090 など

■ その他の範囲指定例

| 範囲 | 意味 |
|-------------|------------------|
| [a-z] | 小文字アルファベット |
| [A-Z] | 大文字アルファベット |
| [ぁ-ん] | ひらがな (全角) |
| [A-Za-z0-9] | 英数字 (大文字・小文字・数字) |

4. 使用例 (1/3)

■ 金額の値を数値のみにする

下記の変換リストは数値のみの値にする際に活用する個別読取結果変換リストの例です。

変換リスト名
number + -

number

項目名の修正

住所の修正

西暦和暦変換

単位の修正

店名の修正

変換リスト

| 変換対象(正規表現) | 変換後の文字 |
|----------------------|--------|
| [UuOoCc(€@)] | 0 |
| ¥[0-9]{1,10}[¥ !¥ :] | 1 |
| [Zz] | 2 |
| fi [fbEG] | 6 |
| [SsB] | 8 |
| [gq] | 9 |
| [^0-9] | |
| | |
| | |

- ・ 似た文字などに誤読した場合でも数字に変換するようにします。(例：1Z30U→12300)
- ・ [^0-9]の正規表現で数字以外の文字を除去します。(例：¥12,300→12300)

4. 使用例 (2/3)

■ 日付のフォーマット統一

個別読取結果変換リストで日付のフォーマットを統一することが可能です。

下記の変換リストは年は西暦に、日付を数字で統一する際に活用する個別読取結果変換リストの例です。

| 変換リスト名 | 変換リスト | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------|--|------------|--------|-----------------|------|-----------------|------|-----------------|------|-----------------|------|-----------------|------|----------------|------|----------------|------|----------------|------|-----------------|------|---------------|------|--------|--|
| 西暦和暦変換 | <table border="1"> <thead> <tr> <th>変換対象(正規表現)</th> <th>変換後の文字</th> </tr> </thead> <tbody> <tr> <td>(?:令和 ^R ^)^075</td> <td>2023</td> </tr> <tr> <td>(?:令和 ^R ^)^076</td> <td>2024</td> </tr> <tr> <td>(?:令和 ^R ^)^077</td> <td>2025</td> </tr> <tr> <td>(?:令和 ^R ^)^078</td> <td>2026</td> </tr> <tr> <td>(?:令和 ^R ^)^079</td> <td>2027</td> </tr> <tr> <td>(?:令和 ^R ^)^10</td> <td>2028</td> </tr> <tr> <td>(?:令和 ^R ^)^11</td> <td>2029</td> </tr> <tr> <td>(?:令和 ^R ^)^12</td> <td>2030</td> </tr> <tr> <td>[年/]([1-9])[月/]</td> <td>0\$1</td> </tr> <tr> <td>[月/]([1-9])日?</td> <td>0\$1</td> </tr> <tr> <td>[^0-9]</td> <td></td> </tr> </tbody> </table> | 変換対象(正規表現) | 変換後の文字 | (?:令和 ^R ^)^075 | 2023 | (?:令和 ^R ^)^076 | 2024 | (?:令和 ^R ^)^077 | 2025 | (?:令和 ^R ^)^078 | 2026 | (?:令和 ^R ^)^079 | 2027 | (?:令和 ^R ^)^10 | 2028 | (?:令和 ^R ^)^11 | 2029 | (?:令和 ^R ^)^12 | 2030 | [年/]([1-9])[月/] | 0\$1 | [月/]([1-9])日? | 0\$1 | [^0-9] | |
| 変換対象(正規表現) | 変換後の文字 | | | | | | | | | | | | | | | | | | | | | | | | |
| (?:令和 ^R ^)^075 | 2023 | | | | | | | | | | | | | | | | | | | | | | | | |
| (?:令和 ^R ^)^076 | 2024 | | | | | | | | | | | | | | | | | | | | | | | | |
| (?:令和 ^R ^)^077 | 2025 | | | | | | | | | | | | | | | | | | | | | | | | |
| (?:令和 ^R ^)^078 | 2026 | | | | | | | | | | | | | | | | | | | | | | | | |
| (?:令和 ^R ^)^079 | 2027 | | | | | | | | | | | | | | | | | | | | | | | | |
| (?:令和 ^R ^)^10 | 2028 | | | | | | | | | | | | | | | | | | | | | | | | |
| (?:令和 ^R ^)^11 | 2029 | | | | | | | | | | | | | | | | | | | | | | | | |
| (?:令和 ^R ^)^12 | 2030 | | | | | | | | | | | | | | | | | | | | | | | | |
| [年/]([1-9])[月/] | 0\$1 | | | | | | | | | | | | | | | | | | | | | | | | |
| [月/]([1-9])日? | 0\$1 | | | | | | | | | | | | | | | | | | | | | | | | |
| [^0-9] | | | | | | | | | | | | | | | | | | | | | | | | | |

- ・ 和暦を西暦に変換します。（例：令和7→2025、R08→2026）
- ・ 1～9（月日）を頭に0をつけて2桁にします。（例：年7月→07、/9→09）
- ・ [^0-9]の正規表現で数字以外の文字を除去します。（例：2030年12月31日→20301231）
- ・ 日付のフォーマット統一の設定例は[こちら](#)をダウンロードしてご参照ください。

4. 使用例 (3/3)

■ 環境依存文字の置き換え

「キ_口」や「m²」のような環境依存文字はAIReadの対応外の文字種のため、標準で読み取ることは出来ません。ですので、下記の変換リストのように誤読した文字を個別読取結果変換で置き換えて出力します。

変換リスト名
単位の修正 + -

number ↑

項目名の修正

住所の修正

西暦和暦変換

単位の修正

店名の修正

変換リスト

| 変換対象(正規表現) | 変換後の文字 |
|------------|----------------|
| [キ*][口口] | キ _口 |
| [mfnM]2 | m ² |
| | |
| | |

- ・カタカナで読み取られた場合の「キ」「口」と、誤読可能性のある「*」や「口(くち)」の文字を変換対象にしています。(例：*口→キ_口)
- ・「m」「2」と、mの誤読可能性のある「f」や「n」、「M」を変換対象にしています。(例：M2→m²)



アライズイノベーション株式会社

東京都中央区勝どき3-13-1

Mail : airead_support@ariseinnovation.co.jp

AIRead Official : <https://airead.ai/>

AIRead Support : <https://airead.tech/>

AIRead Manual : <https://manual.airead.ai/>