

文脈を伴う多様なデジタル展示のインタラクティブ編纂

近藤 悠太郎^{†1} 川嶋 稔夫^{†1}

デジタルアーカイブを一般の市民に向けて公開して行くには、単に個々の資料を公開するだけでなく、文脈に沿って資料を配列し解説を加える事が望ましい。そこで本研究では、地域アーカイブを説明する書籍等のテキストをあらすじとして、テキストに対し資料のメタデータを利用した推薦を行うことで、Web ページや仮想ミュージアム等を媒体とするデジタル展示を半自動的に構成する手法を提案する。

Context-based Authoring of Virtual Exhibition for Digital Archives

YUTARO KONDO ^{†1} and TOSHIO KAWASHIMA^{†1}

Authoring of archived data along a story related to the content is desirable when we publish collections stored in digital archives. In this report, we propose an interactive method of authoring for web-based exhibition or virtual digital museum using document-based recommendation approach.

1. 背景と目的

近年は低コストで高精細にデジタル化が可能になった事を背景に、大量の地域の歴史資料が収集・デジタル化され、大規模なデジタルアーカイブをインターネットで公開する試みも行われている。しかし、これらの資料は大量であるにも関わらず、資料を羅列して公開しているだけで一般の市民が気軽に楽しめるものではない。現状のデジタルアーカイブにおいて、いわば資料は1点ずつ倉庫に置かれた状態であると言える。

アジア歴史資料センター^{*1}は、「インターネット特別展」と称して、あるテーマの歴史的背景とともに関連する資料に解説を加えた Web ページを公開している。従来より親しみやすい公開方法ではあるが、このようなコンテンツを作成するのは一般に手間がかかり、公開できる対象も資料全体のごく一部に過ぎない。特に地方において、今後さらにデジタルアーカイブを積極的に活用してゆく為には、限られた人材やコストの条件で、大規模なアーカイブの全貌をいかに公開していくかが課題になってくる。

本研究では、デジタルアーカイブの資料を説明する書籍等の文書に基づいて、半自動的に文脈を伴ったデジタル展示を構成することを目的とする。本研究によって、特別な知識を持たない一般の市民も文脈に沿って並べられたデジタルアーカイブの資料を楽しみながら鑑賞してゆくことができると考えられる。

2. 関連研究

デジタル化された収蔵物の鑑賞手段として仮想ミュージアムは広く研究されてきた。近年はセカンドライフを利用した研究¹⁾も多数行われているが、これらは主に3次元空間内でのコミュニケーションなど、ミュージアムの機能や活用方法を問題としており、コンテンツの作成は人手による作業が前提である。

XML ベースで展示物の階層構造を記述し、Web ページや仮想ミュージアムをテンプレートに沿って生成する研究²⁾も試みられているが、知識を持つ学芸員が手作業で階層構造とその見せ方を入力するもので自動的な構成とは言えず、その表現も単調なものに限られている。

大量の写真等のデータから、ユーザが楽しめるコンテンツを、低コストで作成する研究は広く行われている。例えば個人が撮影した写真群から、撮影時間や画像自体の内容に基づいてアルバムを自動的に自動的に構成する研究³⁾がある。しかし、このような研究は最近のマルチメディア環境を前提としているため、デジタルアーカイブに収蔵される歴史的な資料群に対して直接適用するのは難しい。

大量の歴史的な資料から、自動的にコンテンツを作成するアプローチは、十分に検討されていないのが現状である。近藤らは、アーカイブを説明する文書をストーリーラインに見立てて、仮想ミュージアムを構成する方法を提案している。実験では函館町物語を節ごとに分割し、TFIDF 法を使ってアーカイブの資料を推薦した結果に考察を加えた。しかし、文書とアーカイブの資料の性質から、他と比べて極端に多い資料が特定の節に集中し文脈を感

^{†1} 公立はこだて未来大学
Future University-Hakodate

^{*1} <http://www.jacar.go.jp/>

じにくくなる事や、節での単語の使われ方によって本来推薦されるべき資料が見逃されるなど、望ましい結果が得られない場合があった。本報告ではその対策として、資料の推薦にあたり、赤石の手法を用いて、偏って推薦された資料をさらに振り分け、解説が必要な話題に関連した文書を発見してミュージアムに組み入れる事で、より自然な展示の構成を目指した。

3. 提案手法

本研究では学芸員が博物館・美術館で展示を構成するプロセスを参考に半自動化のアプローチを考える。博物館・美術館で実際のミュージアムが構成される過程と、本研究で空間デザインを自動化する手法を図1に示す。

実際のミュージアムで展示が構成される過程では、まず学芸員が自身の知識を元に、情報を収集しながら展示のアイデアを考案して、ストーリーラインにまとめ、ストーリーラインに沿って資料が収集されて、展示の実施に至る。

本研究では、ストーリーラインの元になるキュレーターの知識に相当するものとして、展示対象とするアーカイブを説明した書籍等の文書を用いる。

更に、その文書を意味単位に分割して、分割した文書の部分に、順序や相互の関連性など、展示の中で持つ意味を付与してゆく。この操作がストーリーラインの作成である。

続いてストーリーラインに基づく資料の収集の工程では、意味単位に分割した文書と、個々の資料が持つ目録情報の比較を行うことで、文書の部分に対して展示すべき資料の推薦を行う。

最後に、展示の構成の段階では、構成したい媒体それぞれの方法で、文書の部分に推薦された資料群を、文書の部分の関連性や意味によって配置して、レイアウトを決定する。

デジタル展示の構成に於いては、資料の収集や展示に物理的な制約は存在しないので、文書の分割から資料の配置までの繰り返しを行うことで、展示の完成度を高めて行く。

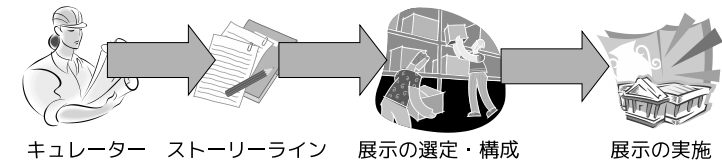
本研究ではこの手順でデジタル展示のデザインを自動化する。

4. デジタル展示構成のプロセス

テキストに基づいてデジタル展示をデザインするプロセスは3段階に分けられる。

- (1) 文書を適度な大きさの意味単位に分割し、相互の関連性や個々の性質を付与し、展示単位を定義する
- (2) 展示単位中の意味で分割した文書と、資料の目録情報を比較して、展示単位に資料の

ミュージアムの展示の構成



半自動的に行うアプローチ

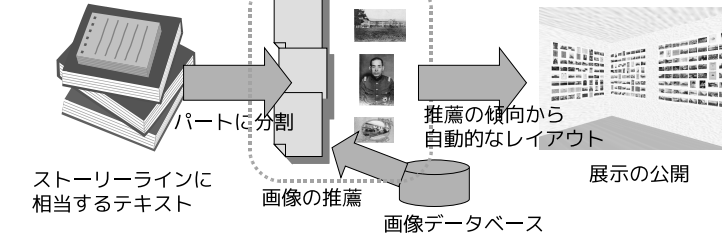


図1 半自動構成のコンセプト

推薦を行う

- (3) 関連性や性質を使って資料のレイアウトを決定し、展示単位をデジタル展示に変換する

4.1 展示単位

本研究では、抽象的な展示の単位のモデルを次のように定義する。展示単位は以下の要素から構成される。また展示単位は内部により小さな展示単位の集合を持つ。

- 文書の部分
- 他の展示単位への関連性のリンク
- 下位の展示単位の集合

最終的に変換の行程で、この展示単位が、デジタル展示の中で1つのWebページや、3次元ミュージアムの展示室内の展示コーナーに変換される。

4.2 文書の分割・階層・関連性の付与

まず展示単位の文書の部分を指定する方法である。もちろん人手により、妥当だと思われる箇所まで文書を分割し、章や節のような階層的な構造を定義しても良いが、文書の性質を利用して半自動的に文書を展示単位にする方法も考えられる。

4.2.1 文書構造の分析

一般に書籍等の文書は、章や節といった形式的な記述順序と階層構造を持って記述されている。また章や節の内部に、更に細かい意味的な構造を持つ場合もある。このような形式的または意味的な記述順序と階層構造は、時間、場所、分類など記述される内容の変化を伴っていると考えられる。

4.2.2 形式的構造の利用

そのため文書の形式的な章立てを、そのまま展示単位として扱う方法が考えられる。例えば章、節をそれぞれ別の階層の展示単位として、章同士を出現順序によって1方向リンクで次の章へ関連付ける。節についても同様にする。

4.2.3 意味的構造の発見

形式的なものではなく、意味的な切れ目をシステムにより検出する方法も考えられる。例えば節が内部に意味的な構造を含んでおり、後述する資料の推薦の行程で、多量の資料が推薦されたり、不明確な推薦が行われるなど、都合の悪いものだったとする。このような場合、話題の変化により更に細かい展示単位を発見して推薦を行うことで、展示をより文脈に沿ったものにできる可能性がある。

本研究では話題の発見に赤石⁵⁾の系列的分割の手法を用いる。この研究では、単語の共起のみに基づいて物語を分解し、分解した物語を別の文脈で再構成する手法が提案されている。系列的分割の手法は、記述順序に沿って話題の変化を見つけて文書群を分割する方法である。文書群を入力として、分割された文書群の一部分と、そのトピック(単語)が吸引力という重要さを表す指標と共に得られる。

4.2.4 関連文書の提示

これまで文書構造を反映した展示単位を構成する手法を紹介した。しかし、単に文書に沿ったミュージアムを構成するだけでなく、掘り下げが足りないテーマについて、関連文書から補足的な展示を構成したい場合が予想される。これは、本来の記述順序とは別に、補足となる展示単位を、既存の展示単位群に組み入れる操作である。

システムでこの作業を支援するため、前処理として関連文書を赤石の場面的分割法により分割し、分割された文書ごとにトピックを求めておく。場面的分割法は、記述順序を問わずに、文書群から話題を抽出する方法である。入力と出力は系列的分割法と変わらない。

ある部分を補足したい必要が生じた場合、その部分のトピックを調べ、トピックがわかっている文書群のうち同じトピックの文書を提示すれば、学芸員が展示を構成する際の手がかりになると考えられる。

4.3 展示単位への資料の推薦

各展示単位の持つ文書の部分を推薦の単位(パッセージ)、資料の目録から生成したクエリと比較を行って、展示単位に展示すべき資料の推薦を行う。

パッセージに対する資料の推薦にはTFIDFを重みとしたベクトル空間法を用いる。推薦結果は類似度順でソートし、学芸員に提示し、学芸員は展示する資料を選択する。この際に推薦結果に偏りがある場合、学芸員は適切な画像を指定して関連フィードバックによって再推薦を行うこともできる。

文書に付属する索引は、文書の内容において重要な単語であり、それらの語をTFIDF法の辞書として扱うことで、推薦の精度を向上させられると考えられる。しかし、索引は文書によって性質が異なるので、索引を辞書として利用する場合は、索引語がかえって推薦を妨げないか、検討する必要がある。

4.4 展示単位のデジタル展示への変換

文書から作成された展示単位は、この工程でデジタル展示に変換される。本稿ではWebページと3次元ミュージアムの2種類の展示への変換方法を以下に示す。

4.4.1 Webページ

Webページへの変換は、基準として指定した階層の展示単位を1つのWebページとして変換する。基準となる階層の上位の階層は、ディレクトリとなる。下位の階層はWebページ中にセクションとして埋め込まれる。各ページには文書の部分と、推薦された画像が表示される。

4.4.2 3次元ミュージアム

ミュージアムは展示室の連続で表現する。展示室が変換の基本となる展示単位である。最上位の階層など、文書の大まかな構造を展示室の連続である鑑賞通路の曲がり角とする。展示室に割り当てた展示単位の下位の展示単位は、展示室の進行方向に沿って再帰的に展示スペースとする。展示スペース内には、展示スペースの展示単位の推薦資料を配置する。展示室や展示スペースを構成する際には、パッセージの性質や推薦資料数を空間の幅や大きさに対応させる事で、ミュージアムのウォークスルー時に文書の内容をユーザが感じられるようにする。

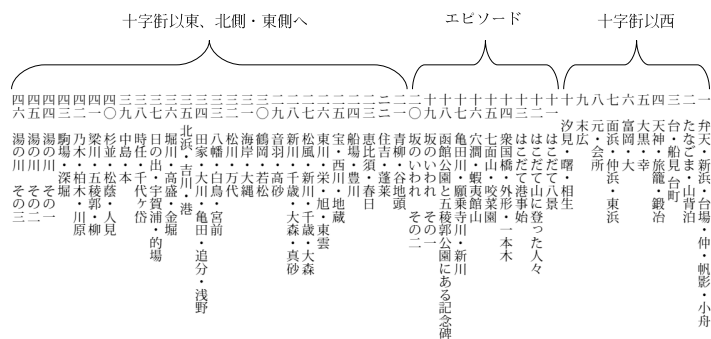


図 2 (三) 函館町物語の節構成

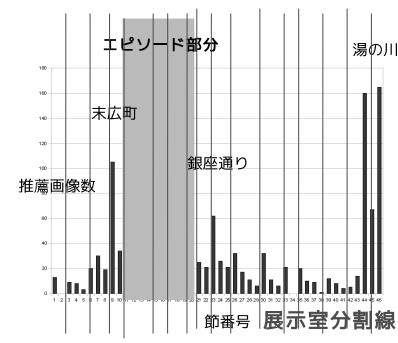


図 3 推薦画像数と展示室の分割
閾値(類似度80)以上の画像のみを集計、全1138件

5. 実験

実験として、文書に函館の町の由来と沿革を紹介する「新編＝函館町物語」*1の中核部分である「(三) 函館町物語」(全体の 1/5 程度)、その関連文書として「新編＝函館町物語」のストーリーライン以外の部分を、展示する資料として函館市中央図書館が Web で公開している「函館の絵葉書」*2を用いて、ミュージアムの構成を試みた。

(三) 函館町物語の構成は図 2 の通りである。全 46 節からなっており、節ごとに函館の町の歴史を紹介している。1 節から 10 節で函館の十字街以西の町を紹介し、21 節から 46 節で十字街以东、さらに北側、東側の町を紹介している。11 節から 20 節は町の成り立ちではなく、函館の坂や港を話題とする、エピソード部分が挿入されている。本書には索引が付属しており、索引の正規化を行った上で一般名詞を除外し、全 1916 語のうち 1782 語を TFIDF 法で処理する辞書として用いた。

「函館の絵葉書」は函館市中央図書館による Web サイト「デジタル資料館」の絵葉書のアーカイブである。今回はこのうち函館の写真を用いた 6351 点だけを推薦対象の絵葉書とした。すべての絵葉書には目録が付属しており、書籍の文書と比較するために、タイトルと解説を連結した文字列を絵葉書のクエリとして用いることにした。目録の例を表 1 に示す。

*1 元木省吾著, "新編＝函館町物語", 幻洋社, 昭和 62 年発行

*2 <http://archives.c.fun.ac.jp/postcards/>

6. 実験結果

6.1 展示単位の指定

(三) 函館町物語は、全 46 節から構成されるため、まず各節を基本となる展示単位とした。しかし、(三) 函館町物語には、前述の通りエピソード部分が挿入されている。このエピソード部分のある種の章立てと見なして、節を包含し全体を 3 つのパートに分ける上位の階層の展示単位を定義した。

6.2 資料の推薦

図 3 は節を展示単位として、設定した閾値 (80) を越える類似度で推薦された資料数と、展示単位の分割を示したものである。

大量に資料が存在する部分は、記録がそれだけ残っている部分であり、いわば話題の盛り上がりであると考えられる。そこでグラフ中の谷を境にして、個々の推薦単位に山が含まれるように、全体を 16 の展示単位に振り分け、展示を順に見ていく事で、函館の町の盛衰を感

表 1 目録の例

Table 1 An Example of Catalog Data

タイトル	函館市柏野尋常小学校校舎前景
西暦	1928
和暦	昭和 3 年
解説	末広堂高田製 函館市柏野尋常小学校 昭和三年十月 開校記念絵葉書 函館市柏野尋常小学校 (3 枚中 1 枚)
資料番号	pc000001-001
地域	北海道 (函館)

じられるよう工夫した。

6.3 展示単位の分割

推薦結果を見ると、未広町の節には大量の画像が推薦されているため、展示室内部で話題に沿って資料を展示する事にした。図4は未広町の16文を、系列的分割法により処理した結果である。ここでは3パートに分割され、内容を確認するとそれぞれ未広町の成り立ち、繁栄、そして衰退と解釈できる内容に分割されていた。

各パートに対して、元々未広町の展示単元に推薦された資料を、各パートと資料の間で類似度を計算し、最大のパートへ振り分けた。比較的良好な結果が得られたのは後半で、未広町衰退の要因だった大火についての資料が推薦された。しかし、実際には8割近い資料は未広町に強く推薦されたもので、推薦画像は振り分けが行えなかった。

振り分けが不十分な原因は、資料の時代に偏りがある事に加えて、例えば写真に写っている建物があっても目録に記載されないなど、テキストと目録が目目している粒度の違いが原因だと考えられる。

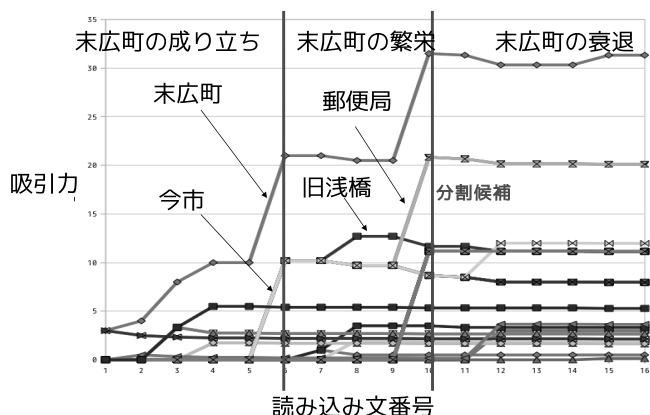


図4 系列的分割法による未広町の分割

6.4 十字街の補足展示単位の追加

前処理で、関連文書の全文節を入力として、場面的分割法によりストーリーの抽出を行った結果、115の独立したストーリーが得られた。

更に未広町の推薦結果を詳しく見ていくと、未広町の一角である十字街は、かつて未広町

でも特に栄えた場所であるが、函館町物語では、未広町の解説に「未広町」という単語だけが繰り返し使用されており、「十字街」はほとんど出現しないため、TFIDFや赤石の手法でも、十字街が重要語とはならず、「十字街」のみを含む資料が推薦されることもなかった。

そこで、関連文書から作成したストーリーのうち、十字街をメインピックとするストーリーを文書の部分とする展示単位を作成し、推薦を行った所、ストーリーライン単体では見逃されていた、55点の十字街についての資料が推薦された。この展示単位は、未広町から次の節へ続く順路とは分岐する形で、未広町を含む展示単位からリンクを設定した。

本実験で構成したウェブページのうち未広町の各パートを図5に、作成したミュージアム全体の平面図を図6に示す。



図5 未広町の Web ページ

7. 考察とまとめ

様々なデジタル展示を構成するために、抽象的にストーリーラインを取り扱うための枠組みとして、展示単位を定義し、順に処理を行う事でデジタル展示を構成できる事を確認した。現在は Web ページと簡易な3次元ミュージアムが出力できるに止まっているが、展

論文誌 Vol.21, 2006.

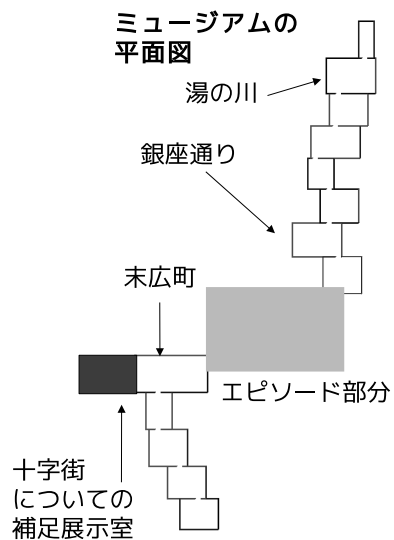


図6 ミュージアムの平面図

示単位を年号や座標など定量的な属性を定義できるように拡張すれば、デジタル展示への変換段階で、展示単位をマッピングして平面上で可視化したり、表形式にまとめて年表を作成するなど、更に多様な形式の出力が実現できると考えている。

今後は、函館に関する知識を持つ人に実際にミュージアムを構成してもらい、評価を行いたいと考えている。

参 考 文 献

- 1) Design an Immersive Interactive Museum in Second Life, Q. Zhu, Digital Media and its Application in Museum and Heritages, 2008.
- 2) ARCOLite-an XML based system for building and presenting virtual museum exhibitions using Web3D and augmented reality, M. White, Theory and Practice of Computer Graphics, 2004.
- 3) Mixed-initiative photo collage authoring, J. Xiao, Proceeding of the 16th ACM international conference on Multimedia, 2008.
- 4) デジタルアーカイブのための3次元ミュージアムの半自動的構成, 近藤悠太郎, 川嶋敏夫, 情報処理学会研究報告 Vol.2010-DD75 No.2, 2010.
- 5) 文書群に対する物語構造の動的分解・再構成フレームワーク, 赤石美奈, 人工知能学会