

画像文字選択のための「テキスト動線」編集・利用システム—Sijima

津田 光弘

iPallet

OCR で判読が困難な文字画像であっても、テキストのように文字や行を選択できる利用者インタフェースを説明。データモデル「テキスト動線」を設計し、HTML5 の SVG と Web Storage API を用いたプロトタイプ Sijima を試作し、このインタフェースの実現性を確認した。

The "Text-Vein" application — Sijima to select characters and lines in a text image

Mitsuhiro Tsuda

iPallet

This report describes the user interface to select character positions and lines of any digital image that is difficult to identify automatically by OCR Technology etc., like a text. For this, I designed the "Text-Vein" model, made a prototype application by SVG and Web Storage API of HTML5, and confirmed its feasibility.

1. はじめに

デジタル化された画像中の文字を、Web ブラウザや電子書籍アプリのテキストと同じようにマウスやタッチ操作で選択したい、という目標を利用者インタフェースとして開発した報告を行う。今日では多くの貴重な資料が Web で画像として閲覧できるようになったが、その利用については大部分はまだ「影印本を見ている」という域にある。内容をテキスト化し得る対象とそうでないものでは活用面で大きな隔たりがある。主に欧米で著作権処理された古い書籍をデジタル化後 OCR でテキスト化する取り組みが進んでいるが、このような資料は Web で一般情報と同列で扱えるため利点がある。しかしもう少し古い手書きの資料は現在の OCR 技術ではテ

キスト化は困難で、専門家が翻字を行わない限り画像の段階にとどまる。このような現状を見る限り二者択一しか無いのだろうか？高度な技術に依らず、テキスト化もせず、ローテクな方法は意味が無いのだろうか？

以下では、画像中の文字や文章を簡単に「選択」するための仕組みとして考案したテキスト動線 [1] の説明と共に、その実現性を検証するためのプロトタイプシステム Sijima（しじま）について報告する。

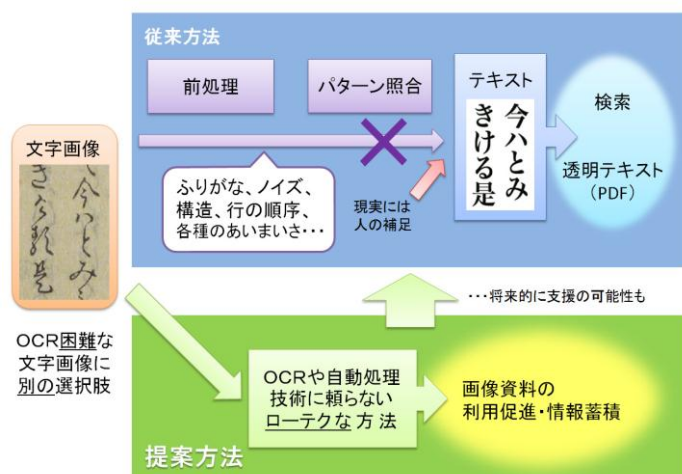


図 1. 課題と目的

2. 「選択」について

文字が書かれた古い書物や資料を画像データを用いて閲覧するとき、私たちはもちろん画像の中の文字の流れを見ている。これらは図 2 のように様々だが、書式や文字がどのようなものであれ、たとえ未知の外国語で意味が理解できなくても、おおよそ、その文字の流れを認識して追うことができる。そして、追えるのであれば、画像であっても Web ブラウザのテキストのようにマウスやタッチ操作で「選択」したいと考えることは自然なことであり、選択することから記録、検索、伝達などの情報行動にも結びつくだろう。

従来、テキストを用いず画像で類似の目的を実現するには PDF や Web マップで経験できるように図形アノテーションを手作業で用いる方法 [2] がある。しかしこの方法は図 3 (a) のように画像を読み難くし選択範囲も利用者次第で自由度が大きく統制がとれない。別の方法に、文字画像の自動識別過程のデータを用いて文字列を多数のブロック情報を用いて選択するインタフェースが報告されている。[3] 文字箇所にも拘束的にガイドを対応付けることができるが、自動化技術であり図 2 のような広い対象では課題があると思われる。

さらに今後、データの Web の時代に向け画像データの内容も構造化し情報化してゆくことが予想できるが、その基盤データとして文字や図がどこにあるか定める必要もある。そこで、新たな概念のガイド手法を考案した。

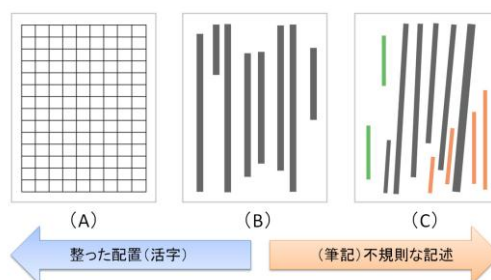


図 2. 対象 (模式図)

3. テキスト動線

文字画像に必要な十分なガイドを対応付け、画像の文字箇所を間接的に選択するデータモデルの提案を行った。[1] 以下で簡単に説明する。

図 3 (a)の一般的なアノテーションは模式化すると(b)のように表せる。丸印から四角形へマウスを動かすことで文字画像の範囲を指定し選択できる。これらは HTML の DOM (Document Object Model) 要素を用いてマウス操作等で選択が可能なオブジェクトにできる。しかし、(c)のように文字列の方向が資料に対し斜めであるとその範囲は大きくなり、隣接する文字列と干渉し正しく選択が行えない。HTML5 では SVG (Scalable Vector Graphics) によって(d)のように文字列に沿って記述が可能である。(d)(e)のような SVG を作成するために必要十分なデータは図の黒い直線群 (破線共) であり、これらを「テキスト動線」として定義した。[1] 図形アノテーションと異なり、テキスト動線は事前にそのデータを作成しておき、あとは利用者の目的に委ねる。文字画像の種類を問わず対応できる利用者と画像間のインタフェースである。

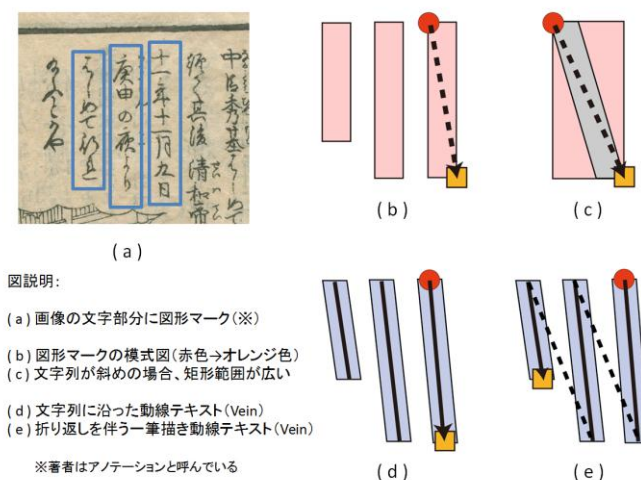


図 3. テキスト動線モデル経緯

4. データモデル

テキスト動線の汎用データモデルを図 4-1 に示す。資料には多様な構造があるが、個々の行は文字の並びの中心を折れ線（ポリライン）で表し得る。日本語縦書き、右上から左下へ折り返しを伴う表記を仮定する。書き文字の場合は行の向きが一樣でない場合もあるが、複雑な構造であっても行を分割すると個々には単純となり最終的に座標データの配列としてまとめ得る。座標系は Web での利用のため、左上を原点、右方向と下方向を正の向きとする。

図の点線外枠はこのテキスト動線の範囲を表す。文字列の本体位置に相当する折れ線はこの枠の右上にある開始点からの相対座標として記述する。テキスト動線の目的は範囲を拘束の下で「あいまいに」指定できるガイドであって、資料の内容をこれで具体的に記述することではないため、割注や振り仮名などの構造はこのデータモデルの積極的な記述対象としない。データは最低限次のパラメータを持つ配列として構造化でき、XML や JSON 形式で記述できる。

- ・ ID（対応関係および順序の識別子）
- ・ 座標値の配列（範囲を導く矩形座標値、テキスト動線のポリライン座標値）

開始点座標 P_a とポリラインの先頭座標 P_1 によって文字幅の $1/2$ 値が得られる。この幅は行の干渉が生じないように、また、一連の文字列に対し代表的な値を設定する必要がある。この P_a と P_1 等の関係から横書きへの対応や文字方向も単位ごとに類似形で行える（図 4-2）。複数行の連続選択もこのデータモデルの配列を用いてプログラム可能であることを確認した（図 4-3）。

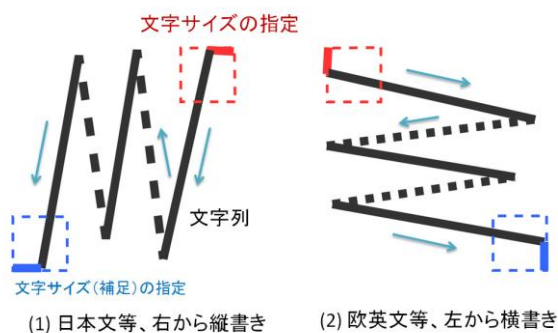
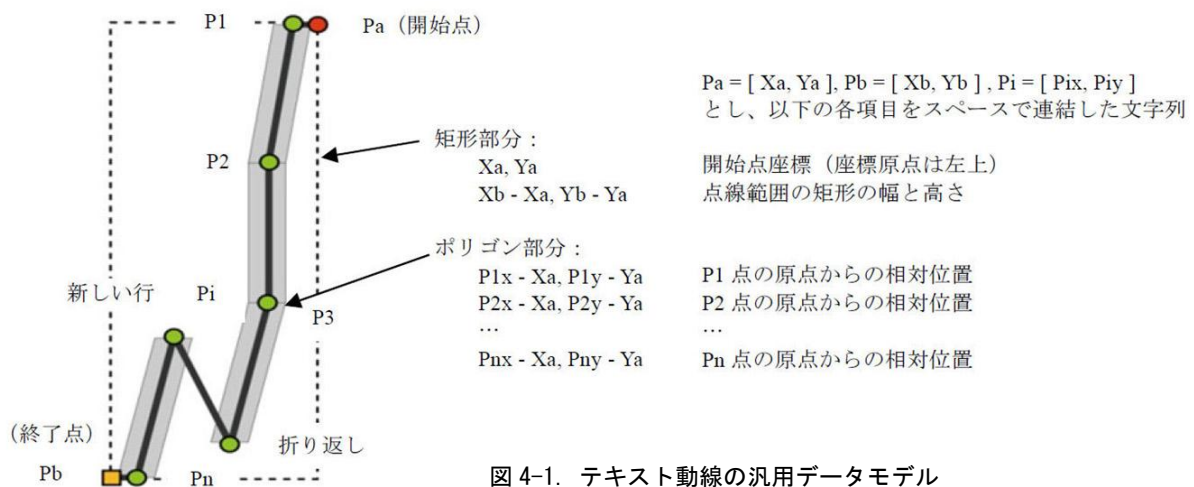
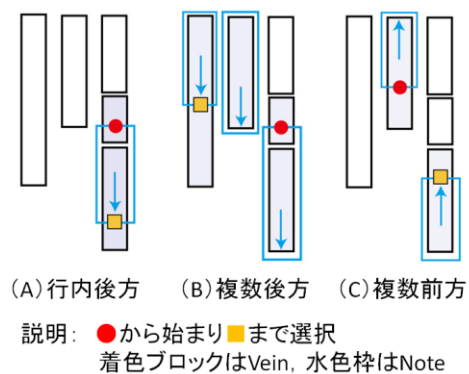


図 4-2. 文字列方向への対応



5. サービスの形態

デジタルアーカイブ画像を公開している提供者が画像と共にテキスト動線データとサービスを提供すると想定する。

テキスト動線データは、利用者が作成することもできるが、まったくの個人利用は別として、各画像に複数の選択基盤データが結びつくことは、後のデータ共有を考える際は注意を要する。

サービスを Web で行うとき、テキスト動線のシステムは次の形態を考え得る。

- 1) 提供者が画像と共に、テキスト動線データ、選択後のデータの保存用DB等のすべてを提供
- 2) 提供者が選択データの保存を除く

利用システムを Web サーバに準備し、利用者側で選択結果のデータを管理
今回の検証用プロトタイプではサービス提供者の運用負担が少ない 2) を想定する。

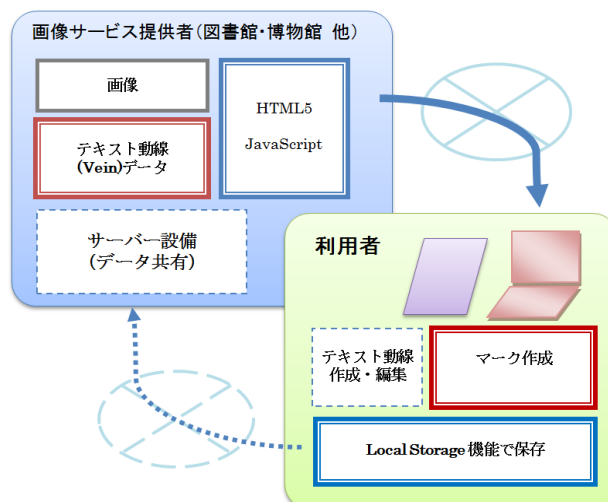


図 5. サービスと Sijima

6. プロトタイプ—Sijima

実現性検証のため、テキスト動線データの編集から選択範囲の保存管理までを簡単なプロトタイプ—**Sijima**（しじま：開発名）として作成した。システムを2つのパートに分けテキスト動線データの編集は既存のアノテーション編集アプリを応用し、利用システムは **HTML5** ベースの **Web** アプリケーションとしてサーバ側に配置した。

6.1. テキスト動線データの作成

編集は通常の図形アノテーション編集方法と同様であり、汎用的には前述のデータモデルをポリラインで作成しテキスト動線のデータの形式に変換する。注など文字の小さな箇所には矩形型、規則的な文字配列のためにグリッド型の一括編集ツールも追加した。データモデルは矩

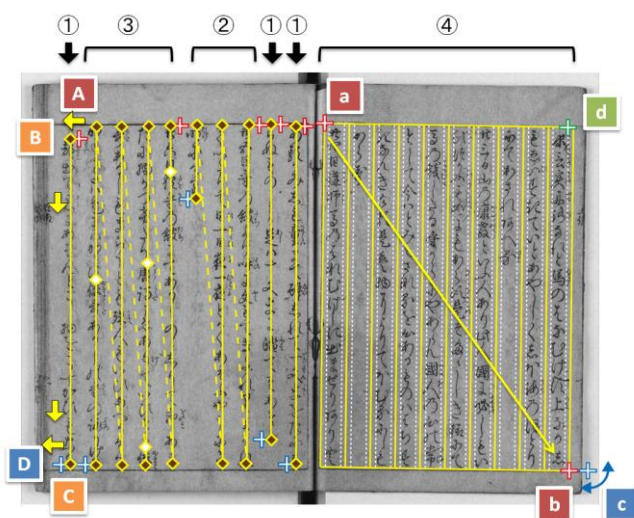


図 6-1-1. テキスト動線の作成

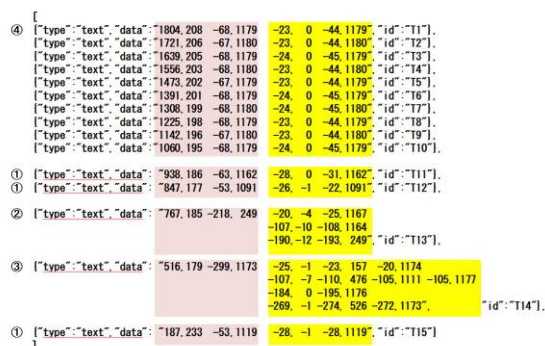


図 6-1-2. テキスト動線データ例 (JSON)

形とポリラインの複合構造のため、挿絵、頭注、画像の縮小時などは矩形情報を用いて選択箇所を簡素化し得る。編集過程で座標の修正や行の連続選択のための順序を変更できる。

6.2. 表示

HTML5 と jQuery を動作環境としてテキスト動線データを Web ブラウザで表示する。ページ読み込み後、JSON 形式のテキスト動線データをサーバから読み込みデータの準備をする。図 6-2-1 に仮想的に示すように、文字画像の行に対応するポリラインデータの配列である。このデータから文字列を囲む SVG (DOM) のポリゴンデータを JavaScript で動的に作成する。結果、図 6-2-2 のように図形が基画像にかぶさる。このポリゴンの SVG 個々にマウス(タッチ)イベントを組み込み、レイヤは透明化する。

以上のように透明化 SVG として作成したテキスト動線を Vein (ベイン:静脈や葉脈の意味) と用語定義する。

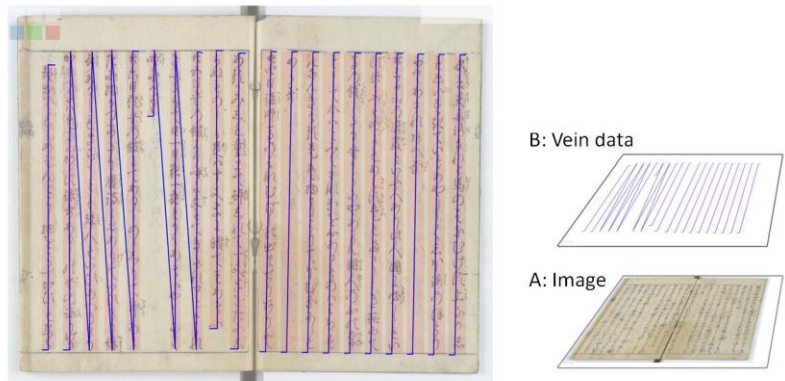


図 6-2-1. テキスト動線 (仮想的表現)

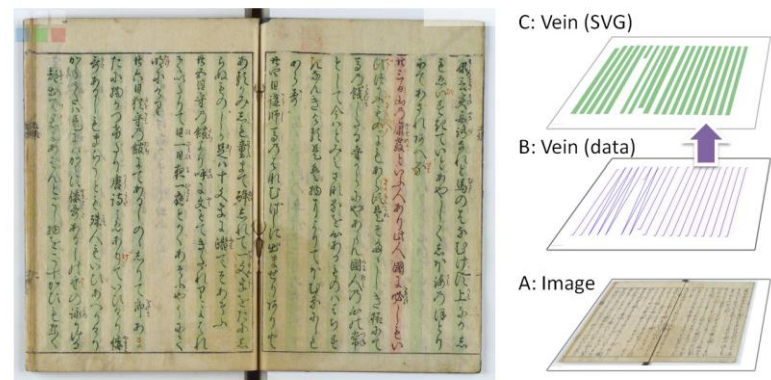


図 6-2-2. テキスト動線の SVG (Vein)

6.3. 選択と保存

SVG 化したテキスト動線(Vein)の利用例を Sijima に実装した。図 6-3-1 のように Vein はマウス (タッチ、以下同様) イベントを感知させるため利用者からみて最上層に配置する。透明化により Vein 自体は見えない。この状態でマウス操作をすると、Vein の情報を反映して操作箇所がハイライトされる。この可視化は Vein の下層で表現する。ハイライト箇所を再度クリックすると簡素なメニューを表示し必要があれば読みやメモ等を記録する。これらの操作は既知の手法を用いて良い。作成したマークやメモは HTML5 の Web Storage API のローカル機能によって各デバイスの所定位置に JavaScript のオブジェクト形式で保存する。この例では個人利用の素朴なメモに過ぎないが、サーバ側で保存サービスを提供すればデータ共有も可能となる。図 6-3-2 に保存された複数のマークの状態、図 6-3-3 に保存データ例を示しておく。

7. まとめ

画像中の文字の選択は当たりまえに実現できていそうだが、その一般化やシステム化がなされていない領域と考えている。新たな情報機器が主役となる今後、デジタルアーカイブ画像を参照し手軽にメモ等を記録できる情報環境の整備は、それ以降の利用のためのデータ蓄積ともなる。仮にローテクな仕組みであっても画像と利用者とのインタフェースが確立できれば資料の活用領域はさらに広がると信じる。まずは教育用途や展示物への応用から進めたい。

本研究は国立歴史民俗博物館基盤研究「デジタル化された歴史研究情報の高度利用に関する研究」(平成22年度～平成24年度)の一部である。

参考文献

- [1] 津田光弘: 無名付箋の画像資料への適用方法について, 情報処理学会研究報告, Vol.2012-CH-95 No.1, 2012.
- [2] 津田光弘: 構造化アノテーションによる文字画像への活字配置とその応用, 人文科学とコンピュータシンポジウム, pp195-198, 2004.
- [3] 猪村元, 田中譲: 図形特徴に基づく手書き古文書画像の全文検索手法, 電子情報通信学会論文誌 D, vol.J92-D, no.11, pp.2022-2032, 2009.

```
"tosa03_T11_p11_908_512_T11_p11_897_955":
{
  "p1": "T11_p11", "p1x": 908, "p1y": 512,
  "p2": "T11_p11", "p2x": 897, "p2y": 955,
  "maxrect": [880, 512, 57, 443],
  "areaval": 25251,
  "type": "spot",
  "url": "http://localhost:8080/vein/pages/tosa03.html",
  "image": "../images/tosa03.jpg",
  "vinedata": "tosa03.json",
  "description": "童(わらハ)まで酔(よひ)しれて",
  "createdate": "2012-10-08",
  "public": "true",
  "identifier": "T11_p11_908_512_T11_p11_897_955"
}
```

図 6-3-3. Web Storage に記録するデータ例
(着色部はキー、key-value 形式)

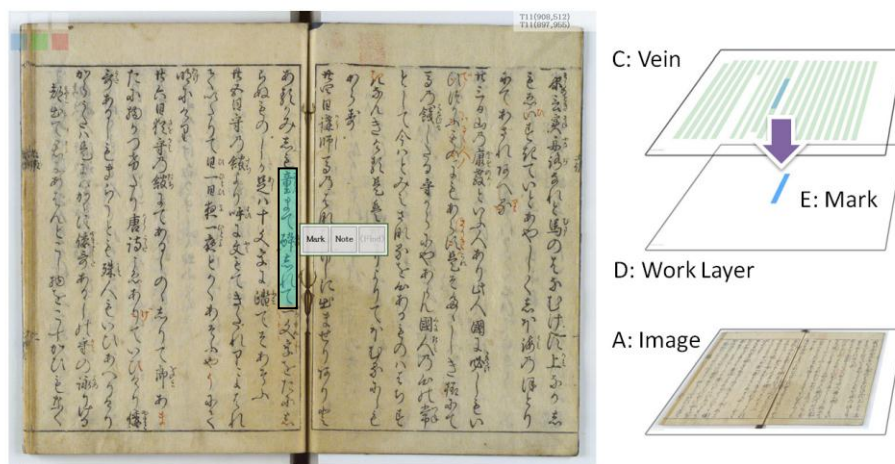


図 6-3-1. テキスト動線を用いて選択

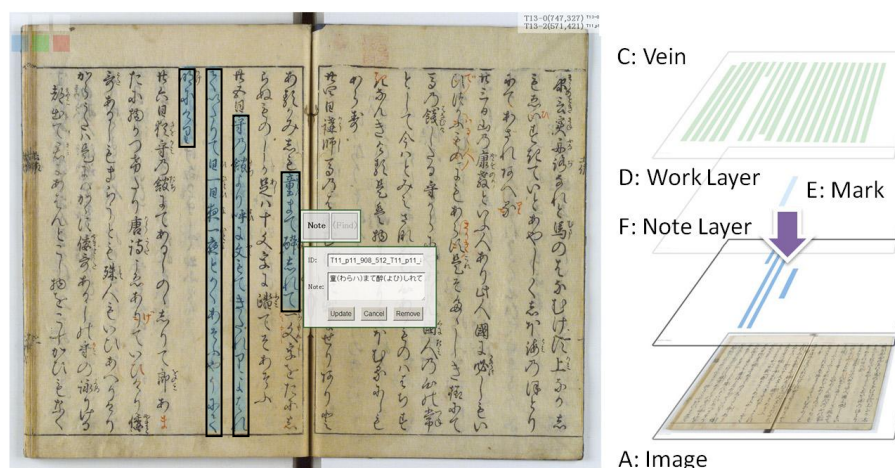


図 6-3-2. マークから記録した情報の読み出し