

# Concepts and Collections: A case study using objects from the Brooklyn Museum

Tim Wray and Peter Eklund

`twrap,peklund@uow.edu.au`

School of Information Systems and Technology  
University of Wollongong  
Northfields Ave, Wollongong, NSW 2522, Australia.

**Abstract.** In this paper, we present a browsing framework for digitised cultural collections. Using a data analysis technique called Formal Concept Analysis (FCA), units of thought can be constructed from a series of objects and their tags. FCA can dynamically generate links in between objects and induce a serendipitous browsing experience using a relatively simple data structure. We evaluate the utility and scalability of our approach to a collection of 15,000 objects from the Brooklyn Museum’s collections. We describe how we use natural language processing techniques and external lexical resources to synthesise key terms from museum documentation. We then combine this term extraction process with FCA to effectively demonstrate links between and within collections of objects. In doing so we present a versatile, generalizable term extraction and browsing framework suitable for digital libraries and archives within the art and architecture domain.

## 1 Introduction

Cultural collections are vast, heterogeneous stores of history that are monumental in their representation of human history and expression. Of particular interest are the philosophical notions on how to best represent knowledge within these collections, beginning from the rigid classification hierarchies that are commonly employed in today’s cultural collections to organic, tag-based, associative approaches. Weinberger [1] examines tags as a form of classification, and notes that there are often multiple relationships among objects within a collection, each of which can be meaningful in their own interpretation. He quotes that “trees can be built from leaves” – meaning that sorting and categorisation can be dynamically induced, either from user communities and stakeholders (social tagging) or from the metadata itself without reliance on an imposed classification schema. In effect, sorting, categorising and relating objects can be organic, dynamic and data-driven. When combined with a consistent knowledge representation structure and controlled vocabulary, these relationships can unify multiple, heterogeneous collections.

Large scale cultural heritage projects such as Europeana<sup>1</sup> and Digital NZ<sup>2</sup> are a step in the right direction in unifying and providing accessibility to collections. As a result of projects such as these, there is a large amount of research conducted in making these collections accessible and semantically inter-related. For example, Schreiber et al. [2] investigate approaches towards enhancing and enriching collection metadata and providing semantic annotation and search facilities to large cultural collections. Klavans et al. [3] describe the nuances and challenges of extracting metadata from cultural collections using natural language processing techniques. Work conducted by Trant [4] report on how audiences can contribute new knowledge to collections in the form of social tagging while latter work by Klavans et al. [5] examine how such data could be exploited in order to assist information retrieval and browsing. These authors recognise the importance of deriving meaning from cultural collections. Their research is well aligned with related work in data visualisation, such as the Visible Archive and commonsExplorer projects [6]. Like our project, these works focus on the discovery of patterns and relationships within collections, rather than traditional targeted search.

In our approach, we discover these patterns and relationships by using a data analysis technique called Formal Concept Analysis (FCA). FCA is the mathematization of conceptual thinking – a way of ordering and relating structured units of thought [7]. A *formal concept* denotes a unit of thought and consists of an extension, the objects that compose that thought, and an intension, the attributes, properties and meanings that apply to all of the objects within the extension. For example, when applied to a collection of works, one may be thinking about “Chinese vases with floral patterns” (the intension, or the attributes) or the actual 17 vases (the extension, or objects). In human conceptual thinking, concepts rarely exist on their own, but rather in relation with many other concepts [8] – as a result neighbouring concepts often play an important role in data analysis and communication. For example, it is inevitable that there would be some sort of link between “Chinese vases” and “vases with floral patterns” – these are *superconcepts* of “Chinese vases with floral patterns”, so called because they represent ‘broader’ concepts with a greater set of objects. Dually, concepts such as “Chinese vases with floral patterns from the Qing dynasty” are *subconcepts* – they provide a more narrow, focused view of the collection. These superconcept-subconcept relationships are one of the core mechanisms in which we use to provide associative links between clusters of objects and as such, it drives our framework for browsing digitised collections.

Over 10 years of research in applied FCA has been dedicated towards new approaches of knowledge discovery within collections. Projects such as ImageSleuth and ImageSleuth2 [9] are precursors to the design of the Virtual Museum of the Pacific [10] in which the current framework is derived from. This research assesses the applicability of the browsing framework towards a large data-set of 15,000 objects from the Brooklyn Museum’s collections, using an automated

<sup>1</sup> <http://www.europeana.eu/portal/>

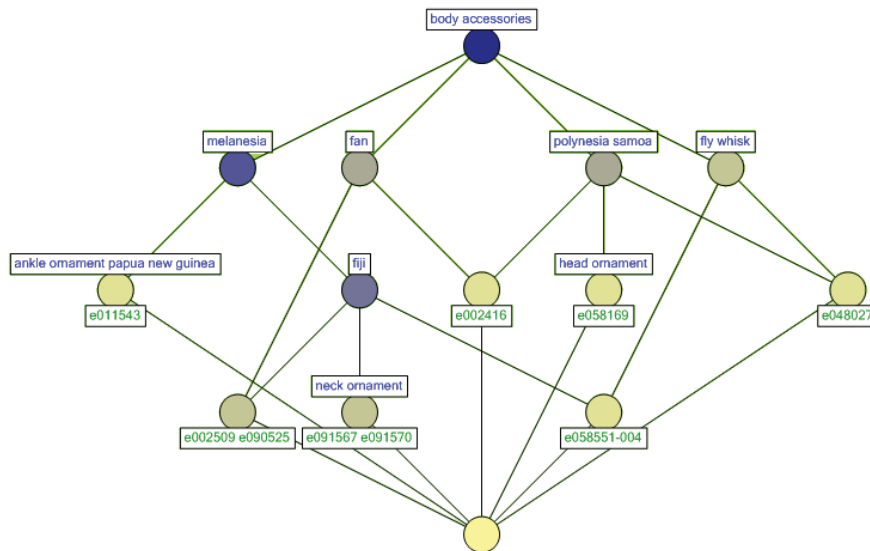
<sup>2</sup> <http://www.digitalnz.org/>

term extraction approach to derive the required key terms for analysis. It refines and assesses the applicability of the content-based retrieval component of the framework, and its contribution lies in its applicability to a large, real world data-sets.

The structure of this paper is as follows: Section 2 provides a brief introduction of Formal Concept Analysis as applied in our case study, and its significance as a tool for linking groups of objects. In Section 3, we describe how we extract key terms from the Brooklyn Museum’s API in order to provide a suitable data structure for analysis. In Section 4, we describe results of our application of Formal Concept Analysis to those terms, highlighting issues with respect to scalability and complexity along with the results of the description of a prototype collection browser. The paper concludes with a discussion on useful applications and extensions of our work.

## 2 Formal Concept Analysis

Formal Concept Analysis (FCA) [7] is a core feature of our framework that is used to derive relationships among objects. Central to the theory of FCA is the notion of the *formal concept*, and its resulting algebraic structure, the *concept lattice*. To clarify the theory of FCA, we will use a Pacific collection of objects as an example.



**Fig. 1.** A concept lattice for a small collection of Pacific objects. Labels above the nodes denote attributes (or tags) and labels below the nodes denote registration IDs from the Museum’s content management system

A formal concept  $(A, B)$  represents a unit of thought, where  $A$  is a set of object identifiers and  $B$  is a set of attributes, or ‘tags’, that describe the objects. For example, the concept “Fijian fans” can be represented by  $(A, B)$  where  $A = \{e002509, e090525\}$  and  $B = \{\text{body accessory, fan, melanesia, fiji}\}$ . Formal concepts can be ordered and arranged in a specialisation hierarchy. A concept  $(A, B)$  is a sub-concept of concept  $(C, D)$  if  $A \subseteq C$  (or equivalently,  $B \supseteq D$ ). Using this definition, more specific concepts have fewer objects and more attributes. For example:  $(A, B) < (C, D)$  where:

$$(A, B) = \{\{e002509, e090525\}, \{\text{body accessory, fan, melanesia, fiji}\}\}$$

$$(C, D) = \{\{e090525, e002509, e058551-004\}, \{\text{body accessory, melanesia, fiji}\}\}$$

The set of all formal concepts, together with the specialisation relation, forms the *concept lattice*. The concept lattice is an algebraic structure that shows hierarchies and relations between formal concepts (Fig. 1). It is derived from the *formal context*, which is a list of objects and their tags, represented as a cross-table (Table 1) and formally denoted as a triple  $\mathbb{K} := (G, M, I)$  where  $G$  is a set of formal *objects*,  $M$  is a set of *attributes* and  $I$  is an *incidence* relation between the objects and the attributes.

**Table 1.** The formal context, or cross table, used to generate the concept lattice in Fig. 1. Note that the core data structure can be expressed as a series of objects and tags.

$\mathbb{K}$	body accessories	fan	head ornament	ankle ornament	fly whisk	neck ornament	melanesia	polynesia	fiji	papua new guinea	samoa
e002509	×	×					×		×		
e090525	×	×					×		×		
e058551-004	×				×		×		×		
e091567	×					×		×	×		
e091570	×					×	×		×		
e002415	×	×						×			×
e002416	×	×						×			×
e058169	×		×					×			×
e058169	×				×			×			×
e011543	×			×			×			×	

In Fig. 1, nodes represent formal concepts. Labels above the nodes represent attributes, (or tags) that describe the object, and labels below the nodes represent the database identifiers of those objects. The set of attributes for a particular formal concept is inferred by gathering all of the attribute labels as

one would traverse upwards on the line diagram, starting from the node represented by the formal concept and ending at the top node. For example, based on the interpretation on this line diagram, one can infer that objects ‘e002509’ and ‘e090525’ are similar to objects ‘e091567’ and ‘e091570’, in that they share common attributes ‘fiji’, ‘melanesia’ and ‘body accessories’ and that they are close to one another. This observation, in part, drives the foundation of the similarity and distance metrics that we use to provide an order ranked list of similar formal concepts for a given object [11].

The similarity metric is a measure based on the number of common objects and the number of common attributes (tags) of two given formal concepts  $(A, B)$  and  $(C, D)$ :

$$\text{similarity}((A, B), (C, D)) := \frac{1}{2} \left( \frac{|A \cap C|}{|A \cup C|} + \frac{|B \cap D|}{|B \cup D|} \right).$$

The distance metric is a measure based on the overlap of the objects and attributes of two concepts, normalised with respect to the size of the formal context, where  $G$  is the total set of objects and  $M$  is the total set of attributes. For two concepts  $(A, B)$  and  $(C, D)$ , the distance metric is as follows:

$$\text{distance}((A, B), (C, D)) := \frac{1}{2} \left( \frac{|A \setminus C| + |C \setminus A|}{|G|} + \frac{|B \setminus D| + |D \setminus B|}{|M|} \right).$$

When combined, these two metrics can be used to provide a list of similar formal concepts to a given object, ordered from ‘most similar’ to ‘least similar.’ As we are comparing formal concepts, a similarity query can derive both matching and nearby objects (e.g. “An American sculpture that depicts youth”) or clusters of objects (e.g. “6 Contemporary sculptures that are made with bronze”). Section 4 of this paper describes how we use these similarity metrics to provide an order ranked list of objects and object clusters from the Brooklyn Museum’s collections. However, in order to do so, we need to build the formal context by extracting key terms from the objects.

### 3 Term Extraction: Building the Formal Context

Term extraction algorithms, such as Yahoo’s Term Extraction Web Service<sup>3</sup>, are commonly employed to assign keywords to documents based on their content. Our term extraction method is built based on the work of Klavans et al. [3] who discuss the application of computational linguistics to museum collections along with current state-of-the-art algorithms developed by Medelyan [12], Frank et al. [13] and Witten et al. [14]. We employ external lexical resources, such as WordNet [15] and the Getty’s Art and Architecture Thesaurus<sup>4</sup> to provide semantic background knowledge for the term extraction process. Like many

<sup>3</sup> <http://developer.yahoo.com/search/content/V1/termExtraction.html>

<sup>4</sup> <http://www.getty.edu/research/tools/vocabularies/aat/about.html>

natural language processing applications, we employ a pipeline architecture for term extraction, shown in Fig. 2.

We source a collection of 15,000 objects using the Brooklyn Museum’s API. These objects are an amalgamation of 12 collections from the museum. The completeness of the object records vary considerably – some objects have full descriptions and interpretive labels to a depth and standard typically found within exhibition catalogues and are often procured for exactly that purpose. These descriptions often provide the cultural context of the object, how it was used, where it came from and its significance. Given the time and cost associated with their research, objects of these descriptions would naturally only occupy a small portion of the collection. Therefore, 1000 objects were selected as objects having *exhibition quality metadata*. Likewise, the entire collection of 15,000 objects were documented, in the very least, with notes and details of its medium, title, culture and classification – denoted as *basic metadata*. As the amount of metadata present within an object determines the kinds (and types) of terms that could be extracted from them, we create two instances of our framework to accommodate these two classes.

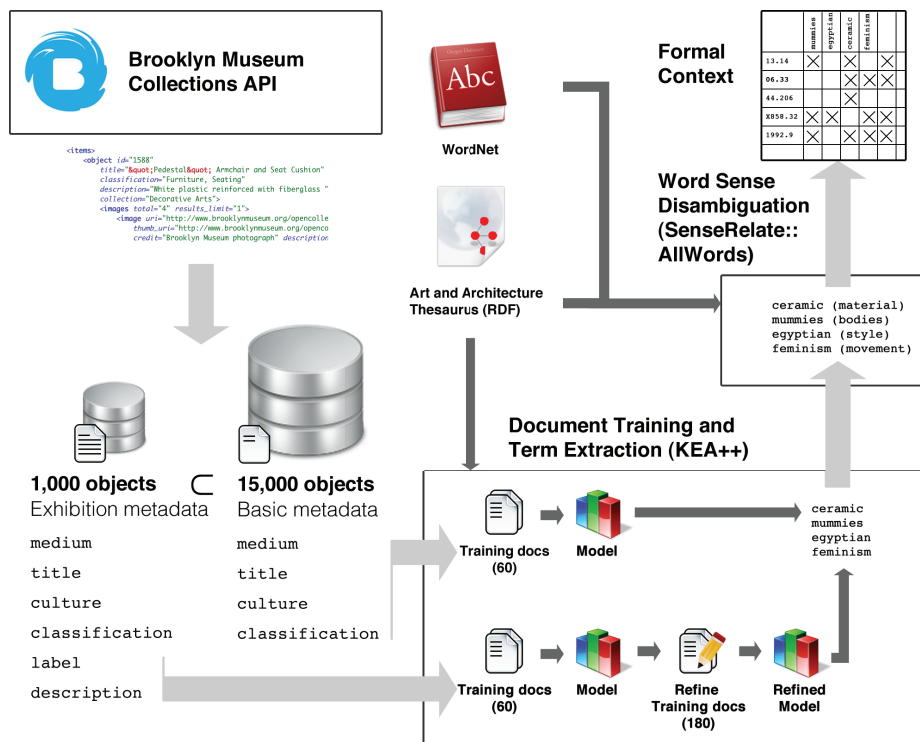


Fig. 2. Overview of the term extraction process used to generate formal contexts, shown anti-clockwise from the top-left

To perform the term extraction, we use a program called KEA++. KEA++ has proven to be a high performance extraction program [12] that combines *keyphrase extraction*: identifying features and prominent keyphrases from a document and *keyphrase assignment*: where terms are selected from a controlled vocabulary using a trained model. We employ the Getty’s Art and Architecture Thesaurus (AAT) as the controlled vocabulary. The AAT contains over 34,800 unique concepts under 33 hierarchies for describing object categories, materials, activities and functions, styles and periods and other abstract phenomena associated with material culture and artworks. It can be used as a single ontology for unifying disparate collections and digital archives. Where appropriate, we use specific hierarchies to perform term extraction on certain types of data fields. For example, the object’s ‘medium’ data field (shown in Fig. 2) would employ only a sub-section of the thesaurus, mainly the ‘physical attributes’ and ‘materials’ hierarchies. This is to reduce the likelihood of a document being assigned an incorrect term due to overstemming (e.g. ‘painting (visual works)’ was often incorrectly assigned instead of ‘paint (medium)’). For basic metadata, each data field (‘medium’, ‘title’, ‘culture’ etc.) was provided with a set of 60 training documents. For the exhibition quality metadata fields, 60 training documents were used to generate a model that produced 180 documents, which were then refined to produce the final model.

For each object record, KEA++ generates a set of candidate terms. However, many of these terms are ambiguous – over 16% of terms extracted from the collections referred to more than one sense within the AAT. For example, the term ‘gold’ refers to two senses of the word, referring to both the material and the color property of an object. As described by Palmer et al. [16], the common linguistic problem of *word-sense disambiguation* is a particularly challenging one. To solve this problem, we adapt a method proposed by Klavans et al. [3] that uses an external algorithm called SenseRelate::AllWords [17]. This algorithm is a Perl module that identifies the correct WordNet sense of each word in a sentence, using the surrounding text as its context. This AAT sense is then selected by performing a word overlap of the definitions of the AAT record and the WordNet sense – the AAT sense with the highest match is assigned to that word.

Once the terms are extracted and disambiguated, we then use them to construct the formal context. As hierarchical term relationships are naturally embodied within FCA, we exploit the broader-narrow relationships within the AAT to enrich the formal context with parent tags so that for example, ‘streetscapes’ → ‘cultural landscapes’ → ‘landscapes’. These hierarchical relations complement the similarity and distance metrics described in Section 2 as these metrics favour objects that share attributes with common parents so that for example, ‘streetscapes’ is notionally similar to ‘suburban landscapes’.

The final step is to prune the formal context in order to reduce its complexity. Although FCA is theoretically robust, applications that employ it for data analysis and communication commonly apply a number of techniques to remove extraneous data points while retaining meaningful representation of its information space [18]. It is also necessary to employ these complexity reduction

measures given the high computational cost of FCA-based operations with respect to the size of the formal context [19]. While more elaborate approaches for reducing complexity in fully formed concept lattices exist [20] [18], our approach needs to rely on more rudimentary measures of complexity reduction as each similarity / distance operation traverses only part of the data-set as required. We use an approach called *context reduction* – it removes rarely occurring tags, which, despite their ‘insignificance’, reduces the size and complexity of the formal context considerably. This makes sense as the “aboutness” of the objects are dictated by the attributes that they have in common, rather than the attributes that they don’t have in common. We remove tags that do not belong to a threshold percentage of objects, with the threshold value set by default at 0.05%.

## 4 Results and Scalability of our Approach

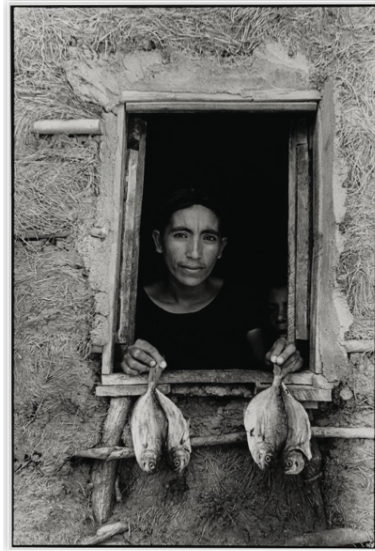
A key design requirement of our framework is to induce an explorative browsing experience by computing the similarities and differences between objects, deriving natural pathways within collections, and highlighting key concepts – showing collections within collections. Furthermore, its implementation needs to be scalable with a performance requirement to suit real time interactive browsing over the Web. To show the results of our work, we have developed a light-weight prototype collection browser<sup>5</sup>, shown in Fig. 3. The browser shows a detailed catalogue description, with links to conceptually similar objects and object clusters.

In the example shown in Fig. 3, the extracted terms of the artwork: {photographs, rituals, women, power, fishing} are used to compute the following similar formal concepts, order ranked from most similar to least similar:

- { women, photographs, power }  
2 objects (Similarity: 0.55, Distance: 0.99)
- { women, fishing }
- 2 objects (Similarity: 0.45, Distance: 0.99)
- { women, power }
- 5 objects (Similarity: 0.30, Distance: 0.99)
- { photographs, power }
- 6 objects (Similarity: 0.28, Distance: 0.99)
- { women, photographs }
- 7 objects (Similarity: 0.27, Distance: 0.99)

<sup>5</sup> Two prototype collection browsers are publicly available for the two collections:  
1,000 objects with exhibition quality metadata:  
[http://epoc2.cs.uow.edu.au/brooklyn\\_r\\_1000\\_ws/similarity/](http://epoc2.cs.uow.edu.au/brooklyn_r_1000_ws/similarity/)  
15,000 objects with basic metadata:  
[http://epoc2.cs.uow.edu.au/brooklyn\\_m\\_15000\\_ws/similarity/](http://epoc2.cs.uow.edu.au/brooklyn_m_15000_ws/similarity/)





## Cuatro pescaditos (Four Fishes), Juchitán, Oaxaca

Graciela Iturbide

Graciela Iturbide is one of the best-known Mexican photographers of the last four decades. The images in this gallery represent series from different parts of Mexico, of which the most important is her breakthrough photoessay Juchitán of the Women (1979-86). In a documentary style notable for its humanistic grace, the series focuses on the indigenous Zapotec people in the town of Juchitán, in southeastern Mexico, where women dominate all aspects of social life, from the economy to religious rituals. The most emblematic image of the series, Our Lady of the Iguanas, shows the power and dignity of a Zapotec woman, who carries on her head live iguanas that form a bizarre crown. Four Fishes shows a woman displaying fish for sale from the private space of her home, the clay and straw of the wall echoing the scales of the fish.

Like her teacher, the photographer Manuel Alvarez Bravo (at one time the husband of Lola Alvarez Bravo, whose work hangs nearby), Iturbide portrays Catholic traditions intertwined with pre-Hispanic rites and superstitions, showing a culture in constant flux. Approaching her subjects directly and frontally, Iturbide represents a dreamlike reality with great compassion, or, to use the artist's own word, "complicity."

### related objects



A Little Taste Outside of Love  
Mickalene Thomas

This photograph depicts women and is associated with power



The Fishing Party  
Winslow Homer

This work is associated with fishing

### related object clusters



3 works that depict women and are associated with power



4 photographs that are associated with power



5 photographs that depict women

Fig. 3. Screenshot of the prototype collection browser with links to conceptually similar objects and object clusters

From these results, the algorithm derives all objects as a unique set<sup>6</sup>, and clusters them according to their member concept. Each similar formal concept consists of the object we are comparing plus member objects of that same formal concept, i.e., the first two results indicate individual objects tagged { women, photographs, power } and { women, fishing }, respectively. These objects are presented as ‘related objects’ as shown in Fig. 3. Other formal concepts are shown as ‘object clusters’ with a series of thumbnails indicating other focal points of interest within the collection.

We employ natural language labels to describe the objects and object clusters. These labels are generated from the tags from the formal concepts’ tags, while their semantics are inferred from their hierarchical membership within the AAT. For example, for a given set of tags { women, photographs, power }, one may assume that they describe photographic works that depict women and are also associated with power, given that ‘women’ exists in the ‘Agents’ hierarchy; ‘photographs’ exists in the ‘Visual Works’ hierarchy and that ‘power’ exists in the ‘Associated Concepts’ hierarchy. Within each hierarchy, its member tags indicate what aspect of an artwork they describe. However, a problem with this approach lies in the inherent ambiguity of whether a term is object-oriented (describing the object itself, its properties) or subject-oriented (describing what the work is about or what it depicts) [21] [22]. In some cases, terms such as ‘water’ could refer to both a work that is *made with* water or a work that *depicts* water features – an apparent shortcoming of many tag based systems. Currently, the AAT only recognises water in the former sense, and further curation of these sorts of tags may be necessary to prevent these semantic ambiguities.

Performance and scalability are important factors for real world implementation. As theorized by Carpineto and Romano [19], the computational cost of FCA-based operations increases as the size of the formal context gets larger. The results of our performance testing have indicated that dynamically performing these computations is unsuitable for a collection of more than 200 objects, with average query times approximating 60 seconds on the full collection of 15,000 objects. To solve this problem, we have adopted a caching method where similar formal concepts are pre-computed and cached with each object record. The system updates these caches as new objects are added, or their tags change.

## 5 Conclusion and Future Work

We have presented a term extraction and browsing framework as applied to the Brooklyn Museum’s collection, using objects and tags as a core data structure. We have also developed a prototype browsing application to demonstrate our framework. It is scalable to a collection of 15,000 objects and it can dynamically generate links to neighbouring objects and object clusters, expressed in natural language. With a focus on *concepts* rather than *objects*, it follows a contemporary

<sup>6</sup> Similar formal concepts have a high overlap of common objects. Based on user feedback, we’ve adopted a design decision to not show duplicate objects within the UI.

data-driven approach of collections browsing, and it can be suitably adopted for experiments and applications in collections visualisation.

Given that we use a common vocabulary for tagging objects, this work could be extended to cover multiple collections from different institutions with an assessment on if or how our framework could scale, along with how it can adapt to the varying kinds of metadata each collection presents. Tags present a simple and versatile data structure that can be provided or derived from free text. *Semantic tagging* [23] introduces an interesting possibility of solving the previously mentioned semantic ambiguity problem described in Section 4.

Social tagging in museum collections is gaining traction and has proven to add worthwhile community knowledge to museum collections [4] – for example, the Brooklyn Museum provides programs such as “Tag You’re It!”<sup>7</sup>, and these social tags are commonly used on their website to assist searching and browsing. As an extension of our work, leveraging social meta-data not only closes gaps in museum documentation and opens up interpretation to visitors, but it can also induce dynamic relationships among objects, allowing for a self-evolving and community-driven approach to the display and interpretation of collections.

## References

1. Weinberger, D.: Taxonomies to tags: From trees to piles of leaves. Release 1.0 **23**(2) (2005)
2. Schreiber, G., Amin, A., Aroyo, L., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Omelayenko, B., van Osenbruggen, J., Tordai, A., Wielemaker, J., Wielinga, B.: Semantic annotation and search of cultural-heritage collections: The multimedial e-culture demonstrator. *Web Semantics: Science, Services and Agents on the World Wide Web* **6**(4) (2008) 243 – 249 *Semantic Web Challenge 2006/2007*.
3. Klavans, J., Sheffield, C., Abels, E., Lin, J., Passonneau, R., Sidhu, T., Soergel, D.: Computational linguistics for metadata building (climb): using text mining for the automatic identification, categorization, and disambiguation of subject terms for image metadata. *Multimedia Tools and Applications* **42** (2009) 115 – 138 10.1007/s11042-008-0253-9.
4. Trant, J.: Tagging, Folksonomy and Art Museums: Results of steve.museum’s research. Technical report, University of Toronto (2009)
5. Klavans, J., Stein, R., Chun, S., Guerra, R.D.: Computational Linguistics in Museums: Applications for Cultural Datasets. In Trant, J., Bearman, D., eds.: *Museums and the Web 2011: Proceedings, Archives and Museum Informatics* (2011)
6. Hinton, S., Whitelaw, M.: Exploring the digital commons: an approach to the visualisation of large heritage datasets. [http://www.bcs.org/upload/pdf/ewic-ev10\\_s3paper2.pdf](http://www.bcs.org/upload/pdf/ewic-ev10_s3paper2.pdf) (2010)
7. Wille, R., Ganter, B.: *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, Berlin (1999)
8. Wille, R.: Formal concept analysis as mathematical theory of concepts and concept hierarchies. In Ganter, B., Stumme, G., Wille, R., eds.: *Formal Concept Analysis*. Volume 3626 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg (2005) 47–70

<sup>7</sup> [http://www.brooklynmuseum.org/opencollection/tag\\_game/start.php](http://www.brooklynmuseum.org/opencollection/tag_game/start.php)

9. Eklund, P., Ducrou, J., Wilson, T.: An intelligent user interface for browsing and search MPEG-7 images using concept lattices. In: Proceedings of the 4th international conference on concept lattices and their applications. LNCS 4923, Springer-Verlag (2006) 1–22
10. Eklund, P., Wray, T., Goodall, P., Bunt, B., Lawson, A., Christidis, L., Daniels, V., Olffen, M.V.: Designing the Digital Ecosystem of the Virtual Museum of the Pacific. In: 3rd IEEE International Conference on Digital Ecosystems and Technologies, IEEE Press (2009) 805–811
11. Saquer, J., Deogun, J.S.: Concept approximations based on rough sets and similarity measures. In: Int. J. Appl. Math. Comput. Sci. Volume 11. (2001) 655 – 674
12. Medelyan, O.: Automatic keyphrase indexing with a domain-specific thesaurus. Master’s thesis, University of Waikato (2005)
13. Frank, E., Paynter, G., Witten, I., Gutwin, C., Nevill-Manning, C.: Domain-specific keyphrase extraction. In: Proceedings of the 16th International Joint Conference on Artificial Intelligence, San Francisco, CA, Morgan Kaufmann (1999) 668 – 673
14. Witten, I., Paynter, G., Frank, E., Gutwin, C., Nevill-Manning, C.: Kea: Practical automatic keyphrase extraction. In: Proceedings of the 4th ACM Conference on Digital Libraries, Berkeley, CA, ACM Press (1999) pp. 254 – 255
15. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA, USA (1998)
16. Palmer, M., Ng, H.T., Dang, H.T.: Word sense disambiguation: algorithms, applications and trends. In Edmonds, P., Agirre, E., eds.: Text, Speech and Language Technology. Kluwer Academic Publishers, Netherlands (2003)
17. Pederson, T., Kolhatkar, V.: Wordnet::senserelate::allwords: a broad coverage word sense tagger that maximizes semantic relatedness. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Demonstration Session. NAACL-Demonstrations ’09, Association for Computational Linguistics (2009) 17 – 20
18. Kuznetsov, S., Obiedkov, S., Roth, C.: Reducing the representation complexity of lattice-based taxonomies. In Priss, U., Polovina, S., Hill, R., eds.: Conceptual Structures: Knowledge Architectures for Smart Applications. Volume 4604 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2007) 241 – 254
19. Carpineto, C., Romano, G.: Concept Data Analysis: Theory and Applications. John Wiley & Sons (2004)
20. Stumme, G., Taouil, R., Bastide, Y., Lakhal, L.: Conceptual clustering with iceberg concept lattices. In: Proceedings of GI–Fachgruppentreffen Maschinelles Lernen’01. Volume 763., Universität Dortmund (2001)
21. Chen, H.: An analysis of image retrieval tasks in the field of art history. Information Processing and Management **37**(5) (2001) 701 – 720
22. Choi, Y., Rasmussen, E.M.: Searching for images: the analysis of users’ queries for image retrieval in american history. Journal of the American Society for Information Science and Technology **54**(6) (2003) 489 – 511
23. Marchetti, A., Tesconi, M., Ronzano, F.: Semkey: A Semantic Collaborative Tagging System. In: Proceedings of the WWW Workshop on Tagging and Metadata for Social Information Organisation. (2007)