Generating Paths through Cultural Heritage Collections

Samuel Fernando, Paula Goodale, Paul Clough, Mark Stevenson, Mark Hall, Eneko Agirre
Outline

• Paths through cultural heritage collections
• User study of path creation behaviour
• Automatic enrichment of paths
  • Preliminary experiments
Introduction

Museums have guided tours with narratives organised by theme or chronology while digital libraries are often limited to a simple keyword search.
Claude Monet

A path linking the popular artworks, inspirations and peers of the artist Claude Monet.

Overview

Claude Monet → Painting Style → Impressionist paintings in Oil

Selected Artworks

Poplars (Les Peupliers) → Springtime (Le Printemps)

Related Artwork...

Study of Poplars → St Tropez

Related Artists → Auguste Renoir

LaTeCH 2013, ACL Workshop, Sofia, Bulgaria
User scenarios

Paths

- Curated collections
- Leisure users
- Classroom activities
- Local historians
- Researchers
Creating a path (1)

Select items from the collection (Grand Tour of Rome)

- Constantine Arch
- Palatine Hill
- Colosseum
Creating a path (2)

Arrange items to form a path or narrative (in this case for a geographical tour)

Colosseum → Palatine Hill → Constantine Arch
Creating a path (3)

Add background text /titles etc.

Grand Tour of Rome

Colosseum → Palatine Hill → Constantine Arch

Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Lorem ipsum dolor sit amet, consectetur adipiscing elit.
User study

• 22 users
  • Mixture of subject experts, students, general users

• 2 hour session including
  • Familiarisation with the system
  • Short information seeking tasks (5 min x 4)
  • Simulated work task on path creation (30 min)
  • Feedback questionnaires
Path creation task

“Imagine you are a student who has been asked to **create a path** as part of a university assignment. You have been asked to use primary source materials to create a mini online exhibition suitable for a target group within the general public and/or school visitor categories. Your goal is to introduce a **historical or art-focussed topic** in a popular, accessible way, and to encourage further use and exploration of cultural heritage resources.”
Example Paths

Example titles:
- Canals in England
- Brooches through the ages
- Art and Criticism of John Ruskin

![Pie chart showing the distribution of items in paths]

- Places (23%)
- Art (23%)
- History (32%)
- Other (22%)

Number of items in path:
- Min=5, Max=29
- 59% with 6-10 items
- Mean=10.7
Selecting items for paths

- Serial searching (33%)
- Serial browsing (39%)
- Primarily exploring (6%)
- Mixture search/explore (22%)
Types of ordering

- Chronology (32%)
- Narrative (23%)
- Geography (9%)
- Theme (9%)
- Interestingness (5%)
- No ordering (22%)

LaTeCH 2013, ACL Workshop, Sofia, Bulgaria
Adding contextual information

- No context added (41%)
- For some items (27%)
- For all items (32%)
Automatically adding context

A significant number of users are unwilling to add contextual information to the paths. However cultural heritage experts place great importance on context. Therefore we explored automatic approaches.

- Sequence of items (S)
- Key Wikipedia article (K), the most relevant article for S
- Sentences from K summarising S
Two methods for summary generation

First-n: Choose the first $n$ sentences from $K$

Scored: Find key topics $W$ from $S$ and use this to select sentences from $K$.

$W$ comprises titles of Wikipedia articles (from using Wikipedia Miner).

Score each sentence in $K$ by occurrence of words from $W$ multiplied by confidence values.

Return $n$ highest scoring sentences
First-n example

Path: “Bronze age metallurgy” comprising items and debris from Bronze Age.

The “Bronze Age” is a period characterized by the use of copper and its alloy bronze as the chief hard materials in the manufacture of some implements and weapons.

Chronologically, it stands between the Stone Age and Iron Age.
Scored example

K = “Bronze Age”, W = Bronze Age (0.9), Copper (0.9), Alloy (0.8), Tin (0.6), Bronze (0.6)

Highest scoring sentence: A region could be in the Bronze Age either by smelting its own copper and alloying with tin or by trading for bronze from production areas elsewhere (score 3.8)

2nd highest scoring: The Bronze Age in Ireland commenced around 2000 BC, when copper was alloyed with tin and used to manufacture Ballybeg type flat axes and associated metalwork (score 2.4)
Evaluation

• 5 annotators independently rated the automatically generated background text for each of the 22 paths.

• The text was rated for Relevance, Coherence and Contextualisation

• Ratings were on a 5-grade scale from A (Excellent) to E (very poor)
Results

[Bar chart showing the results for Relevance, Coherence, and Contextualisation with categories Excellent, Good, Average, Poor, and Very Poor across different measures such as First-n and Scored.]
Summary

• Studies analysing different behaviours when creating paths, and properties of created paths

• Experts rate importance of context in path, but users tend not to add this when creating their own paths

• Automatic methods for adding context show good results using a simple approach
Future work

• Build on work for automatic creation and enrichment of paths
• Automatic methods for selecting key Wikipedia articles
• Filter out information from summaries that is already present in path
Questions?

Paths system:
http://explorer.paths-project.eu
References

*Expert interpretations of paths are discussed in:*


*Mapping from items to Wikipedia:*

Introducing paths

We introduce the idea of paths into digital library collections. These may be created by experts or end-users.

Path Title (+keywords, description)

Grand Tour of Rome

Colosseum  Palatine Hill  Constantine Arch

Item title (+keywords, description)
Overview

User studies

- subject experts
- students
- general users (subject novices)

What do people want from paths? How might they use them?

Analyse properties of created paths

Problem: users do not add background text

Propose automatic methods to address this

22 users in total
Expert interpretations of paths (Goodale 2012)

Hypertext trails

- Search history
- Information seeking journey
- Linked metadata

Paths
- Through route
- Augmented reality
- Starting point / way in

Encouraging exploration and enriching data

Learning process
- Transaction process

Process oriented