Using Formal Concept Analysis to Create Pathways through Museum Collections

Tim Wray, Peter Eklund**

Faculty of Engineering and Information Science
University of Wollongong
Northfields Avenue, NSW 2522, Australia

Abstract. This paper presents *A Place for Art* – an iPad app that allows users to explore an art collection via semantically linked pathways that are generated using Formal Concept Analysis. The app embraces the information seeking approach of exploration and is based on the idea that showing context and relationships among objects in a museum collection augments an interpretive experience. The fundamental interaction metaphor inherent in *A Place for Art* relies on Formal Concept Analysis so the interface has embedded within it the semantic clustering features of machine learning in artificial intelligence.

Keywords: Intelligent Interfaces, Formal Concept Analysis

This paper presents "A Place for Art", a working artefact developed by a team of developers of which the authors are members and that has been reported more extensively elsewhere [17] albeit not to a AI or FCA audience. The work can be framed as a contemporary extension of work using FCA for Information Retrieval[6, 3].

A Place for Art showcases a collection of contemporary and Australian indigenous works from the University of Wollongong's Art Collection. It is a digitized companion piece to the print publication of the same name [9]. The app provides access to 77 works and accompanying short essay pieces that feature the history of the collection and significance to its local region. The key result is a semantic navigation concept driven interaction paradigm that has Formal Concept Analysis (FCA) at its heart. A demonstration of the design – an iPad app – is a very important companion to this written text and the reader is encouraged to download, install and run the app while reading the paper.

1 Pathways through an Art Collection using FCA

Interaction and navigation in A Place for Art relies on users creating and exploring their own path through the collection: an approach that is well supported by the literature on information seeking in museum collections. For example: Skov[13] found that online visitors demonstrated exploratory behaviors such as

** On leave in 2014 to the IT University of Copenhagen, Denmark

serendipity and, when finding the unexpected, exhibited meaning making qualities, i.e. following paths and making implicit connections between objects. Further, Goodale et al. [7] conceptualize the pathway as a guiding metaphor to characterize the design of digital artifacts that support creative and divergent exploration of cultural data-sets. In their findings, the authors elaborate that pathways can be both used as a means to creatively explore a collection and also to structure the relationships among the objects.

Previous work has experimented with semantically linking museological content [12, 1], such as the semantically enriched search platform produced by Schreiber et al. [11]. In our work, we build a custom user interface that embraces the pathway metaphor by allowing the user to navigate clusters of related art content and 'branch off' at specific points of interest. To do this, we employ Formal Concept Analysis (FCA) to derive clusters of related artworks, and then exploit FCA's relational properties to generate semantically linked pathways. In this paper, we apply the terms *convergence* to describe the way FCA can be used to cluster related objects, and *divergence* to the way a visitor could potentially move through different sets of object clusters based on their shifting points of interest. We elaborate on this *convergence* - *divergence* approach to navigation, along with a brief overview of FCA in the following section.

1.1 Formal Concept Analysis, Convergence and Divergence

Formal Concept Analysis (FCA) was developed in the early 1980s as a mathematization of the human cognitive constructs of concepts and concept hierarchies [14]. Concepts are understood as basic units of thought shaped by observations of existing phenomena and formed in dynamic processes within social and cultural environments ([15], p. 2). According to its philosophical definition [15], a concept is composed of a set of objects as its extension, and all attributes, properties and meanings that apply to those objects as its intension. As an example, if one considers the idea of works that depict heavy industry and the Illawarra[1] (its intension), which we derived from the text analysis of the printed catalogue, there are 7 paintings (its extension) in the collection which have these attributes (see Table 1); a concept is therefore defined as the simultaneous perception of its intension and extension, i.e., the compositional qualities of those paintings (as attributes) and the actual paintings (as objects) defined via those attributes.

In FCA, concepts are mathematized as formal concepts defined as a pair (A, B) where A and B respectively are elements of the set of objects G (the formal concept's extension) and the set of attributes that describe those objects M (the formal concept's intension). Concepts are never perceived in isolation, but in the context of existing phenomena. By interpreting concepts in context, one can derive implications and perceive their relational and spatial properties to other concepts. For the purposes of interpreting works in A Place for Art, we consider groups of artworks that share equivalences as concepts that are a part of the 77 works that compose its collection, its context. In FCA, the context is mathematized as a formal context defined as a $\mathcal{K}(G, M, I)$ where G

[1] The Illawarra is the name of a region 80-160km south of Sydney, Australia.

Table 1. A sample of formal concepts from A Place for Art

Formal concept, expressed in natural language	**No. of objects**
paintings that depict the Illawarra	8
works that evoke identity issues and social critique	6
surreal works that depict animal imagery	6
vibrant and abstract paintings	11
intricate works that depict nature	6
vibrant works that evoke a sense of calm	6
works that depict heavy industry and the Illawarra	7

and M respectively describe its set of objects and attributes and I describes the associations between them. Formally, $I \subseteq G \times M$ is a binary relation where $(g, m) \in I$ is read object g has attribute m.

A valuable layer of meaning is added when concepts are perceived in context. One way of inferring meaning is by deriving attribute implications. These implications provide the assertion that within a given context, if all objects with X attributes also possess Y attributes, then X infers Y. Applying this form of inference gives the ability to gain insights into the implicit relationships and phenomena within the collection. In A Place for Art, it infers that, for example, all works that depict natural landscapes are painted with coarse brush strokes, or that all the depictions of heavy industry in the A Place for Art collection also all take place in the Illawarra. Using the latter assertion as an example, these attribute implications are formed by the way formal concepts are constructed: for a give formal concept (A, B), that has an attribute set $M =$ {'heavy industry'}, let G be composed of all objects that possess M, giving the result of: $G =$ {'Waiting, Port Kembla', 'Foundry Men', 'Steel Works BHP',...} Now let M be all attributes common to objects in G, giving the result of: $M =$ {'heavy industry', 'the Illawarra'} G and M are then combined as (A, B) to create a closed concept. The additional attributes that were derived from this operation give rise to their implication, in this case, 'heavy industry' $\rightarrow$ 'the Illawarra'. In A Place for Art, formal concepts are computed using the PCbO algorithm [8] but the choice of algorithm is not important.

Within its context of 77 artworks, there are a total of 330 formal concepts, 7 of these formal concepts are expressed in natural language are shown in Table 1. By deriving clusters from data and inferring association rules, formal concepts provide the mathematical realization of – what we term – *a convergence*: the way a group of otherwise disparate works of art are represented as a meaningful whole. A Place for Art also employs purpose built algorithms for describing concepts in natural language.

The examples shown here and in Table 1 are direct outputs generated from these algorithms. When a formal concept is expressed in natural language, the algorithm takes its intension and orders it based on their semantic qualifiers and parts of speech, such as whether they depict the work itself ('painting', 'screenprint', etc.), are adjectival ('surreal', 'vibrant', etc.) or are otherwise appended as clause fragments ('identity issues', 'a sense of calm', etc.). Using basic princi-

ples of grammar and sentence construction, these attributes are then conjoined to produce a statement. The algorithms also take into consideration whether the natural language statement should be expressed in a singular or plural form, given that, according to the principles of FCA, individual objects are also formal concepts. These natural language statements are used to convey the semantic meaning of the convergences as human-readable, narrative-like statements.

Meaning is further conveyed when concepts are observed in relation to one another. Concepts are inherently spatial and relational, as connections of concepts are networked to create a concept lattice [14] or are spatially conveyed via a measure of their concept similarity and distance. One common method of constructing a knowledge space in FCA is via the exploitation of the subconcept/superconcept relationship. Within a context, the complete set of formal concepts– ordered by this relationship – induces a concept hierarchy – an implicitly structured collation of human knowledge that can be represented visually as a concept lattice or line diagram.

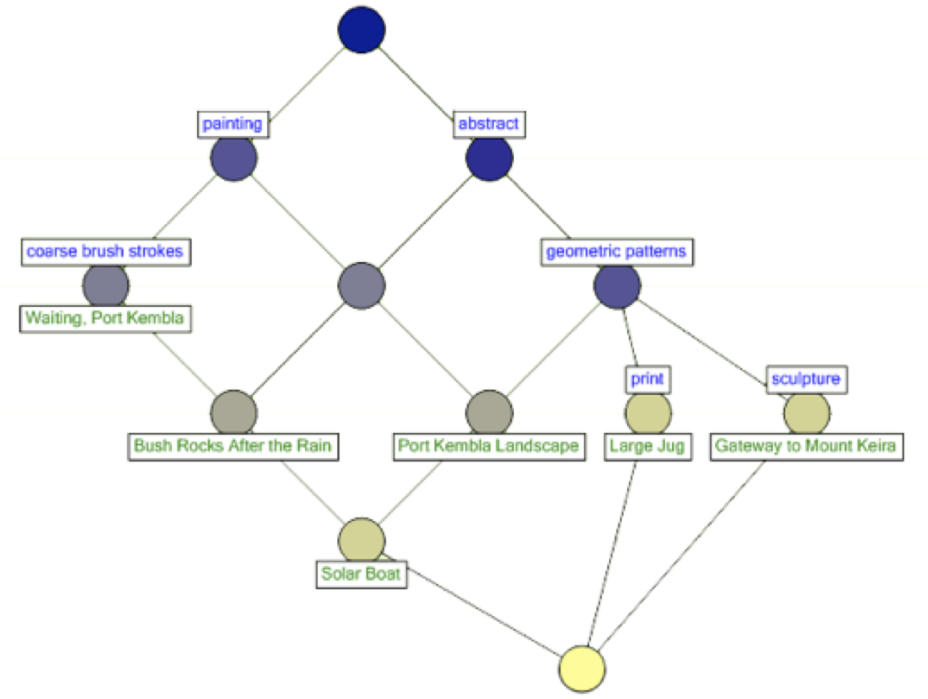

Fig. 1. A line diagram showing a small selection of artworks from A Place for Art.

Following the example in Fig. 1, the concept 'abstract paintings with geometric patterns and coarse brush strokes' – depicted by the artwork titled 'Solar Boat' as it appears bottom-right in the concept lattice is a subconcept of abstract paintings with geometric patterns – that also includes the artwork Port Kembla Landscape, appearing to its top-right – which in turn is also subconcept

of both abstract paintings and abstract works with geometric patterns. In FCA, a formal concept (A, B) is a subconcept of (C, D) (expressed $(A, B) < (C, D)$) if $A \subseteq C$ and $B \supseteq D$. Likewise, a formal concept (A, B) is a superconcept of (C, D) (expressed as $(A, B) > (C, D)$) if $A \supseteq C$ and $B \subseteq D$.

Relations between concepts can be understood in terms of conceptual neighbors – concepts that are more general or more specific to one another within the concept hierarchy. A concept (A, B) is said to be the lower neighbor, of concept (C, D) if $(A, B) < (C, D)$ such that there is no concept (E, F) that gives rise to $(A, B) < (E, F) < (C, D)$. Likewise a concept (A, B) is said to be the upper neighbor, of concept (C, D) if $(A, B) > (C, D)$ such that there is no concept (E, F) that gives rise to $(A, B) > (E, F) > (C, D)$. Following the running example, the concepts 'abstract paintings with geometric patterns and coarse brush strokes', 'abstract paintings and abstract works with geometric patterns' are all conceptual neighbors of 'abstract paintings with geometric patterns'.

Concepts can also be related in terms of similarity: i.e., certain concepts can be considered conceptually similar based on sharing some common objects and attributes, with the mathematics of such described in Formica [5]. Furthermore, concept similarity provides a fast approximation for identifying a concept s neighbors. In Table 2, for example, for the formal concept 'abstract paintings with geometric patterns', its immediate lower neighbor 'vibrant and abstract paintings with geometric patterns' is identified as its most similar concept. It is also partially similar to the more distant 'energetic and vibrant paintings', more so than the notionally relevant paintings that depict the Illawarra and the almost irrelevant works that depict animal imagery.

These concept similarity metrics are also valid for comparing object-to-concept as well as concept-to-concept relationships, since, according to the mathematics of FCA, formal concepts can represent individual objects called object concepts. For instance, consider the print 'Illawarra Flame Tree and Bowerbird' (Fig. 2) which is, according to the natural language description of its object concept, "an intricate and vibrant print that depicts animal imagery, the Illawarra and nature and has red and blue tones" Using the concept similarity metrics described above, we can observe the multiplicity of contexts that this object can be interpreted in, and determine which concepts are 'most' similar to the artwork (shown below in Table 3).

Table 2. Concepts similar to abstract paintings with geometric patterns

Formal concept, expressed in natural language	**Similarity Score**
vibrant and abstract paintings with geometric patterns	0.80
energetic and abstract paintings with geometric patterns	0.73
abstract paintings	0.60
energetic and vibrant paintings	0.32
paintings that depict the Illawarra	0.16
works that depict animal imagery	0.03

Examining the objects in these contexts and ordering them by relevance gives the ability to induce new objects and draw equivalencies between then.

This provides the basis of how *divergences* work in A Place for Art. Divergent exploration is based on the idea that every turning point within the collection should infer new objects and enlighten new connections. Hence, divergences have two design criteria: a) they should infer new objects based on similarity of an object of interest and b) the resulting pathways should always infer new objects that have not yet been presented previously by prior convergences. This approach avoids repetition and circularity navigating the information space, where the sum of divergences affords a gradual unveiling of the collection by highlighting new works of interest.

Table 3. Concepts similar to "Illawarra Flame Tree and Bowerbird".

Formal concept, expressed in natural language	**Similarity**	**#objects**
intricate prints that depict animal imagery and nature	0.50	2
works that depict animal imagery and have red tones	0.38	2
intricate works that depict nature	0.21	3
works that have blue tones	0.13	8
works that depict animal imagery	0.11	11
vibrant works	0.08	23

Fig. 2. "Illawarra Flame Tree and Bowerbird".

When considering what objects to show in a divergence, all the other objects represented by the 'pivot point' are determined, along with the total set of objects in prior convergences. Based on the object depicted in its pivot point, an object concept is constructed, in which a set of formal concepts containing that object are retrieved. Using concept similarity metrics, the formal concept that is selected is the one that has the highest similarity score containing objects not part of a prior set of convergences. These new objects are then reconstructed as a formal concept, so that any additional attributes are implied from this reduced set, which is then presented as an adjoining pathway from the pivot point.

2 Refections on Divergence/Convergence

The convergent-divergent interaction paradigm is also supported by the theory and philosophy of Formal Concept Analysis in two ways. Formal concepts – just like the human concepts they are modeled on "express subjectivity and emotions" [15]. In the context of museum collections, this provides the ability to model human meaning and thought in the form of sentiments and conjectures; within the A Place for Art, the idea that certain works have 'warm tones' or 'evoke a sense of calm'. It also offers a way of creating inferences and structures from these conjectures without their explicit encoding in other formal knowledge representation schemas.

The second implication concerns the relational qualities of the conceptual structure, a quality best observed from the concept lattice (Fig 1). Wille [14] introduces the notion of conceptual landscapes as a metaphor to describe the inherently spatial properties of human knowledge, drawing parallels to Murray's [10] landscape paradigm. Whereas Murray's perspective refers to spatial qualities of navigating information spaces, Wille's conceptual landscapes describe the way humans produce, communicate and consume knowledge. Yet, like Murray, Wille alludes to the metaphorical adoption of landscape motifs that dictate the way humans interact with information spaces, and shares the view that computers are a medium, rather than a container for the storage and display of data. He argues "The idea of a landscape is becoming increasingly influential in the field of knowledge representation and processing. Especially, the frequently used term of 'navigation' suggests this idea is becoming a leading metaphor. That view is also supported by the development of computers as a medium. This development shows that it is time for explicating the pragmatic landscape paradigm for knowledge processing" (Wille[16]).

From this proposition, Wille defines the practice and discipline of Conceptual Knowledge Processing [4, 14] as a set of techniques that make use of a variety of conceptual structures to augment human activities in knowledge representation, processing and communication. Within this framework, Wille defines the act of identification – the positioning and contextualization of objects, concepts or data elements in relation to other objects, concepts or data elements and exploration – understood as the act of seeking without a goal, or where the item in question is vague or not well known.

3 Conclusion

The mathematization of the convergent-divergent paradigm of navigation and its implementation in A Place for Art highlight the spatial properties of its interaction and of its underlying conceptual structures. These structures recognize the inherently polyvalent nature of knowledge and their interpretation that orbits and interpretation of museum objects [2], and, through the convergent-divergent interaction paradigm, A Place for Art intends to afford creative exploration of these structures in a non-didactic way. Evaluation of the interface among user communities reinforces the success of the approach and these results are described in the presentation at FCA4AI but are to be fully published elsewhere.

References

1. L. Aroyo, N. Stash, Y. Wang, P. Gorgels, and L. Rutledge. CHIP Demonstrator: Semantics-Driven Recommendations and Museum Tour Generation. In K. Aberer, N. Choi, K-S.and Noy, D. Allemang, K. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, editors, *The Semantic Web*, volume 4825 of *LNCS*, pages 879–886. Springer Berlin / Heidelberg, 2007.
2. S. Cairns. Mutualizing museum knowledge: Folksonomies and the changing shape of expertise. curator. *The Museum Journal*, 56, 2013.
3. C. Carpineto and G. Romano. *Concept Data Processing: Theory and Practice*. Wiley, 2004.
4. P. Eklund and R. Wille. Semantology as basis for conceptual knowledge processing. In *Formal Concept Analysis*, pages 18–38. Springer Berlin Heidelber, 2007.
5. A. Formica. Concept similarity in formal concept analysis: An information content approach. knowledge-based systems. *Knowledge-Based Systems*, 21(1):80–87, 2008.
6. R. Godin, R. Missaoui, and A. April. Experimental comparison of navigation in a galois lattice with conventional information retrieval methods. *International journal of Man-Machine Studies*, 38(5):747–767, May 1993.
7. P. Goodale, P. Clough, N. Ford, M. Hall, M. Stevenson, S. Fernando, N. Aletras, K. Fernie, O. Archer, and A. Polo. User-Centred Design to Support Exploration and Path Creation in Cultural Heritage Collections, 2012.
8. P. Krajca, J. Outrata, and V. Vychodil. Parallel recursive algorithm for formal concept analysis. *6th International Conference on Concept Lattices and their Applications*, pages 71–80, 2008.
9. A. Lawson and C. Judd. *A Place for Art*. University of Wollongong Press, 2012.
10. J. Murray. *Inventing the Medium*. MIT Press, Cambridge, MA, 1012.
11. G. Schreiber, A. Amin, L. Aroyo, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, B. Omelayenko, J. van Osenbruggen, A. Tordai, J. Wielemaker, and B. Wielinga. Semantic annotation and search of cultural-heritage collections: The MultimediaN E-Culture demonstrator. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):243–249, 2008.
12. A Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. Finin, K. Thirunarayan, J. Wielemaker, M. Hildebrand, J. van Ossenbruggen, and G. Schreiber. *Thesaurus-Based Search in Large Heterogeneous Collections*, volume 5318, pages 695–708. Springer Berlin / Heidelberg, 2010.
13. M. Skov. *The Reinvented Museum : Exploring Information Seeking Behaviour in a Digital Museum Context*. PhD thesis, The University of Copenhagen, Royal School of Library and Information Science, 2009.
14. R. Wille. Conceptual landscapes of knowledge: A pragmatic paradigm of knowledge processing. *Classification in the Information Age*, pages 344–356, 1999.
15. R. Wille. Formal concept analysis as mathematical theory of concepts and concept hierarchies. In R. Wille B. Ganter, G. Stumme, editor, *Formal Concept Analysis*, volume 3626, pages 47–70. Springer Verlag, 2005.
16. R. Wille. Restructuring lattice theory: An approach based on hierarchies of concepts. In S. Ferre and S. Rudolph, editors, *Formal Concept Analysis*, volume 5548, pages 314–339. Springer Verlag, 2009.
17. T. Wray, P. Eklund, and K. Kautz. Pathways through information landscapes: Alternative design criteria for digital art collections. In *The 34th International Conference on Information Systems (ICIS)*. AIS Press, 2013.