

1-1-2010

## Information retrieval and social tagging for digital libraries using formal concept analysis

Peter W. Eklund  
*University of Wollongong*, peklund@uow.edu.au

Peter Goodall  
*University of Wollongong*, pgoodall@uow.edu.au

Timothy Wray  
*University of Wollongong*, twray@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/infopapers>



Part of the [Physical Sciences and Mathematics Commons](#)

---

### Recommended Citation

Eklund, Peter W.; Goodall, Peter; and Wray, Timothy: Information retrieval and social tagging for digital libraries using formal concept analysis 2010, 1-6.  
<https://ro.uow.edu.au/infopapers/1501>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# Information retrieval and social tagging for digital libraries using formal concept analysis

## Abstract

This paper reports a novel semantic web application developed to deliver a collaborative tagging system for a digital on-line museum. The key features of our application – called the Virtual Museum of the Pacific – concern the browsing and retrieval interface based on Formal Concept Analysis, the extensible distributed data model to support collaborative tagging and its web services implementation.

## Keywords

Information, retrieval, social, tagging, for, digital, libraries, using, formal, concept, analysis

## Disciplines

Physical Sciences and Mathematics

## Publication Details

Eklund, P. W., Goodall, P. & Wray, T. (2010). Information retrieval and social tagging for digital libraries using formal concept analysis. 2010 8th IEEE-RIVF International Conference on Computing and Communication Technologies: Research, Innovation and Vision for the Future, RIVF 2010 (pp. 1-6). Piscataway, New Jersey, USA: IEEE.

# Information Retrieval and Social Tagging for Digital Libraries using Formal Concept Analysis

P. Eklund, P. Goodall, T. Wray  
 Centre for Digital Ecosystems  
 University of Wollongong, Australia  
 peklund,twray@uow.edu.au,pjgoodall@gmail.com

**Abstract**—This paper reports a novel semantic web application developed to deliver a collaborative tagging system for a digital on-line museum. The key features of our application – called the Virtual Museum of the Pacific – concern the browsing and retrieval interface based on Formal Concept Analysis, the extensible distributed data model to support collaborative tagging and its web services implementation.

## I. INTRODUCTION

The Virtual Museum of the Pacific (VMP) is designed around the use of Formal Concept Analysis [20] to organize digital library content. It creates an associative network based on semantic themes according to a museological view of content. The system allows the extensible creation of new views, the addition of new attributes and contains a social tagging interface that encourages the Museum’s stakeholders to annotate content and build custom views of the collection according to their communities of interest.

While similar efforts with social tagging in Museums are well known [1], [18], our results present an alternative approach, specifically for interaction with the Australian Museum’s Pacific collection leveraging metadata tags from a collection management system. The Virtual Museum of the Pacific is implemented as a rich internet application that marshals Web Services (WS) to deliver a rich user experience from a distributed and heterogeneous store including relational database, high resolution image files, folksonomy, wiki text, video and audio content. The effect is to create an end-to-end digital ecosystem [9] with access control and data model that accommodates formal metadata and informal taxonomy, distributes rich media and captures social tagging content.

The paper is structured as follows. First, we provide some preliminaries of Formal Concept Analysis (FCA). Our paper surveys some of the literature leading to the VMP design. We then explain the data providence of the objects within the museum collection and discuss some of the issues in translating it in a format suitable for explorative browsing using Formal Concept Analysis. We describe some of the browsing and tagging features of the VMP, and then conclude with the results of the first round of stakeholder evaluations.

### A. Formal Concept Analysis

Formal Concept Analysis [20] is a well-known technique for data analysis that involves the synthesis of a formal concept as a collection of objects that exhibit a common set of attributes.

TABLE I  
 THE FORMAL CONTEXT  $\mathbb{K} := (G, M, I)$  CONTAINING INFORMATION ABOUT THE OBJECTS ( $G$ ) AND THEIR ATTRIBUTES ( $M$ ) FOR THE “BODY ACCESSORY” SCALE.

$\mathbb{K}$	body accessories	ankle ornament	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	×			×			×			×	

Formal Concept Analysis (FCA) offers a formalization of concepts understood in the philosophical tradition: a concept is a unit of thought constituted by its extension (a set of objects) and its intension (a set of attributes). At its simplest level of understanding FCA is an unsupervised machine learning technique that induces a mathematical lattice through algorithms that examine a set of objects and their attributes. A formal context  $\mathbb{K} := (G, M, I)$  is a triple 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.  $I \subseteq G \times M$  is a binary relation where  $(g, m) \in I$  is read “object  $g$  has attribute  $m$ ” and is often written as  $gIm$ . A formal context can be represented as a cross-table where the rows represent  $G$ , the columns represent  $M$  and the incidence relation  $I$  is represented by a series of crosses as shown in Tab. I.

A formal concept of this context can be derived by taking an object and collecting all attributes describing it, then collecting all objects described by those attributes. For instance, taking the object  $e002509$ , let  $B$  be its attributes,  $B = \{\text{body accessories, ankle ornament, melanesia, fiji}\}$ . Now, let  $A$  be the objects which have all

attributes in  $B$ .  $B = \{e002509, e090525\}$  so the formal concept  $(A, B)$  is  $\{\{e002509, e090525\}, \{\text{body accessories, ankle ornament, melanesia, fiji}\}\}$ .

Concepts can be arranged in a hierarchy. A concept  $(A, B)$  is a sub-concept of another concept  $(C, D)$  if  $A \subseteq C$  (or equivalently  $B \supseteq D$ ). Via this definition, more specific concepts have fewer objects and more attributes, e.g.  $(A, B) < (C, D)$  where  $(A, B) = \{e002509, e090525\}$   $\{\text{body accessories, ankle ornament, melanesia, fiji}\}$  and  $(C, D) = \{e090525, e002509, e058551-004\}$   $\{\text{body accessories, melanesia, fiji}\}$ . Further, let  $(G, M, W, I)$  be a many valued context and for each  $m \in M$  let  $\mathbb{S}_m := (G_m, M_m, I_m)$  be a scale for  $m$ . The *derived context* of  $(G, M, W, I)$  with respect to *plain scaling* with the scales  $(S_m | m \in M)$  is then  $(G, N, J)$  where:

$$N := \prod_{m \in M} \{m\} \times M_m \text{ and}$$

$$(g, (m, n)) \in J \Leftrightarrow \exists w \in W (g, m, w) \in I \text{ and } (w, n) \in I_m.$$

Conceptual scales [10] are a powerful tool that can be used to store views that partition data being analysed. For example, Fig. I is a sub-content combining two conceptual scales for “body accessory” and “origin”: this can be seen by the fact that “body accessory” is a attribute common among all the objects in the context table and that the mutually exclusive attributes “polynesia” and “melanesia” appear in each of the objects in the context. Within a system based on Formal Concept Analysis multiple views can be stored and combined to effectively query the data, combining the “body accessory” and “origin” scales results in the line diagram of the concept lattice (shown in Fig. I) and there are 2 of 29 predefined conceptual scales used in the Virtual Museum of the Pacific which include scales such as “indigenous names”, “body enhancements”, “clothing”, “container”, “hunting fishing weapon”, “pastime”, “raw material”, “spiritual”, “tool”, “textile and equipment” etc. Other scales (called perspectives in the UI) can be defined by the stakeholders depending on their interests. The scales can be thought of as facets of the domain which (when deployed) produce a computable sub-context of the cross-table.

### B. Digital Library Annotation, Classification and Semantics

From our experience with the Australian Museum’s (AM) ethnographic collection, we observe that without the initial classification, performed over many years by curatorial staff, tens of thousands of objects would be unreachable. Therefore, an initial formal starting-point helps communities of interest create their own contextual semantic annotations over diverse subsets of the collection.

By treating annotations as important secondary documents [6], we can include historical provenance records: descriptive cards, registers, computerised collection management records, photographs and multimedia. In some contexts the notion of a ‘primary document’ may eventually lose currency as scholarly descriptions become more important than the physical objects themselves.

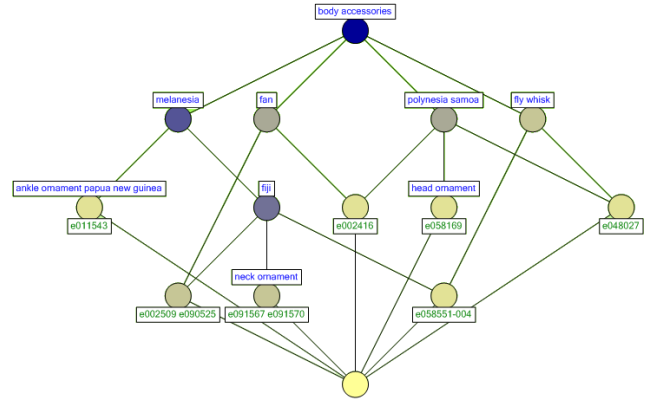


Fig. 1. The line diagram  $(\mathfrak{B}(\mathbb{K}))$  of the cross-table presented in Table. I

The AM reverted from initial taxonomic complexity where objects were effectively ‘lost’ in the classification system as the principle of uncertainty is compounded by the breadth and complexity of the taxonomy, this means that few attempts at classification an object produce the same result. As a result the museum’s formal classification schema is now simple and less error-prone. Social tagging may produce surprisingly stable structures [11], which may lead to emergent and dynamic taxonomies [2], [4], [16] although there is an open problem in ensuring that terms are used consistently within the tagging process – previous applications that use unstructured and non-normalised meta-data to construct formal contexts have resulted in concept lattices that are almost unusable for navigation and search [8]. However term stemming, the use of and lexical databases such as WordNet [17] can be used overcome the problem of synonymous tags. It will always be a challenge to curate ‘correct metadata’ – perhaps it is best to accept that description is a product of socio-historical mechanisms and will always be contested and dynamic. Modern analysis may help develop social annotations as a equal partner to formal classification – especially where paid resources are scarce and enthusiasm abounds.

Digital Libraries should be a place for sharing and developing knowledge, through enabling enhancement and manipulation of resources and metadata in a seamless way [13]. The representation of these complex semantic systems as working IT architectures will be a challenge for some time [12], [14].

### C. Data Provenance

The Virtual Museum of the Pacific is a prototype social media system so the selection of a suitable set of 427 objects (from about 60,000) from the Australian Museum’s Pacific Collection is an important part of the process. The project’s anthropologist selected the objects based on (i) their knowledge of Pacific material culture in general (ii) knowledge of the specific collection (iii) with the aim to create a representative ‘sample’ of Pacific objects that are relevant and interesting for community members, researchers and general public. Further, the selection represents (1) a number of overlapping categories

which have similar attributes, but may come from a variety of geographical areas (2) objects constructed from a variety of materials and techniques, so that the descriptors have enough variety, as well as overlap. A further constraint is that moving objects in preparation for photography involves at least two people so objects that are too large, too fragile or too heavily are excluded. In short the criteria for the selection of objects from the Australian Museum’s Pacific Collection is based on the following: (i) the objects cultural relevance; (ii) ensuring a variety of objects throughout Pacific (such as headrests, fans, musical instruments, body adornment); (iii) including a variety of materials and techniques; (iv) anticipating interest from stakeholders; (v) the relevance of certain objects for our collection (rarity, identity of the donor).

#### D. Classification and Ontology

The metadata used by the Virtual Museum of the Pacific for navigation and discovery within the collection is imported from the Australian Museum’s Collection Management System (CMS). The current CMS is the third effort by the Museum to computerize its records of the Pacific Collection.

To understand the evolution of the Pacific Collection’s metadata we give an overview of the life cycle of records. The Museum acquired the objects in its Pacific Collection from many sources over the last 150 years. The process of adding an object to the collection is reasonably uniform and involves creating a registry entry to associate collection metadata with the object, and instantiates a registration number, i.e.. the object names in Tab.I, e002509, e090525 etc are registry numbers for some of the objects in the Pacific Collection. At some later point in time an index card system was created which included the object’s provenance, more detailed descriptive text, and (on its reverse) the object’s physical measurements. Later again, as objects were added to one of the three generations of CMS used by the Museum, objects were further described. During this phase, a simple, practical corporate taxonomy was developed. The Museum’s *Archaeology and Anthropology* taxonomy is two-level, with 27 categories and 709 item types distributed across those categories. The taxonomy describes objects in the collection using an organizational warrant, by which we mean we mean that it is ‘warranted’ within the ‘organizational’ context of the Museum. It is common for older museums with large natural history and anthropological collections to have developed taxa in this way, Colomb [5] provides good coverage on the relevant issues surrounding the design of classification systems.

## II. INTERACTION AND COMPLEXITY

From information collected during preparation of the 427 objects for the prototype of the VMP, about 70% of the objects in the Pacific Collection have been entered into the CMS (although all objects have an index card) and nearly all objects need metadata cleaning to bring them up to a uniform exhibition standard. This involves normalizing spelling and thesaurus checking, for instance testing whether “mother of pearl” or “pearl shell” should be used or whether a “dagger” should be

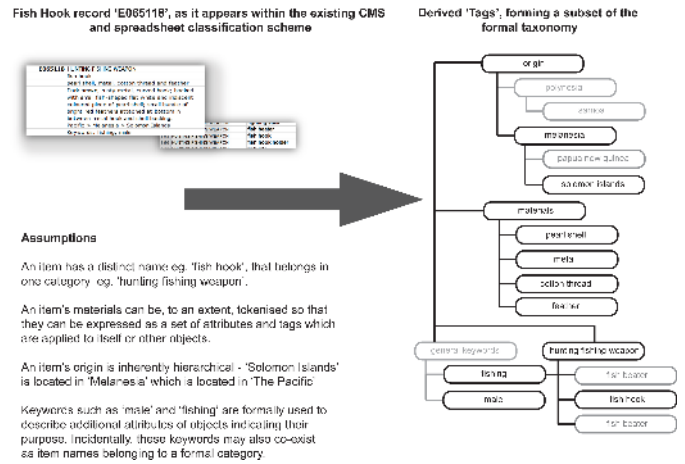


Fig. 2. Inferred taxonomy from existing Australian Museum data model.

tagged as such or with a preferred term “knife”. About one hour’s effort per object is required for basic metadata cleaning, and another hour to write an interpretive label (like a descriptive card in a museum exhibition case). Additionally, a process must be designed and checked where ‘tags’ are extracted from narrative-like data as it appears within the existing CMS to a formal taxonomy, where rules are based on inferences derived from the existing Museum thesaurus along with contextually placed keywords and terms such as ‘materials’ or ‘indigenous names’ within the objects’ metadata fields. Fig. 2 demonstrates this process. A rigorously implemented control for quality assurance must be implemented and maintained to ensure that the metadata entered within the existing CMS is compatible with this process designed for extracting concrete, uniform and hierarchical tags into the Virtual Museum of the Pacific. So, while the metadata adds enormously to the value of an object for research and Web-based exploration, there are significant costs in establishing an adequate information base for it. A key aspect of the VMP is costing the information acquisition, metadata cleaning necessary to bring the entire 60,000 objects of the Pacific Collection in the Museum to exhibition standard.

#### A. Complexity

Recall, a *formal context*  $\mathbb{K} := (G, M, I)$  is a triple 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. The formal context  $\mathbb{K}$  can be pictured as a cross-table and since the complexity of computing the concept lattice is  $O(|C||M|(|G| + |M|))$  [3] (where  $C$  is the number of formal concepts), so too a fundamental concern in computing concept lattices is reducing  $|M|$ . Also, since the cardinality of the object set  $|G|$  is fixed by the size of the objects collected from the Virtual Museum of the Pacific, complexity is managed by reducing  $|M|$ , the size of the set of attributes in  $\mathbb{K}$ . This is one important motivation for effort spend normalizing object metadata.

However, to reiterate the design notes from Section I-B, the Virtual Museum of the Pacific approach is to compute

only the lattice neighbor of the current formal concept so  $C$  is not the the number of formal concepts in the complete lattice but much smaller. Computing a formal concept from an object involves taking its attribute set and finding all associated objects and symmetrically for attribute set. This can be performed quickly using two relations keyed on object and attribute ids. However, we also compute the lower and upper neighbors of the formal concept which is worst case in the order of the number of attributes, so the overall algorithm is  $O(|G||M||G + M|)$ . With the Virtual Museum of the Pacific currently configured with 427 objects and 1198 attributes the impact of this complexity result on system performance presents no difficulty. Further reducing complexity is the use of conceptual scales (or perspectives) which induce sub-contexts that relieve complexity in both object and attribute dimensions.

The issue of scaling the Virtual Museum of the Pacific to 60,000 objects presents issues that will need to be considered to reduce the overall complexity of computing formal concepts. These include (i) a technique called ‘context clarification’ [19] where both objects and attributes that always coexists are collapsed to a singularity; (ii) pre-computation of structures involving partitioning data into sub-contexts using the existing 15 pre-defined conceptual scales (“general keyword”, “indigenous names”, “archaeology”, “body enhancements”, “clothing”, “container”, “hunting fishing weapon”, “pastime” etc). This divide and conquer approach subsequently allows for the (iii) distribution of workload by computing formal concepts on each sub-context and then using the algorithms developed for large virtual file system and Formal Concept Analysis [15] or (iv) apply other (as yet unresearched) approaches involving Formal Concept Analysis and data/task partitioning and distributed algorithms.

In summary, within the present scope of the Virtual Museum of the Pacific complexity is manageable but further distributed systems architectures will need to be explored for a VMP with 60,000 objects in its scope. In practice, fully automated data and knowledge acquisition over all 60,000 objects is impossible and the nature of curatorial processes involved provide a long lead-time in which to solve scalability issues.

### B. Browsing by Attributes/Tags

One of the Virtual Museum interfaces is shown in Fig. 7. By showing upper and lower neighbors of the current concept and allowing navigation to these concepts, users can refine or generalise their position in the information space. Because of the algorithms used in Formal Concept Analysis, no tag will appear unless it results in an object set, this follows from the fundamental theorem in Formal Concept Analysis that the concept lattice is complete with the top of the lattice representing an extension containing the set of all objects and the bottom of the concept lattice containing no objects or objects processing all attributes. This means it is impossible to navigate via the attribute controls to a concept with no result in its extent (or object set).



Fig. 3. The control vocabulary and pick list as the user types the text ‘kava’.



Fig. 4. The resulting formal concept from Fig. 3 shows the association rule induced by the search with the text ‘All kava bowl objects are also container, pastime, wood and food preparation.’

### C. Search and Inference

Navigation via control tags provides a complete vocabulary for searching the VMP but there are many tags (1198 in total) and because the tags are extensible the system can support many more over time. The system therefore needs to support a direct search capability. This is shown in Fig. 3 where the user searches for ‘kava’ and the resulting completions from the control are displayed. An important design idea is that a user should never be able to navigate to an empty result set so only tags that are known to describe objects are auto-completed. This functionality will need to be extended and adapted as the need to use text search against wiki content, which does not form part of the control vocabulary, emerges. At this point we plan to use standard text search but have also developed a theoretical framework for relevance ranking for formal concepts in Formal Concept Analysis described in [7].

A key feature of browsing with Formal Concept Analysis is that the system can infer the maximum set of attributes that satisfy search criteria. In Fig. 4 we can see that all ‘kava bowls’ also include the attributes ‘container’, ‘pastime’, ‘wood’ and ‘food preparation’. The maximal feature clustering ability of Formal Concept Analysis allows association rules to be inferred in the VMP which aids learning from the collection. An individual who knows nothing of ‘kava bowls’ can infer that all are wooden food containers related to a pastime.

### D. Query by Example

In addition to established methods such as keyword search and inclusion of attributes that move the view to upper/lower neighbors, the VMP includes a query-by-example function. This is shown in Fig. 5 where we see the macro view of a specific necklace from the Pacific Collection. The associ-



ated tags for the image are shown above the image ‘body accessories’, ‘cotton’, ‘fiji’, ‘melanesia’, ‘neck ornament’, ‘waseisei’, ‘wasekaseka’, ‘whale teeth’.

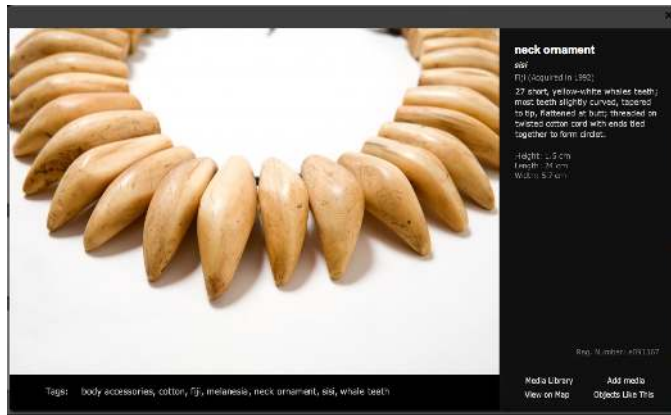


Fig. 5. Macro view of the a necklace object from the Pacific Collection. The text above the image lists the attributes associated with the object. Below the image is the ‘Objects like this’ control which activates query by example and the ‘Map’ control which uses the Google map Web service to show the origin of the object based on its geocode.

Also note in Fig. 5 a perspective (or facet) called ‘body accessory’; is checked, this can be seen by ‘body accessory’; being in the attribute set for the image. The ‘body accessory’ perspective provides a context for the query-by-example search by selecting the ‘objects like this’ link in Fig. 5 which activates the query by example Web service resulting in Fig. 6.



Fig. 6. The results of query-by-example (‘Objects like this’) Web service from Fig. 5. As is evident all the results are necklaces.

### E. Faceted Browsing

We have already introduced the idea of facets in the VMP via the ‘body accessory’ example presented for query-by-example search. Perspectives are pre-defined conceptual scales that can be combined to define the attribute set that induces a sub-lattice that forms a reduced information space and a more focused thematic view of the collection data. Fig. 7 shows two facets combined, namely ‘hunting fishing weapon’ and ‘origin’. In this case the user has navigated to the formal concept induced by the attributes ‘fish hook’ and ‘polynesia’ (polynesian fishhooks) that induces the three objects displayed. The attributes induced as lower neighbors of the formal



Fig. 7. Shows two facets (called perspectives – left) activated ‘hunting fishing weapon’ and ‘origin’. The screenshot also shows the upper and lower neighbor attributes and the current formal concept, the three images, induced by polynesian fishhooks.

concept shown are ‘samoa’ and ‘cook islands’, these being the only attributes that can be used to drill down and reduce the display set with this combination of facets/attributes.

### F. Social Tagging

The Virtual Museum of the Pacific is supplemented with the ability to tag and annotated objects (images and their metadata) using a control vocabulary and a wiki. Further, a set of management interfaces allows users to create “perspectives” that correspond to Formal Concept Analysis’s concept scales described in the earlier section. In addition the user can edit and add new attributes that can then be used to create further perspectives. New objects can also be added allowing the virtual collection content to grow. The program therefore represents an extensible museum content management system with a flexible mechanism for browsing and exploring the set of objects. Its social media capabilities give a rich Web 2.0 interface that we have argued is a web-based digital ecosystem [9].

## III. PRELIMINARY EVALUATION RESULTS

A preliminary evaluation of the VMP took place at the Australian Museum which examined system usability along with the potential for it to be a social platform for collection access and collaborative annotation. Interviews with 16 staff members with backgrounds of collection management, anthropology, education and senior management took place where they were ask to perform scripted tasks within the VMP. The evaluation was directed at staff experience of the system to date, technical issues, particularly the navigation of the VMP application, and at gaining prospective views on the VMP as a virtual domain of the Australian Museum’s Pacific collections.

Participants of the evaluation stated that the system needs to implement non-hierarchical, didactical ways to keep the visitor interested, and to acknowledge the short attention spans of its users. The intention of using Formal Concept Analysis as a means of navigating the collection would theoretically present a non-hierarchical representation of formal concepts and thus promote explorative search behaviours as one browses the lattice structure. However, given that the source of the meta-data was sourced from the Australian Museum’s thesaurus it

was evident from the evaluation results that in some cases there were visual cues of its hierarchical taxonomy, particular with respect to how perspectives ‘cluster’ broad categories of object-types, for instance, we have unique perspectives for ‘body accessories’ and ‘container’ which correspond to ‘categories’ within the thesaurus. Participants also stated that some of the terminology does not reflect the social requirements of 21st century audiences is further evidenced by categories or tags such as “domicile” and therefore stated that perspectives or categories in which the object descriptions or metadata are currently divided need to be revised in order to be relevant for a broader audience. Despite this, participants generally appreciated the unique style of navigation that the VMP presents, although many stated that users may be discouraged in exploring objects within the VMP due to the relative unfamiliar style of navigation along with its steep learning curve. There were suggestions to implement a video tutorial or some form of contextual help so that new users of the system can become familiar with the style of navigation.

Participants stated that the VMP would work best in an exhibition context as an interactive tool and would be suitable for hosting blogs and discussions concerning the objects and the site itself. Furthermore participants stated that the use of visuals and animation to convey navigation rather than the use of hierarchical text labels would make the VMP more appropriate and accessible to its audiences, and that the use of Formal Concept Analysis and its non-hierarchical means of exploring an otherwise flat (or near-flat) taxonomy of objects partially achieves this objective.

Overall, participants felt the VMP could make a useful addition to the Museums representation practices and knowledge generation, once the issues of terminology and accessibility are resolved. Furthermore, the need to make parts of the collection visible to the public that otherwise would remain in storage was almost unanimously pointed out as a progressive museum strategy and in accordance to international developments. Here, the possibility for creator communities to actively participate in the knowledge creation opens not only up to the preservation of traditional knowledge for their respective communities but provides the museum with the opportunity to update mostly 19th century annotations and descriptions.

There are presently over 500 registered users of the Virtual Museum of the Pacific.

#### IV. CONCLUSIONS

Our paper presents a social tagging application of a digital library with novel search and discovery features based on Formal Concept Analysis. We have attempted to convey the key search and navigation features – the style of the interaction and the user experience is very important for its intended use – but we also discussed our distributed data model that supports extensible social tagging and web services implementation. We examined issues relating to system scalability and explained and characterized the providence of the museum data and interests of its stakeholders. We have also presented the first results of the preliminary stakeholder evaluation trials.

#### ACKNOWLEDGMENTS

This work is supported by the Australian Museum and by the ARC linkage grant LP0884075. The authors wish to thank fellow team members, Prof. Amanda Lawson, Friederike Krishnabhakdi-Vasilakis and Jon Ducrou who contributed material support and ideas in the production of this paper.

#### REFERENCES

- [1] M. J. Addis, K. Martinez, P. Lewis, J. Stevenson, and F. Giorgini. New ways to search, navigate and use multimedia museum collections over the web. *Museums and the Web*, Vancouver, Canada, 2005.
- [2] C. H. Brooks and N. Montanez. Improved annotation of the blogosphere via autotagging and hierarchical clustering. In *Proceedings of the 15th international conference on World Wide Web*, pages 625–632. ACM New York, NY, USA, 2006.
- [3] C. Carpineto and G. Romano. *Concept Data Analysis: Theory and Applications*. John Wiley & Sons, 2004.
- [4] S. Christiaens. Metadata mechanisms: From ontology to folksonomy... and back. *Lecture Notes in Computer Science*, 4277:199, 2006.
- [5] R. Colomb. *Information Spaces: the Architecture of Cyberspace*. Springer, London, 2002.
- [6] P. Constantopoulos, M. Doerr, M. Theodoridou, and M. Tzobanakis. On information organization in annotation systems. *Intuitive Human Interface, LNAI*, 3359:189–200, 2004.
- [7] F. Dau, J. Ducrou, and P. Eklund. Concept Similarity and Related Categories in SearchSleuth. In *Proceedings of the 16th International Conference on Conceptual Structures (ICCS’08)*, LNCS 5113, pages 255–268. Springer-Verlag, 2008.
- [8] J. Ducrou. *DVDSleuth: A Case Study in Applied Formal Concept Analysis for Navigating Web Catalogs*, pages 496–500. LNCS 4604. Springer-Verlag, 2007.
- [9] P. Eklund, T. Wray, P. Goodall, B. Bunt, A. Lawson, L. Christidis, V. Daniels, and M. V. Offfen. Designing the Digital Ecosystem of the Virtual Museum of the Pacific. In *3rd IEEE International Conference on Digital Ecosystems and Technologies*, pages 805–811. IEEE Press, 2009.
- [10] B. Ganter and R. Wille. Conceptual scaling. In F. Roberts, editor, *Applications of Combinatorics and Graph Theory to the Biological and Social Sciences*, pages 139–167. Springer-Verlag, 1989.
- [11] S. Golder and B. Huberman. The structure of collaborative tagging systems. *Journal of Information Science*, 32(2), 2006.
- [12] D. B. Krafft, A. Birkland, and E. J. Cramer. Ncore: architecture and implementation of a flexible, collaborative digital library. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, pages 313–322, Pittsburgh PA, PA, USA, 2008. ACM.
- [13] C. Lagoze, D. B. Kraft, S. Payette, and S. Jesuroga. What is a digital library anyway? *D-Lib Magazine*, 11(11), 2005.
- [14] C. A. Lee. Taking context seriously: a framework for contextual information in digital collections. *UNC SILS TR-2007-04*, 2007.
- [15] B. Martin and P. Eklund. Asymmetric Page Split Generalized Index Search Trees and Formal Concept Analysis. In *5th Int. Conf. on Formal Concept Analysis (ICFCA 2007)*, LNCS4390, pages 82–97. Springer-Verlag, 2006.
- [16] E. Michlmayr. A case study on emergent semantics in communities. In *Proceedings of the Workshop on Social Network Analysis, International Semantic Web Conference (ISWC)*, 2005.
- [17] G. Miller. WordNet: A Lexical Database. *Communications of the ACM*, 38(11):39 – 41, 1995.
- [18] J. Trant and B. Wyman. Investigating social tagging and folksonomy in art museums with steve.museum. In *Proceedings of the International World Wide Web Conference*, 2006.
- [19] R. Wille. Methods of conceptual knowledge processing. In *Prof. of the 5th Int. Conf. Formal Concept Analysis (ICFCA 2006)*, LNCS4390, pages 1–29. Springer-Verlag, 2006.
- [20] R. Wille and B. Ganter. *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, Berlin, 1999.