Elise Conradi

to_be_classified:

A Facet Analysis of a Folksonomy

Abstract

This research examines Ranganathan's postulational approach to facet analysis with the intention of manually inducing a faceted classification ontology from a folksonomy. Folksonomies are viewed as a source to a wealth of data representing users' perspectives. An in-depth study of faceted classification theory is used to form a methodology based on the postulational approach. The dataset used to test the methodology consists of over 107,000 instances of 1,275 unique tags representing 76 popular non-fiction history books collected from the LibraryThing folksonomy. Preliminary results of the facet analysis indicate the manual inducement of two faceted classification ontologies in the dataset; one representing the universe of books and one representing the universe of subjects within the universe of books. The ontology representing the universe of books is considered to be complete, whereas the ontology representing the universe of subjects is incomplete. These differences are discussed in light of theoretical differences between special and universal faceted classifications. The induced ontologies are then discussed in terms of their substantiation or violation of Ranganathan's Canons of Classification.

Acknowledgments

There is a Vedic episode, bearing on "Ideas and words run together in pursuit of the Absolute". Word came back first. "Where is the Idea?", it was asked, "I could not keep up. Therefore, I have returned. But the idea was going further ahead." At this moment, the idea comes back with a similar story, "I too could not keep up. The difficulty became unbearable after I got separated from the words". But the word said, "And yet you did go ahead of me. Unless I keep close to you, I become inert." (Ranganathan, 1967, p.327)

I would like to thank my teachers and fellow students for engaging me in stimulating discussions in and out of the classroom. They added fuel to my ideas and contributed to an extremely rewarding learning experience throughout the course of working on this thesis. I would also like to thank Tim Spalding and LibraryThing for an exciting dataset with which to work; the Interlibrary Loan department at the Oslo University College Library for procuring countless documents about facet analysis from the past 60 years; Jan Stephan Muryn for multi-faceted support along the way; and Theodor Conradi Muryn for demanding that I retain perspective. Last but not least, I am very grateful to my advisor, Ragnar Nordlie, for helping me to reunite my runaway ideas with words, and for prodding me along when my words verged on inertia.

Oslo, December, 2009
Elise Conradi

TABLE OF CONTENTS

1   Introduction

Today's digital information landscape is like a wonderful beast of a forest, growing wilder and wilder with each passing second, enticingly defiant of attempts at controlled cultivation. In terms of sheer size, estimated 161 billion gigabytes digital information was generated in 2006. This number is thought to increase sixfold by 2010, making it equivalent to more than 7 million times the amount of physical information currently stored at the Library of Congress (IDC, 2009; Library of Congress, 2007). The landscape encompasses a plethora of different content types, ranging from representations of absolutely everything that is sellable on the internet to full-texts of scanned physical books and articles. It also includes digitally born objects in an assortment of traditional and emergent genres, an increasing amount of which is either partially or wholly user-generated. These different content types exist in a variety of formats, including text, sound recordings, still images and movies, and in an even wider variety of file formats. To describe the landscape as complex is pertinent, albeit an obviously gross understatement.

Finding one's way around this huge and complex forest is indeed a daunting task. Information organization professionals are increasingly rediscovering tools that were originally developed within library and information sciences (LIS) for the purposes of structuring, organizing and labeling physical information, and faceted classification in particular has been gaining in popularity over the past few years (Adkisson, 2003; La Barre, 2006). That faceted techniques should be appealing in today's digital information landscape is hardly surprising. Faceted classification was initially developed in the 1930s precisely to cope with the growing size and increasing complexity of the information resources of the time (Ranganathan, 1961, p. 83) and studies of faceted techniques applied to digital information indicate that they are both scalable and highly suitable to digital environments (Ingwersen & Wormell, 1992; Ellis & Vasconcelos, 1999). There is, however, documented confusion regarding the theories behind facet analysis techniques.

Concurrently, an entirely new classificatory tool has emerged on the landscape: the folksonomy. Also called distributive or social classifications, folksonomies are easily generated by users or creators of digital information from the bottom-up, resulting in cost- and labor-efficient ways to label and categorize colossal collections of information resources. A by-product of the tool is that

they provide new insight into precisely which terms users choose to describe resources. They fail, however, to explicate the semantic and syntactic relationships between these terms (Mathes, 2004; Hammond et al., 2005).

In this thesis, I will explore how facet analysis techniques can be used to explicate relationships in a folksonomy. Specifically, I intend to examine how the postulational approach to facet analysis can be used to attempt to manually induce a faceted classification ontology from a flat tag space. In so doing, I hope to find out what types of faceted structures will emerge and whether these structures substantiate faceted classification theory. To accomplish this, I have delved into the canonical literature on faceted classification and used this as a basis upon which to perform a facet analysis of a folksonomy.

1.1   Background

At the center of this research are two popular and highly different classificatory systems: faceted classifications and folksonomies. Faceted classification theories are here invoked in order to illustrate and examine facet analytical techniques. Folksonomies are here viewed as a source to a wealth of data representing users' perspectives.

1.1.1   Facet analysis: unclear theoretical underpinnings?

The use of facets to organize information is commonplace on the web today, in particular on commercial websites. Already in 2003, a study found that some form of faceted classification was used in 69% of 75 leading e-commerce sites (Adkisson, 2003). More recently, an increasing number of non-commercial actors have begun to use facets on their websites. In 2006, a random sampling of commercial and non-commercial websites from four categories in the Open Directory Project showed that facets were used in 37% of the 65 websites that had integrated search components (La Barre, 2006, p. 161). Befittingly, this trend is seen most clearly in the library sector, where the use of facets for the refinement of results lists has become a common feature in most second generation OPACs and in all next generation catalogs.

Parallel to the increase in popularity of faceted techniques is a widespread call for clearer information about the theories that lie behind them. La Barre documented this exigency in her 2006 doctoral thesis in which she examined the use of faceted analytico-synthetic techniques

(FAST) on websites. As part of her analysis, she interviewed 18 information organization professionals who regularly used faceted techniques with the goal of answering the following question: "In what ways do the products of those who make explicit claims to utilize or be informed by faceted analytico-synthetic technique (FAST) conform to or depart from the theory as described in LIS literature?" (La Barre, 2006, p. 127). She noted an observation by one of the interviewees that was representative of many of the members of the group: "I think we use faceted classification, and everybody understands it more or less, but nobody has really formulated it for us in a way we can understand. The practice we have needs to be theorized a bit and formalized" (Interviewee 2, 2005 from La Barre, 2006, p. 153). This observation is corroborated by a host of forums and blogposts in which discussions of facet analysis and faceted classification abound. La Barre and Cochrane (2006) registered comments from several information architecture mailing lists, websites and blogs exemplifying the confusion surrounding facet analysis and the need for a clarification of its theories.

Even within LIS, there appears to be a dearth of clear knowledge about the theories and principles behind facet analysis techniques. Spiteri (1998) summarized the situation as follows: "Although *Prolegomena* (by Ranganathan) is readily available to LIS students, the same cannot be said for its contents… [Furthermore,] the CRG does not present its complete set of facet analysis principles in any one source, which means that LIS educators and students are required to consult a variety of works written by different members of the CRG." Her oft-cited work, "A Simplified Model for Facet Analysis," is an attempt to remedy the situation by serving as a guide to LIS students and information architects alike to understanding the postulates and principles of Ranganathan and the CRG. The model, however, says very little about how to actually perform a facet analysis. Giess etal (2007) make the following observation: "In essence, many Library Science texts are evaluative as opposed to generative. The methodologies expressed in the more applied texts tend to provide broad overviews … [instead of] discussing practical applications and examples of facet analysis."

One source of confusion regarding facet analysis is the fact that its techniques are used in different ways, for different purposes and on different information resources. Within the canonical literature on faceted classification theory, facet analysis is alternately described as a technique to construct faceted classification schemes and as a technique to classify documents.

The former is essentially a type of domain analysis performed by a classificationist, while the latter is usually a subject analysis performed by a classifier; both actions are interdependent and essential to faceted classification theory. As will be discussed, a facet analysis of a folksonomy with the intention of unearthing a faceted classification ontology combines aspects of both of these types of facet analysis. A clear understanding of faceted classification theory is therefore an essential component of this research.

1.1.2 Folksonomies: unearthing the "wisdom of the crowd"

Since their inception on the web in 2003 with the tagging system Del.icio.us, folksonomies have become an enormously popular way to categorize large amounts of information resources. Folksonomies emerge from the aggregation of textual labels called tags that are affixed to digital objects of various formats within sites that allow for tagging. Depending on the system, tags are either generated by the creator or owner of the content, by the users of the content, or by a combination of the two. The former is called a *personomy*, while the latter two are called *folksonomies*. The term *folksonomy* was coined in 2004 by Vander Wal, who explained "that if you took "tax" (the work portion) of taxonomy and replaced it with something anybody could do you would get a folksonomy" (2007).

Folksonomies have been criticized by those advocating top-down approaches to organizing information. It is argued that the uncontrolled vocabulary of tags causes too many recall and precision problems (primarily due to ambiguity, polysemy and synonymy) to make them useful as information retrieval tools, and that the flat structure of folksonomies prevent users from seeing relationships between information items (Rosenfield, 2005; Petersen, 2006). Whether one subscribes to these remonstrances or not, the fact remains that tags beget a new layer of flat metadata in which huge collections of information resources are described and categorized in manners that presumably give most meaning to individual users. Furthermore, studies of folksonomies have revealed stable trends and patterns in the ways in which large user groups tag items (Golder & Huberman, 2006; Kitt & Campbell, 2006), unearthing what Weinberger (2006) has called the "wisdom of the crowd". Analysis of user-generated metadata can therefore provide invaluable insight to librarians and other information organization professionals into precisely how large groups of users view and describe digital information resources.

In order to understand the construction of a folksonomy, Lambiotte & Ausloos (2005) present a tripartite model consisting of three main components: users, tags and resources. This model has been useful in attempts to explain the relationships between aggregated tags, tagged resources and the community of users (Mika, 2007). This is particularly the case for analyses of so-called broad folksonomies. Vander Wal (2005) distinguishes between broad folksonomies and narrow folksonomies, explaining that a "broad folksonomy has many people tagging the same object" whereas in a narrow folksonomy, an object is tagged "by one or a few people." In an in-depth analysis of a large set of tag distributions, Halpin et al. (2007) show that "tagging distributions tend to stabilize into power law distributions."

Quintarelli (2005) discusses broad folksonomies in terms of the Power Law distribution, stating that the "power law reveals that many people agree on using a few popular tags but also that smaller groups often prefer less known terms to describe their items of interest." He argues that this makes broad folksonomies ideally suited to reveal "trends in large groups of people describing a corpus of items." The consensus of meaning achieved through the aggregation of large sets of tags has been referred to as "emergent semantics" (Marchetti et al., 2007).

The folksonomy used in this research acts primarily as a platform on which to examine the postulational approach to facet analysis. It is hoped that a facet analysis of a broad folksonomy will be able to take advantage of "emergent semantics" and reveal underlying conceptual categories and facets to which the folksonomy's aggregated tags belong. In this way, it is hoped that facet analysis techniques can be used to manually expose a faceted classification ontology in the flat tag space.

1.2 Relevant research

There are three areas of research that are relevant to this study: research on facet analysis techniques, research on folksonomies and research on the combination of the two.

1.2.1 Facet analysis research

Research on facet analysis and faceted classification abounds. Here, focus is placed on research aimed at improving the understanding of Ranganathan's postulational approach to facet analysis, especially as it is applied to new information resources. As such, the studies described here could

also be viewed as supplicates to the theory chapter of this thesis, and where pertinent, they are described in greater detail there.

One of the most recent analyses of theoretically based faceted analyses is (Giess et al., 2008). Based on a review of the canonical literature on faceted classification, they propose a concrete methodology for facet analyses within the field of engineering. Their stated goal is "to identify where the explication of the theory of facet analysis is insufficient for its application in an engineering context, and further … to demonstrate how a faceted classification scheme may be generated for the organisation of an engineering document corpus."

La Barre's doctoral dissertation (2006) offers a framework for "a set of resources and guidelines … as a way to begin discussion about effective practices and to move toward codification of standards and guidelines for faceted organization and access." The guideline is constructed by listing the most common questions she encountered in her interviews with information organization professionals and attempting to answer them with theories and techniques cited primarily from the canonical literature on facet analysis within LIS. She suggests research on the "use of facets with unstructured data" as a viable area for further exploration of this topic.

The Integrative Levels Classification (ILC) Project (2004) examines the use of free facets in a classification based on integrative levels. As such, it is an attempt to extricate facets from subjects and rather present them as being aspects of concepts. The Dandelion bibliography of facet analysis (Hong, 2006) uses ILC to classify a collection of "references, both printed and on the Web, concerning facet analysis theory and its application to knowledge organization."

Denton (2003) writes that "a survey of the literature on applying facets on the web shows that librarians think it a good idea but are unsure how to do it, while the web people who are already doing it are often unaware of S.R. Ranganathan, the Classification Research Group, and the decades of history behind facets." To alleviate the situation, he "attempt[s] to bridge the gap by giving procedures and advice on all the steps involved in making a faceted classification and putting it on the web." Throughout his oft-cited work, he draws on examples from faceted classification theory. In a poster submitted to the American Society of Information Science & Technology (ASIS&T) Information Architecture Summit, Louie (2003) diagrammatically shows

how faceted classification theory from LIS can be used to provide structure for Information Architecture.

In a project titled "Towards a knowledge structure for high performance subject access and retrieval within managed digital collections" (2002-2003), Broughton and Slavic examine facet analytical theory (FAT) for use in the classification of digital resources within arts and humanities disciplines. The stated aim of the project was "to investigate the feasibility of using FAT to develop a knowledge structure suitable for the digital environment." Borrowing facet analysis techniques and citation order from the Bliss Bibliographic Classification 2, the resulting classification model (FAT-HUM Classification) provides faceted classification schemes for two disciplines: Religion and the Arts.

The aforementioned research by Spiteri (1998) has as its stated goal "to propose a simplified model for facet analysis that incorporates the principles of facet analysis proposed by both Ranganathan and the CRG." The resulting simplified model "gives an overview of the underlying principles of facet analysis that are common to both these theories, and which reflects common usage amongst the designers of faceted classification systems and IR thesauri."

Facet analysis techniques are examined for use in digital environments and the World Wide Web in (Ingwersen & Wormell, 1992; Ellis & Vasconceles, 1999). Both studies found that Ranganathan's faceted classification theories are ideally suited for digital information structuring. The ongoing Flamenco project at Berkeley explores several aspects of faceted navigation in user interfaces (Stoica et al., 2007; Hearst, 2006; Yee et al., 2003; English et al., 2002).

1.2.2 Folksonomy research

The focus of this research is on the examination of the postulational approach to facet analysis and its application to a folksonomy. The approach will be studied to find out whether it can be used to manually expose a faceted classification ontology from a folksonomy. Therefore, the research is tangentially related to several other recent studies on the identification of ontologies and other structures in flat tag spaces.

Using a subsumption-based model, Schmitz (2006) shows promising results in the automatic inducement of an ontology from Flickr[1] tags. He writes that a refinement of the model using probabilistic methods may "improve upon the accuracy, and also induce a faceted ontology." Dix et al. (2006) analyze semantic relationships in large sets of aggregated tags and test the use of a "semantic halo" to broaden meaning in automatic queries. Mika (2007) uses a tripartite model of folksonomies to discuss the inherent social context of ontologies in "Ontologies are us: A unified model of social networks and semantics." He uses the concept of social context to illustrate two emergent ontologies. Halpin et al. (2007) examine the "short head" in tags displaying power law distributions in order to "show how tag co-occurrence networks for a sample domain of tags can be used to analyze the meaning of particular tags given their relationship to other tags."

There have been several studies on the trends and patterns underlying folksonomies. In "The Structure of Collaborative Tagging Systems" (2006), Golder & Huberman found that "because stable patterns emerge in tag proportions, minority opinions can coexist alongside extremely popular ones without disrupting the nearly stable consensus choices made by many users." In "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices" (2007), Kipp & Campbell arrive at the same conclusion and found additionally that temporal tags suggest "the presence of an extra dimension in classification and organization, a dimension which conventional systems are unable to facilitate."

In an attempt to disprove the flat nature of folksonomies and to examine their effectiveness in terms of retrieval and organization, Kome (2005) studies the relationships between tags. He concludes that implicit hierarchical relationships exist in folksonomies. Beaudoin (2007) examines emergent patterns within Flickr tags. Through an iterative process of analysis, she found that 18 categories of tags consistently emerged.

There are also two master's theses worth mentioning here that have examined aspects of tags from LibraryThing. Smith (2007) compares tags from LibraryThing with controlled vocabularies in "Cataloguing and You: Measuring the Efficacy of a Folksonomy for Subject Analysis". In "Classified: Analysis of user generated metadata in the LibraryThing folksonomy," Sterken

---

[1] http://www.flickr.com/

(2008) examines the differences between librarian-generated tags and non-librarian-generated tags at LibraryThing and then compares them both with Library of Congress Subject Headings.

1.2.3   Facet analysis and folksonomy research

Several recent projects and studies have focused on the relationship between faceted classifications and folksonomies.

Choi (2009) explores how a facet analysis of tags can bring to light the user-perspective in the design of faceted navigation systems. In "Bringing a More Accurate User's Perspective into Web Navigation: Facet Analysis of Folksonomy Tags," the author examined tags from Del.icio.us, compared them with labels from two web directories in the attempt to place them all into predefined conceptual categories.

Weaver (2007) studied the tagging practices of a library community in order to find out if an "examination of user-generated metadata can reveal new approaches to information architecture." In "Contextual metadata: faceted schemas in virtual library communities", he describes "a faceted structure to current approaches for user-generated metadata, adding versatility to search terms."

In "Folksonomies: Power to the people" (2005), Quintarelli argues that "traditional hierarchies for organizing information (or reality) will not be replaced by tags, but through tagging, we are finding new ways of thinking about classification and new applications for organizing and sharing knowledge." Joined by Resmini and Rosati, he developed a "working prototype of a semantic collaborative tagging tool" which is described in "Facetag: Integrating Bottom-up and Top-down Classification in a Social Tagging System" (2007). Facetag is a tagging system which allows users to choose free tags within predefined facets in order to improve retrieval.

There are also a number of non-academic projects that have explored the use of facets with tags. In 2005, the corporation Siderean played around with the idea that tags from Del.icio.us could automatically be extracted and grouped into facets. The result of the thought experiment was a short-lived site called Fac.etio.us[2], which automatically grouped tags into the following facets: organization, activity, place, technology, attribute, genre, tag, contributor, site and date. Other

---

[2] Fac.etio.us, by way of the Internet Archive Wayback Machine:
http://web.archive.org/web/20060526050202/demo.siderean.com/facetious/facetious.jsp

commercial enterprises combining the use of tags with facets are Buzzillions[3], Peter Van Dijck's brainchild MeFeedia[4], and Raw Sugar[5], a "guided, tag-based search engine."

1.3　Statement of research questions

This research attempts to answer four questions:

1. How does one apply the postulational approach to facet analysis to a folksonomy?

2. What types of challenges and problem areas exist in a facet analysis of this type of data?

3. What kinds of facets, conceptual categories and relationships can be identified in the folksonomy chosen for this research, and how are these characterized?

4. Where do the results of a facet analysis of this type of data substantiate faceted classification theories and where do they depart?

It is my contention that these questions remain largely unanswered. It is my hope that answering them will aid in improving the understanding of facet analytical theory, while gaining new insight into user-generated metadata.

1.4　Thesis outline

In chapter 2, the theoretical foundation for a facet analysis of a folksonomy will be laid. Here, I will introduce an in-depth discussion of Ranganathan's theories and I will introduce relevant aspects of his approach to facet analysis. I will take into consideration those aspects of the approach discussed by the Classification Research Group (CRG) that are relevant for the work at hand.

Chapter 3 is devoted to an explication of the methodology. As will be shown, the methodology is based on the theoretical discussions of facet analysis in chapter 2. Here, I will also describe the folksonomy used in the research and the tagging system whence the dataset was selected.

Results of the facet analysis will be discussed in chapter 4. Focus will be placed on the correspondence between the results and the underlying theoretical foundation. Special attention

---

[3] http://www.buzzillions.com/
[4] http://www.mefeedia.com/
[5] http://www.rawsugar.com/

will be given to areas where the two diverge. A discussion of problems and shortcomings will in the dataset and methodology will also be included.

Chapter 5 is reserved for concluding remarks regarding theoretical and practical implications of the results. Recommendations for further study will also be offered here.

Before continuing, a brief discussion about the notion of facets may be helpful.

1.5   The Concept of Facet

The concept of facet has been in use for the organization of knowledge for three quarters of a century in a variety of disciplines and information environments. It has consequently come to be defined in many different ways. Before continuing, I will here attempt to distinguish between some of the most popular uses of the term and state how I intend to use it in this text.

Although the facet concept can be identified in the works of Otlet & Fontaine and Kaiser (Broughton, 2006), Ranganathan is generally attributed with its invention sometime in the 1930s when he first developed his Colon Classification scheme and published his theories related to the scheme, most notably in his *Prolegomena to Library Classification* in 1937. He did not begin to use the word 'facet' until the publication of the 4th edition of Colon Classification in 1952, opting instead to call them 'trains of characteristics,' a term he continued to interchange with 'facet' for the rest of his life (Beghtol, 2008). Facet is succinctly defined by Mills (1960) as "the total subclasses resulting from the application of a single principle of division" (p. 8). This is the definition that will be used throughout this research.

It appears that the concept of facet is sometimes confused with the concept of conceptual category. This seems particularly to be the case in recent literature intended towards information organization professionals. In the third edition of Morville and Rosenfield's oft-cited reference book, *Information Architecture for the World Wide Web* (2007), for example, the term facet completely replaces the term category in the discussion of faceted classification: "[Ranganathan] suggested five universal *facets* to be used for organizing everything" (p. 221, my italics). Likewise, an article in the peer-written information architecture journal, *Boxes and Arrows* describes "the fundamental *facets* that Ranganathan developed" (Steckel, 2002, my italics). The misapplication is also found in (Uddin & Janecek, 2007), (Rabourn, 2003), (Redmond-Neal, n.d.)

and in the Wikipedia article on the Colon Classification (retrieved November 27, 2009). Although it is quite conceivable that the misapplication of the terms is intentional for the sake of simplicity, the distinction between the two has important implications. Facets are used to differentiate between aspects of each individual class in a universe, while categories differentiate between aspects of all of the classes equally in a universe. If facets are assumed to be equal to categories, one loses the distinction between the level of universe and the level of classes in the universe, thus requiring facets to differentiate from a more general level.

Conversely, although not nearly as extensive, it has been suggested by some that tags are like facets (Smadja, 2005). This misconception is likely due to the fact that tags can be used to create synthetic relationships in the same way that these relationships are created in faceted systems. In faceted systems, however, it is the subclasses, or foci, of the facets that are combined to create synthetic relationships. Here, tags will be treated as potential foci belonging to facets.

Figure 1 illustrates the simplified relationships between tags, facets and categories that will be used in this research:
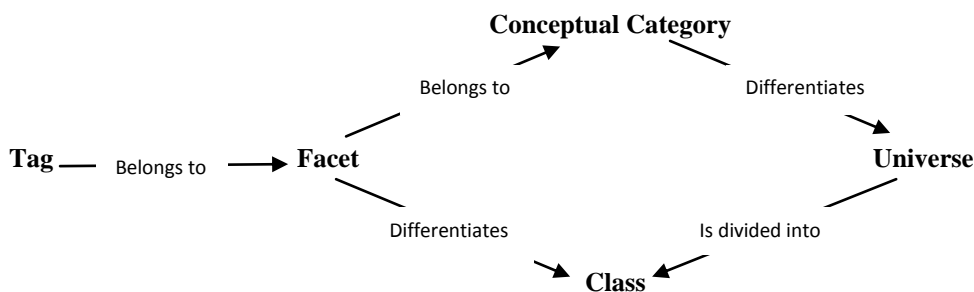
Figure 1: Ontological relationships between tag, facet[6], category, universe and class

Another area in which the concept of facet has caused confusion is in its application. Can facets be facets of *anything*? Broughton (2006) writes about the purist view of faceted classification, which maintains that facets should be regarded as facets of subjects, as they are in library classifications. From this perspective, most of the facets on the web today are not regarded as

---

[6] As will be shown in section 2.2.2.2, the model is slightly different for basic facets.

faceted classifications. Schwartz (2008) points out what seems to be the crux of the problem: the term facet is used differently "in information architecture (IA) and guided navigation, where 'topic' is one among many facets, [than it is in library science and] thesaurus development, where 'topic' is the primary object of facet analysis." In this research, it is assumed that the concept of facet can be applied to aspects of any universe, regardless of whether the universe is topic-based or entity-based. As will be shown in chapter 2, this is consistent with Ranganathan's faceted classification theory. The bottom line is that facets can be facets of *anything*, as long as they are facets *of* something.

2   Theoretical foundation for a facet analysis of a folksonomy

The methodology used in this research is based on the postulational approach to facet analysis, which is inextricably tied to the theory of faceted classification. Furthermore, the results of the facet analysis will mainly be viewed from within the faceted classification paradigm. It is therefore necessary to understand this theory in order to discuss the implications of applying a facet analysis to a folksonomy.

2.1   The faceted classification paradigm

The introduction and development of faceted classification in library and information sciences arguably represents a Kuhnian paradigm shift within knowledge organization (Dahlberg, 1992; Xiao, 1994). Previously, although pragmatic by purview, library classificationists had been highly influenced by traditional philosophical classifications of knowledge, adapting the ontological view that knowledge[7] can be divided into neat, hierarchical categories (Abrera, 1974, p. 21). The first library classification systems (Dewey Decimal Classification, Cutter Classification, Library of Congress, Brown's Subject Classification)[8] all reflect a top-down one-place-for-everything ontological view of the universe of knowledge, most commonly depicted as an upside-down hierarchical tree-like structure. Like traditional library classifications, faceted classification is pragmatically based, but it is grounded in theory and it represents an entirely new ontological perspective within knowledge organization.

Developed by Indian librarian and mathematician Shiyali Ramamitra Ranganathan in response to the increasingly complex nature of the subjects in books and documents in the 1930s which, he perceived, traditional library classifications failed to accommodate, faceted classification offers an alternative to the ontological view that subjects have one and only one placement in a classification scheme. Instead of accepting that the universe of knowledge can only be classified in one way by set categorical paths, Ranganathan proposed a system in which knowledge is represented as being multi-faceted and classifiable from a number of different perspectives. Based on an examination of the literature to be classified, a faceted classification scheme is built

---

[7] knowledge in and of itself or as it is reflected in documents

[8] These classification schemes would later come to be known as enumerative classification schemes in order to distinguish them from analytic-synthetic and faceted classification schemes.

bottom-up, comprising several hierarchical classifications based on different perspectives of the classified items. This theoretically allows for the multiple placements of complex subjects into the classification system and hence, it provides multiple access points to the classified documents.

The principles of faceted classification theory were first introduced by Ranganathan in 1937 in his *Prolegomena to Library Classification* and were based on a "positivistic preparation" which consisted of "ten years' work in forging and polishing the Colon Classification; and eight years' work in teaching the Colon Classification and the Decimal Classification on a comparative basis" (Ranganathan, 1967, p. 22). A "theory after design", he had by his own calculation facet analyzed nearly 100,000 books between 1925 and 1936, providing him with the experience and practical insight he felt was necessary to explicate a new classification theory (Ranganathan, 1961, pp. 85-86). Over the course of the next three decades, he further developed the theory in the publication of two revised editions of the *Prolegomena* (1957 and 1967) and in numerous articles and lectures. The postulational approach to facet analysis was introduced in the second edition and further developed in the third edition.

Meanwhile, the Classification Research Group (CRG) was established in the United Kingdom in 1952 to "discuss the principles and practice of bibliographic classification, unhampered by allegiance to any particular published scheme" (Vickery, 1966, p. 10). By 1955, a clear preference for faceted classification was apparent in the group's collectively submitted memorandum to UNESCO and to the *Library Association Research Committee* entitled "The need for a faceted classification as the basis of all methods of information retrieval." Expanding on Ranganathan's theories and building on the postulational approach to facet analysis, the CRG's subsequent research on facet analysis is primarily scattered amongst the prolific writings of its various members, some of which lead in slightly different directions.

The postulational approach to facet analysis created by Ranganathan and developed by the CRG has informed much of the work on faceted classification during the twentieth century through today, and it provides the theoretical underpinnings of this research, where it will be examined for use in a facet analysis of a folksonomy. Before delving into the postulational approach, however, it is necessary to consider a number of issues when applying a theoretically based facet

analysis to user-generated metadata. These include the concepts of universes, roles, planes and sequence. Their illumination will help show which aspects of faceted classification theory are needed in this research.

2.1.1 Universes in faceted classification theory

Although the theory of faceted classification was created for intended usage within library classification, Ranganathan explicitly writes that the Canons of Classification "govern the classification of any universe of any kind whatsoever" (Ranganathan, 1967, p. 80). Universe is here defined as a "collection of entities, without any special arrangement among them, [and that is] under consideration in a given context". Examples of universes include the finite universe of "furniture in a room", the infinite universe of "all men, past, present, and future", and the growing universe of "books in a library" (ibid, p.54). The facet analysis in this research concerns user-generated metadata representing a finite universe consisting of 76 books. The universe of books (and of any type of document or object that is about something) is an exciting universe to facet analyze because, depending on the purpose of the classification, one of its facets can be the universe of subjects. Ingwersen & Wormell (1992) note that "this implies to accept several classifications for one physical entity."

2.1.1.1 The Universe of Subjects

The universe of subjects is said to be the most difficult of the universes to classify. Ranganathan describes it as a "Continuous Infinite Universe", meaning that the things and ideas included in it are "so packed that it is impossible to extricate any single entity from out of its neighboring ones" (Ranganathan, 1967, pp. 74-75). A subset of the universe of knowledge, the universe of subjects is the primary object of library classification and hence the primary subject of library classification theory. Universal library classification concerns itself with the classification of the universe of all subjects, while special library classification deals with the universe of a limited set of subjects.

Subject is defined by Ranganathan as an "organized or systematized body of ideas," existent in documents as either basic subjects, compound subjects or complex subjects. (Ranganathan, 1967, pp. 82-85). Also called the aboutness of a document, Hjørland (2001) writes that "the subject of a document is that 'something' that subject analysis and retrieval are supposed to identify," thereby

connecting its value to "the future use of the document" and to its relevance for the user. Although Hjørland convincingly argues for a theory of subject from a domain-oriented perspective, in this research, aboutness is assumed to be formed by the consensus of meaning formed through the accumulation of aggregated tags. Aboutness is thus socially and cognitively defined.

Broughton (2006) notes that universal faceted classifications of the universe of subjects actually consist of several different faceted classifications: one for each main subject. The entire universe of subjects is first divided by the classificationist into a number of basic subjects, usually corresponding to the major disciplines[9]. These become the main classes of a universal classification. Each class is then facet analyzed by the classificationist, which means that the initial division into disciplines is "external to the application of the facet analysis proper" (Broughton, 2006). The implication of this is that, although it has been proposed that there are fundamental categories common to all subjects, facets vary from discipline to discipline. As will be seen, the notion of facets as dependent on basic subjects is problematic and will be discussed in greater detail in section 2.4.2.

2.1.2   Roles in faceted classification theory

When discussing the theoretical underpinnings for a facet analysis of a folksonomy, it is be useful to distinguish between the different roles involved in classification. Ranganathan differentiates between what he calls the classificationist and the classifier. The classificationist, he writes, is "one who designs a scheme for classification and provides a set of guiding postulates and principles to fix the position of a newly emerging class by interpolation or extrapolation, as the case may be, in the correct filiatory position, among the already existing classes and for fixing the class number of each such class." The classificationist's tasks include the discerning of core facets in each universe through a facet analysis of the universe and an adherence to a set of normative rules called the Canons of Classification. The Canons of Classification, as will be shown, form the theoretical foundation of a faceted classification.

---

[9] Throughout the rest of this research, discipline is understood to be a branch of knowledge that has traditionally been used as an organizational unit in traditional classifications. When identifying tags indicating disciplines in this research, I have attempted to defer as far as possible to the discipline-division of the universe of subjects as presented in the second edition of the Bliss Bibliographic Classification. This is a highly problematic solution, which will be discussed in more detail in section 4.4.3.

The classifier is "one who classifies a universe in accordance with a preferred scheme for classification and fixes the position of any newly emerging class by interpolation or extrapolation, as the case may be, in the correct filiatory position among the already existing classes and determines its class number, in accordance with the postulates and principles laid down by the classificationist for this purpose" (Ranganathan, 1967, p. 79). The classifier's work pertains to the identification of facets by means of a facet analysis of the objects to be classified. As will be discussed, the classifier's work is guided by adhering to postulates formulated by the classificationist.

It may also be helpful here to add a third role to this discussion of classification theory, namely that of the user of the classification system. The user in this sense is defined as one who uses a classification system for the purpose of information retrieval and discovery. In a folksonomy consisting of user-generated metadata, however, the user's role is mixed: in addition to the aforementioned role, the user is also the classifier. In an analysis of a facet analysis of a folksonomy, focus will be placed on the mixed roles of the classificationist, the classifier and the user. It will be seen that the user has already performed half of the classifier's work in a facet analysis by already having analyzed and described the documents collected. It will also be seen that the initial work performed by the classificationist, i.e. the analysis of the universe to be classified, is absent.

2.1.3 Planes in faceted classification theory

Ranganathan distinguished between three different planes when working on classification: the idea plane, the verbal plane and the notational plane (Ranganathan, 1967, pp. 327-328). Work in the notational plane includes the creation and development of class numbers for use in classification schemes. Work in the verbal plane includes controlling vocabulary to compensate for problems that exist in natural language[10], like homonymy, synonymy and the problem of multiple languages (ibid, pp. 329-330). Work in the idea plane includes the "unbridled freedom … to carry out [the] incisive analysis of a subject into its facets, whatever be their number" (ibid, p. 338). According to Ranganathan, the idea plane is the most important of the planes, but it is

---

[10] All the natural language problems associated with the verbal plane are present in the user-generated metadata used in this research. They are ignored, however, beyond the extent to which they directly affect the facet analysis in the idea plane. For a thorough analysis of natural language problems present in user-generated metadata, see *Folksonomies- Cooperative Classification and Communication Through Shared Metadata* (Mathes, 2004).

also the most difficult plane within which to work because it is susceptible to "occultation by the verbal plane", "inhibition by the notational plane" and an "inherent inertia" to resist either (ibid, pp. 335-336).

Work in all three planes is requisite for the creation of classification schemes. As has been noted, however, facet analyses can be used for many other purposes than the creation of classification schemes. The different planes are thus accorded different relevancies depending on the purpose of the facet analysis. In the facet analysis of a folksonomy performed in this research, one of the main intentions is to analyze a sample of user-generated metadata into its facets. This activity takes place exclusively in the idea plane.

2.1.4   Sequence in faceted classification theory

In faceted classification theory, sequence refers both to the correct sequence of entities in the arrays of a classification scheme and to the correct facet sequence in a classified object. Both are necessary to ensure the consistent collocation of similar items in a classification and can thus be said to govern the linear order of physical objects. It has been argued that linear order is essential for the classification of physical books and documents, but that it loses its importance in a digital world. Broughton remarks that the "the concerns in managing the digital information store are not those of arranging the material, but rather of adequate object description […], providing search tools that support browsing, navigation and retrieval, and, to a more limited extent, the presentation of results" (Broughton, 2006, p. 51). Linear order is outside of the scope of this research. Therefore, the several Canons of Classification and the Principles for Helpful Sequence governing correct sequence in characteristics and arrays, plus the postulates governing correct facet sequence, will all be disregarded here.

The concept of fundamental categories seems to be inextricably tied to sequence in some of the literature on faceted classification theory. Indeed, in some sources, 'fundamental categories' is renamed to 'citation order,' indicating that the sequence of the categories is their most important characteristic. In this research, focus is on the role fundamental categories play in determining inter-facet relationships. This will be discussed in more detail in section 4.1.1.

2.2   The postulational approach to facet analysis

The postulational approach to facet analysis refers to a set of normative rules in the *Prolegomena* consisting of 43 Canons of Classification, 12 Postulates and 22 Principles. The Canons of Classification are based on Ranganathan's five Laws of Library Science[11] and are intended for use by the classificationist "only in the design of a scheme for classification" (Ranganathan, 1967, p. 113). It is advocated that they be followed as strictly as possible. There are eight Principles for the classificationist to assist in implementing the Canon of Helpful Sequence (ibid, p. 183) and the rest are intended to aid the classifier with facet sequence (ibid, pp.412-434).

The Postulates serve to guide the classifier in "book classification or the classification of macro thought" (Ranganathan, 1960, p. 39) and "are really concerned with the analysis of any subject into its kernel ideas and their rearrangement and synthesis." Following the Postulates "result[s] in the arrangement, in a consistent sequence, of all the classes going with any specific Basic Class" (Ranganathan, 1965, p. 62), but they are by no means seen to be immutable. In fact, Ranganathan writes, "it is open to those who deal with any particular universe to choose the particular model whose postulates are helpful in that universe" (Ranganathan, 1967, p. 567). Postulates are thus assumed truths that are agreed upon before the process of classifying begins. They are:

> "something about which we agree – something we more or less know. We go forward, and as we do go forward, and go on classifying we make this or that clearer, and modify this or that, if necessary. That is how Postulational Classification begins" (Ranganathan, 1965, pp.198-199).

15 of the Canons and all of the Postulates and Principles in the *Prolegomena* are intended for work in the idea plane. Seven of the Canons and three of the Postulates are concerned with facet choice and the rest deal with sequence. In the rest of this section, I will first introduce these seven canons for the classificationist and discuss the implications of performing a facet analysis inductively, i.e. of classifying before the classificationist has analyzed the universe. I will then discuss theoretical aspects of the three Postulates proposed by Ranganathan for help in choosing

---

[11] Five Laws of Library Science (Ranganathan, 1967, p.115):
1. First Law. –Books are for Use.
2. Second Law. –Every Reader His Book.
3. Third Law. –Every Book Its Reader.
4. Fourth Law. –Save the Time of the Reader and Save the Time of the Staff.
5. Fifth Law. –Library is a Growing Organism.

and identifying facets, and I will examine how they have been interpreted and developed by various members of the CRG. Finally, I will explicate the postulates that support a facet analysis of a folksonomy and that serve as the theoretical underpinnings of this research.

2.3 Canons and the Classificationist

The Canons of Classification are at the crux of faceted classification theory. They serve to guide the classificationist in the construction of a faceted classification by providing strict rules for the division of any universe into its core facets. They are thus responsible for the ontology representing any given faceted universe. As stated, there are 15 Canons that govern the classificationist's work in the idea plane, seven of which concern the choice of facets. The first four are from the Canons for Characteristic and they explicitly govern the choice of facets. The latter three, from the Canons for Succession of Characteristics and the Canons for Array, implicitly affect facet choice by governing facet arrays. I will here delineate each of the canons and give examples of them from a potential classification of a universe of books. Afterwards, I will briefly introduce a discussion of the implications of performing a facet analysis before the classificationist has analyzed the universe.

The first three Canons of Classification all deal with characteristics of division and are called Canons for Characteristic. Characteristic is defined by Ranganathan as "an attribute or any attribute-complex with reference to which the likeness or unlikeness of entities can be determined and at least two of them are unlike" (Ranganathan, 1967, p. 55). In a universe of books, for example, author may be considered a characteristic of division in which books written by the same author form a sub-aggregate of the universe. Indeed, it has been noted that "the application of a characteristic produces a facet" (Mills, 1960, p. 8), such that "by Author" is a facet of the universe of books.

2.3.1 Canon of Differentiation

The Canon of Differentiation states that "a characteristic used as the basis for the classification of a universe should differentiate some of its entities—that is, it should give rise at least to two classes or ranked isolates" (Ranganathan, 1967, p. 145). Consequently, a universe of books divided by the characteristic 'author' would be in accordance with this canon, while the same

universe divided by the binary characteristic 'contains subject' would violate the canon, presuming here that all books are about something.

2.3.2   Canon of Relevance

According to this canon, "a characteristic used as the basis for the classification of a universe should be relevant to the purpose of the classification" (Ranganathan, 1967, pp. 146-147). Thus, the classificationist should consider the needs of the user when deciding which facets to use when dividing the universe. For example, in a classification of the universe of books intended for use by movers, "by Weight" would be a relevant facet; "by Author" and "by Subject" would be relevant facets in a classification in which the users are readers. Ranganathan notes that there may be many different facets relevant to any given purpose. For practical purposes, he advocates that only some of these should be used for each classification scheme, but that any given universe may contain multiple classification schemes. He also remarks that there aren't any "*a priori* rules for hitting upon the most helpful set of characteristics," but that practice and experience generally help the classificationist to determine which facets are most relevant (ibid).

With the exception of Farradane[12], the members of the CRG used the concept of literary warrant to fulfill the requirements of the Canon of Relevance. Facet selection is thus based upon their preponderance within the literature of any given domain: "The theoretically unlimited number of characteristics by which a subject could be divided is thus restricted to those which are relevant to the work in hand—cataloguing documents" (Vickery, 1960, p. 20).

2.3.3   Canon of Ascertainability

The Canon of Ascertainability posits that a "characteristic used as the basis for the classification of a universe should be definite and ascertainable" (Ranganathan, p. 148). This canon is intended to aid the classificationist in choosing among the many relevant facets of any given universe by making it a precondition that the foci belonging to each facet can be checked. For example, in a universe of books, the facet "by Language," which contains the foci representing the different languages the books in the universe are written in, can be ascertained; the facet "by Mood,"

---

[12] Farradane argued that 'literary warrant' and the 'user's point of view' are only "justifiable for special classifications, as long as it is clearly borne in mind that distortions and arbitrary selections have been made from a theoretical perfect general classification" (Farradane, 1961, p. 127).

however, which contains feelings elicited by the book (like "depressing", "thrilling", etc.) is far more difficult to ascertain.

2.3.4   Canon of Permanence

According to the Canon of Permanence, "a characteristic used as the basis for the classification of a universe should continue to be unchanged so long as there is no change in the purpose of classification" (Ranganathan, 1967, pp. 149-151). In a universe of books, the previously mentioned facet "by Mood," for example, would presumably contain highly impermanent foci, as would a facet based on "by Activity" containing activities elicited by the books (like "half-read" and "wish-list") and a facet based on "By Location" containing places where the books are located (like "at home" and "box C"). Ranganathan recognizes that there always exists a potential conflict between the Canon of Relevance and the Canon of Permanence. For example, if the purpose of the classification is to keep track of the books in one's private library, then the facet "by Location" would be highly relevant, despite the fact that it may not be permanent (ibid).

Although the final three canons are primarily intended for the classificationist's work in the idea plane on the facet arrays in the associated classification scheme, they are interesting here because they provide rules for the content of each facet, thus implicitly affecting facet choice. The first of these is the first of three Canons for Succession of Characteristics:

2.3.5   Canon of Concomitance

The Canon of Concomitance states that "no two characteristics in the associated scheme of characteristics should be concomitant—that is, they should not give rise to the same array of subjects or of isolate ideas" (Ranganathan, 1967, p. 153). In a universe of books, the facets "by Last page number" and "by Number of pages" are concomitant because they both give rise to the same content in their arrays; the facets "by Number of pages" and "by Height" are not concomitant because the content in their arrays are different. The CRG called this canon the Principle of Homogeneity.

The two final canons concerning facet choice are the first two of the four Canons for Array:

2.3.6 Canon of Exhaustiveness

According to the Canon of Exhaustiveness, "the classes in an array of classes […] should be totally exhaustive of their respective common immediate universes" (Ranganathan, 1967, pp. 158-159). In a finite universe of a given number of books, this canon is fulfilled when the content of each chosen facet is exhausted. For example, the facet "by Author" fulfils the Canon of Exhaustiveness when all the authors of the books in the universe are presented. This canon, however, is a bit problematic because it can be difficult to determine when a facet has been fully exhausted. Ranganathan notes that it is possible, although undesirable, to exhaust a facet with the use of the "other-device," i.e. a foci representing everything else in the class that hasn't been individualized. One can, however, also interpret this canon as belonging more to the notational plane than to the idea plane. He notes, for example, that the Colon Classification fulfils this canon by allowing for the interpolation and extrapolation of new classes "in their respective proper places among the already enumerated classes" due to the hospitality of its notation (ibid). In "A Simplified Model for Facet Analysis," Spiteri (1998) chooses to exclude this canon, declaring it "rather difficult to determine and maintain."

2.3.7 Canon of Exclusiveness

The final canon for choice of facets is the Canon of Exclusiveness, which states that "the classes in an array of classes and the ranked isolates in an array of ranked isolates should be mutually exclusive" (Ranganathan, 1967, pp. 160-162). The CRG gave this canon the more descriptive title Principle of Mutual Exclusivity. By mutually exclusive, it is meant that "no two classes of the array can overlap or have an entity in common." This is achieved when one and only one characteristic of division is applied at a time. For example, in the universe of subjects within the universe of books, a foci consisting of the compound "Medieval Europe" is the result of two different characteristics of division applied simultaneously, one based on "time" and the other on "place". "Medieval Europe" is thus the result of a violation of the Canon of Exclusiveness. Adhering to the canon would have resulted in two separate facets, where "Medieval Ages" would be found in one of them and "Europe" in the other.

There is documented some confusion regarding the implications of this canon. Mills suggests that the principle of mutual exclusivity pertains to the relationships of the foci within each array as well as to the differences between arrays. He claims, for example, that an array in the class

"Buildings" formed with the facet "by Number of stories" is mutually exclusive because "one cannot have a high-rise single-story building" (Mills, 2004, p. 11). Although hesitant to attribute it to the principle of mutual exclusivity, Wilson (2006) likewise defines strict faceted classification as one in which foci within an array cannot be combined. His solution to the problem is to create individual binary facets for each focus in arrays where combinations of foci would be natural, as in, for example, the facet "by Flavor" in a universe of pies.

It is indeed difficult to determine precisely what Ranganathan meant with this canon. In the example he gives to illustrate the Canon of Exclusivity a universe of professors is divided into two facets: "by Subject" and "by Rhetorical Ability". Although it is clear that a professor cannot be both *brilliant* and *dull* at the same time, it is not unfathomable that a professor specializes in both *chemistry* and *zoology*. In examples he gives to illustrate some of the other canons, it appears as though mutual exclusivity cannot be meant to pertain to the differences between foci. He states, for example, that "by Author" and "by Subject" are relevant facets in a universe of books, but it does not seem likely that by this, he means that each book can only be assigned one author or one subject. In this research, the Canon of Exclusivity is understood to pertain to the differences between arrays resulting from the use of one characteristic of division at a time rather than to the differences between foci in an array. It is acknowledged, however, that the canon is problematic.

2.3.8 Facet analysis without the Classificationist

In the postulational approach to facet analysis, the classificationist facet analyzes a universe under the guidance of the above delineated Canons of Classification. Based on the resulting classification, the classificationist then proposes postulates to guide the classifier in the identification of corresponding facets in the objects to be classified. For example, a facet analysis of the universe of books based on the Canons of Classification may reveal the following facets: "by Genre," "by Process," "by Author," "by Publisher," "by Illustrator," "by Place published" and "by Year published." The classificationist would then propose postulates for the classifier based on this ontological model of the universe. It may, for example, be postulated that there are five fundamental categories in this particular universe (Personality, Energy, Agent, Space and Time) and that these categories have a predefined sequence and relationship to one another. It may further be postulated that each book has facets corresponding to these categories. Guided by

these postulates, the classifier will thus be able to identify facets in the books that are to be classified and place each facet in one of the postulated fundamental categories, thus revealing inter-facet relationships.

In the facet analysis of a folksonomy performed in this research, the classificationist is absent. This means that there will be no prior facet analysis of the universe to be classified. Rather, the facet analysis will take place wholly on the side of the classifier, and it will be performed directly on the users' descriptions of the objects within the universe. This is essentially a reversal of the classificationist's process of faceted classification. An exciting implication of this is that a faceted classification built with the facets found in a folksonomy would be truly inductive. In this way, it is hoped facet analysis can be used as a method to expose a faceted classification ontology in a folksonomy. It will remain to be seen whether the resulting ontology substantiates or violates the Canons of Classification.

2.4   Postulates and the Classifier

Another implication of using the postulational approach to facet analysis without a classificationist is that the postulates must necessarily be defined by the classifier. As stated earlier, the intention of Ranganathan's postulates is to guide the classifier in a facet analysis of documents. By definition, the postulates are not proven truths. Indeed, they are mutable, but they should be agreed upon before the commencement of a facet analysis in order to ensure consistency. Ranganathan proposed three postulates for work in the idea plane to guide in the choice and identification of facets. In this section, I will examine each of these and discuss relevant theoretical discussions they elicited by members of the CRG. In section 2.4.8, I will show how these discussions, considered in the context of the metadata to be facet analyzed and the universes they represent, can provide the background needed to formulate postulates for use in this research. These postulates will be used to form the methodology for a facet analysis of a folksonomy.

2.4.1   The Postulate of Fundamental Categories

Ranganathan's first postulate for facet analysis is the Postulate of Fundamental Categories, which states that "there are five and only five fundamental categories to which facets belong—*viz*, Time, Space, Energy, Matter, and Personality" (Ranganathan, 1967, p. 399). Fundamental

categories are the implicit backbone of a faceted classification scheme and the main constituents governing the relationships in its ontology, but they are very difficult to define. Ranganathan cryptically refers readers of the *Prolegomena* to dictionary definitions of "fundamental" and "category", only to state that the definition of the word-grouping "fundamental categories" cannot be discerned this way, but can be "defined by enumeration only" (ibid, p. 398). In an exposition on the Colon Classification, he expands:

> "The five Fundamental Categories are Personality, Matter, Energy, Space, and Time. […] What are these five? Here again the dictionary is not of much help. It was well-known even to the ancients that we have to make a beginning with some assumed terms. We do not question their meaning. We believe that we know their meaning and we believe also that others too know the meaning. If there are some who do not know the meaning, they will sooner or later come to know it." (Ranganathan, 1965, p. 198).

He first wrote about fundamental categories in *Library classification, fundamentals and procedure* (1944), citing them as a tool to be used in facet analyses to make faceted classifications more hospitable (Mills, 1960, p. 117). Previously, each class in the Colon Classification was divided into facets exclusive to that particular class, plus a few common facets representing Time and Space. Since classes in the Colon Classification are based on disciplines, the facets were thus discipline-dependent. With the introduction of fundamental categories, Ranganathan proposed that every facet in every class represents one of the five posited fundamental categories. In this way, fundamental categories transcend disciplines; their predetermined citation order and explicated relationships apply to facets in all classes.

In the *Prolegomena*, Ranganathan presents the possibility that there may be more than five or less than five fundamental categories and that these may represent other ideas than Personality, Matter, Energy, Space and Time (Ranganathan, 1967, p. 398). He writes that "there is absolute freedom for everybody to try it out." For it to be accepted, however, the classificationist would have to perform a positivist analysis of "some thousands of assorted articles" and find out if "it helps in mapping the universe of subjects in a helpful sequence along a line." Ranganathan thus makes it clear that, although intended for use by the classifier, the acceptance of his postulate of five fundamental categories is essentially a time-saving device for the classificationist. In this way, they play an extremely important but passive role in the creation of a faceted classification scheme. Technically, the classificationist need not even be aware of them when facet analyzing a

given universe. They are solely intended by Ranganathan for active use by the classifier to aid in showing the correct sequence of and relationships between the facets in the objects to be classified.

The CRG proposed a more active role for the classificationist in determining the postulation of fundamental categories in special faceted classification schemes. Vickery writes that "any such list of fundamental categories should not be used mechanically and imposed upon the subject, but to use it as a provisional guide in approaching a new field can be helpful" (Vickery, 1960, p. 24). Fundamental categories thus aid the classificationist in the process of facet analysis by "providing an outline framework which may fit the field, and give guidance in suggesting possible characteristics which should not be overlooked" (ibid, p. 24). Instead of accepting the existence of a set number of fundamental categories beforehand, Vickery advocates the examination by the classificationist of "a representative collection of terms" in each subject field in order to isolate the relevant conceptual categories to be postulated to the classifier (ibid, p. 20). Postulated categories are thus explicitly connected to each specific subject field and are not "necessarily applicable in all subjects" (ibid, p. 24). For the field of Science and Technology, for example, he found the following categories to be helpful: Substance (product), Organ, Constituent, Structure, Shape, Property, Object of Action (patient, raw material), Action, Operation, Process, Agent, Space, and Time (ibid, p. 23).

The different approaches illustrate a fundamental theoretical difference between a universal classification and a special classification. Universal classifications are generally too big to be reanalyzed for new fundamental categories every time they are constructed; Ranganathan recommends a positivistic approach in which at least 100,000 books are examined. Special classifications, on the other hand, are more confined; their ontological model is thus easier to identify and the classificationist can play a more active role in identifying conceptual categories.

2.4.2 The Postulate of Basic Facet
Ranganathan's second postulate is the Postulate of Basic Facet. This postulate states that:

> "every compound subject has a basic facet. A subject may have two or more basic facets. Then it will be a case of phase relation between the basic facets themselves or between the compound subjects of which they are the respective basic facets, or a case of one of

the subjects figuring as a facet in a compound subject going with the other"
(Ranganathan, 1967, p. 402).

This means that each document is primarily about at least one basic subject, represented by a basic facet. He adds that, in order "to identify the Basic Facet of a compound subject, a general knowledge of the schedules of Basic Subjects is necessary" (ibid.). Basic subjects are defined as "subjects without any isolate ideas as components" (ibid, p. 83). These constitute the main classes and the main subdivisions of each class (see *Figure 2*). In essence, this postulate seals the facets of the documents to the classification scheme used for classifying the documents; each facet is really a facet of a subject within the schedule. This brings up two questions. 1) How does one interpret this postulate for the facet analysis of universes other than the universe of subjects? 2) How does one interpret this postulate for the universe of subjects in the absence of a classification scheme?

Tag — Belongs to ⟶ **Basic Facet** — Represents ⟶ **Class** ⟵ Is divided into — **Universe**

Figure 2: Ontological relationships between tag, basic facet, universe and class

In order to answer both questions, it may be helpful to look at some of the ways that faceted classification schemes can be structured, in order to acquire knowledge of the implicit schedules. As discussed, Ranganathan's theories about universal faceted classification schemes concerned schemes in which the universe of subjects is initially divided into classes by disciplines. In this type of scheme, basic facet is thus related to discipline or subdiscipline and the rest of the facets in the document are facets of discipline. Vickery, on the other hand, focused his efforts on working out methods for the construction of special faceted classification schemes. Here, there is very little initial division of the universe of subjects into classes, and if there is, it is into aspects of the specific subject. Thus, in special classifications, basic facet can be said to represent the specific subject or aspects thereof, and the ensuing facets are all facets of this specific subject.

With a specific enough subject, one can envisage a special faceted classification scheme based on one concept. Indeed, Vickery shows how a facet analysis of the concept "soil" reveals five different facets (Vickery, 1960, p. 21). This answers the first question about how it may be

possible to interpret the Postulate of Basic Facet for the facet analysis of other universes than the universe of subjects; the concept represented by the said universe can be viewed in the same way as is a special faceted classification based on one concept. This opens up another question concerning the ways that faceted classifications of the universe of subjects can be structured: is it at all possible to create a universal classification scheme that is independent of disciplines, so that the notion of basic facet is related to individual concepts and facets are thus facets of concepts?

A large portion of the CRG's work in the 1960s and 1970s was devoted to research funded by NATO on a New General Classification Scheme for a universe of subjects in which the initial division into classes was based on the integrative levels of concepts rather than on disciplines. The facets in this type of scheme would thus be facets of concepts or phenomena rather than of disciplines, allowing for the identification of specific conceptual categories governing a citation order that "holds good across the entire spectrum of knowledge, so that, for example, a formula which controls the order of terms in physics applies equally well in music and politics" (Austin, 1976, p. 164). Foskett (1961) proposed that conceptual categories in the New General Classification could be identified by linguistic traits: facets representing nouns would belong to a category of Things, while facets representing verbs would belong to a Process or Energy category (p.138). He notes that "this is the reverse of the traditional procedure, which is the choice of main class followed by enumeration of the terms in its first facet; here, we enumerate the facet and then try to set boundaries at appropriate points" (Foskett, 1961, p. 139).

Although the CRG concluded after over a decade of research that such a faceted scheme was impossible to achieve, their research on discipline-independent facets and categories provides an important theoretical groundwork for extricating facets from disciplines[13]. The Integrative Level Classification project (2004) has recently resumed the investigation into a nondisciplinary classification scheme. The main coordinator of the project, Claudio Gnoli (2006), argues that "the prevalence of the use of facets in websites for concrete concepts and phenomena" combined with the increasing interdisciplinarity of knowledge, demands that "the meaning of facets cannot depend on a limited list of disciplines."

---

[13] Work on the project laid the foundation for the development of the subject indexing system, PRECIS, by CRG member Derek Austin (Austin, 1984).

In this research, it is hoped that knowledge about the two fundamentally different ways to construct universal faceted classification schemes can be used in the search for basic facets in the tags representing the universe of subjects. Likewise, knowledge of how special faceted classifications are constructed will guide in the search for basic facets in the universe of books.

2.4.3 The Postulate of Isolate Facet

Ranganathan's final postulate for use in choosing facets is the postulate of isolate facet. Here, he posits that "each isolate facet of a compound subject can be deemed to be a manifestation of one and only one of the five fundamental categories. It is generally easy to identify isolate ideas that are manifestations of the Fundamental Categories: MEST. Any isolate idea not found to be a manifestation of any of these four categories, has a good chance to be a manifestation of P. Its manifestation can also be directly sensed in some cases" (Ranganathan, 1967, p. 403).

This is a fairly straightforward postulate concerning the relationship between facets and categories. Although a category can be represented by several different facets in a document, each facet represents one and only one category. For Ranganathan, these categories are PMEST; for Vickery and the CRG, they are whatever the classificationist postulates upon examination of relevant literature.

2.4.4 Summary of the postulates to be used in this research

Based on the above discussions of Ranganathan's original three postulates for the identification of facets, and based on knowledge of the universes to be facet analyzed and the metadata representing them, three postulates will be used in the facet analysis of the folksonomy in this research:

1. An examination of the metadata will reveal conceptual categories to which all the facets in the universe to be classified belong. The prior recognition of Ranganathan's PMEST categories will facilitate the endeavor.

2. An examination of the categories and facets will reveal explicit or implicit basic facets which represent classes in the universe to be classified. Knowledge of potential faceted classification structures will be helpful here. All the facets found will be facets of these classes while the categories hold true for the entire universe.

3. All the explicit or implicit facets found will belong to one and only one of the conceptual categories found. By extension, each tag in the user-generated metadata will belong to one and only one facet.

The practical application of these postulates will be discussed in the methodology chapter, section 3.3.

3   Methodology

In this section, I will first discuss the folksonomy chosen for the dataset in this research along with the tagging system whence it comes. Thereafter, I will describe the facet analysis techniques used on the folksonomy based on the postulates formulated in section 2.4.4.

3.1   Data source

There were two criteria used for choosing a tagging system whence to gather tags for use in the dataset in this research. The first criteria concerned the objects of the tags in the tagging system. Due to the fact that most of the original literature on facet analysis concerns uses of the technique for the organization of the subjects of books or other documents, it was desirable to find a tagging system in which the objects of the tags are books. A facet analysis of aggregated tags in which some of the tags presumably represent the universe of subjects would thus ensure some consistency with the original literature. Websites that allow users to tag representations of books include the on-line sales company, Amazon.com[14], the university library project, PennTags[15], and the social book cataloging sites aNobii[16], Goodreads[17], LibraryThing[18], Shelfari[19] and WeRead[20].

The second criteria used in the selection process concerned the size of the folksonomy. Spalding (2007a) compared the tagging of books at LibraryThing to the tagging of books at Amazon and found that LibraryThing users generate ten times as many tags per book than Amazon users do, despite receiving ten times less traffic. He attributes this phenomenon to motivation, remarking that "tagging works well when people tag 'their' stuff, but it fails when they're asked to do it to 'someone else's' stuff." He also summarizes the importance of the abundance of tags: "to do anything useful with tags, you need numbers …[and] with a larger number of tags, clear patterns emerge" (Spalding, 2007a). At the time of this writing, the entire LibraryThing folksonomy

---

[14] http://www.amazon.com/
[15] http://tags.library.upenn.edu/
[16] http://www.anobii.com/
[17] http://www.goodreads.com/ (At Goodreads, tags are called *shelves*.)
[18] http://www.librarything.com/
[19] http://www.shelfari.com/
[20] http://weread.com/

comprises over 58 million tags[21], making it much larger than all of the abovementioned sites. Tags from LibraryThing were therefore selected for use in this research.

3.1.1 LibraryThing

Launched on August 29, 2005, LibraryThing is a social networking website where users can catalog and share their book collections, thus enabling people with similar tastes in books to connect (LibraryThing, 2009). To date, the website has more than 940,000 members who have cataloged over 45 million books representing nearly 5 million individual works[22].

Creating a collection at LibraryThing is intended to be easy for users. In a single search box, the user may search for a book using title or author keywords, the ISBN or tags. Using the client-server protocol Z39.50, LibraryThing then searches catalog data at the Library of Congress, Amazon and 80 other libraries and allows the user to browse the returned titles and to click on the title he or she would like to add. Title, author, date of publication and a book cover image are then automatically imported to the user's collection. It is also possible to add bibliographic data manually if, for example, the book cannot be located at one of the abovementioned sources.

Users can choose whether to make their collections public or private; public is the default option. All users with public collections are automatically connected to people who have similar collections. One may then post comments to other members, join groups and partake in discussions. Additionally, statistics (called Zeitgeist) about one's collection are gathered based on user-generated Common Knowledge[23] about each book and based on the user's logged activities, including how many total and how many distinct tags the user has used, how many reviews he or she has written and how often books he or she adds to the collection each day. Aggregated statistics for all users are collected and presented on the Zeitgeist overview page[24].

3.1.1.1 LibraryThing users

It has been speculated that a relatively high proportion of LibraryThing users[25] are librarians by profession and that this may distort the tags such that they more closely resemble library subject

---

[21] http://www.librarything.com/zeitgeist (Retrieved December 1, 2009).
[22] http://www.librarything.com/zeitgeist (Retrieved November 27, 2009).
[23] http://www.librarything.com/wiki/index.php/Common_Knowledge
[24] http://www.librarything.com/zeitgeist
[25] LibraryThing users are also known as *thingamabrarians*.

headings than would tags generated by non-librarians. In a master's thesis examining tags in the LibraryThing folksonomy, Sterken (2008) compared tags generated by LibraryThing members in general with tags generated by LibraryThing members who belong to the group *Librarians who LibraryThing*. He then compared the same tags to Library of Congress subject headings and found that only 21.24%[26] of the analyzed tags were equal to the associated subject headings and that "there were no really significant differences between the group of librarians and of the non-librarians" (Sterken, 2008, p. 63). Although his assumption that LibraryThing users who do not belong to the abovementioned group are thus non-librarians is unlikely, the comparison between tags and subject headings is a good indication that LibraryThing tags differ significantly from Library of Congress subject headings. This is consistent with the cursory comparison performed in this research, which will be discussed in more detail in section 3.2.

3.1.2 Tagging at LibraryThing

There are several factors affecting tag choice that are important to consider when analyzing aggregated user-generated metadata. The first concerns the availability of other tags at the time of tagging. In some tagging systems, users are presented with all the tags that previous users have used to describe the resource in question. At LibraryThing, none of the existent tags for the book in question are shown when the book is first cataloged. Although it is possible at any time to click on any book and edit tags while looking at the other users' tags for the book, tagging is usually done at the user's home page, removed from other users' tags. This presumably leads to a more accurate description of what most users find most important about each book, based on the aggregation of "uninfluenced" tags. Checking this hypothesis, however, is beyond the scope of this research.

Until last fall, LibraryThing users could combine tags. This is a service intended to make up for synonymy problems associated with tags and intended to reinforce the consensus aspect of aggregated tags. A user could, for example, decide to combine the Norwegian tag *sakprosa*[27] with the tag *nonfiction*, thus making all the books he or she had labeled *sakprosa* available when others search for the tag *nonfiction*. Although it is possible to see the results of tag-combinations on each tag's page, the service was "taken down for a day or two" on September 8, 2008 and is

---

[26] This percentage was adjusted by Sterken from 36% to take into account the high occurrence of the tag *fiction*.
[27] *Sakprosa* is the Norwegian term for *nonfiction*.

not up at the time of writing[28]. *Figure 3* shows two examples of the results of tag combining, for the tags *non-fiction* and *history*:

Tag info: non-fiction

Includes: non-fiction, *non-fiction, *sachbuch, @nonfiction, A:unfiction, Genre: non-fiction, Non Fictioin, Non Fiction, Non- fiction, Non-Fiction **, Non-Fiction., Non-Fiction;, Non-fictie, Non-fiction , Not-fiction, "non fiction", ^Nonfiction, facklitteratur, genre - non fiction, no-ficcion, nofiction, non fic, non-fcition, non-fic, non-ficion, non-ficition, non-ficiton, non-fictin, non-fictional, non-fictios, non-ficton, non-fistion, non-fitction, nonfic, nonficion, nonficition, nonficiton, nonfictin, nonfiction, nonfiction., nonficton, não-ficção, sachbuch, sakprosa

Tag info: history

Includes: history, history, @history, Hiistory, Hisoty, History. ., ^History, geschichte, geschiedenis, hietory, hisotry, hist, histoey, histoire, historia, history., histpry, histroy, histry, história, hitory, hsitory, hsstory, storia, 歷史

Figure 3: Examples of the results of tag combining.

3.1.3 Folksonomy levels at LibraryThing

Folksonomies are presented on three distinct levels at LibraryThing. At the user-level, all of the individual user's tags are presented. These can be viewed alphabetically or by frequency. At the book-level, all of the tags that have been used by all users for each individual book are presented. By default, only the most popular tags are shown and they are displayed in a tag cloud. It is possible, however, to view all of the tags associated with the book and to see the frequency of each tag. On the Zeitgeist page, two universe-level folksonomies are presented: a list of the 75 most popular tags in the entire folksonomy and a list of the 50 longest tags[29] in the entire folksonomy. These are both presented by frequency. An additional way of viewing tags is on each individual tag's page, where a list of the top books tagged with that particular tag is presented. On the "tag page", it is also possible to see other tags that have been combined with that tag (see *Figure 3*) and a list of related tags, based on how frequently they are used on the same book.

---

[28] http://www.librarything.com/tagcombine.php?tag=tags (Retrieved November 30, 2009).
[29] Tags with more than 20 letters.

The folksonomy used for this research was comprised of tags gathered at the book-level.

3.2   Data collection

The dataset of tags from LibraryThing was constricted to those depicting non-fiction books from a specific domain, namely history. This was accomplished by creating a TagMash[30] with the tags *history* and *non-fiction* (see *Figure 4*). TagMash is a feature offered at LibraryThing to "close some of the gap between tagging and professional subject classifications" (Spalding, 2007b). A semi-automated process, TagMash allows users to create searches with two or more tags, yielding results based on the intersection of the tags. A TagMash created with *history* and *non-fiction* yielded 45 of the most popular tags for each of the 250 most popular books tagged with both the two tags. The fact that only the 45 most popular tags for each book were included in the research strengthens the likelihood that the tags represent a consensus of what users find most important, as the long tail of the Power Law is excluded (Halpin et al, 2007).

---

[30] http://www.librarything.com/blog/2007/07/tagmash.php
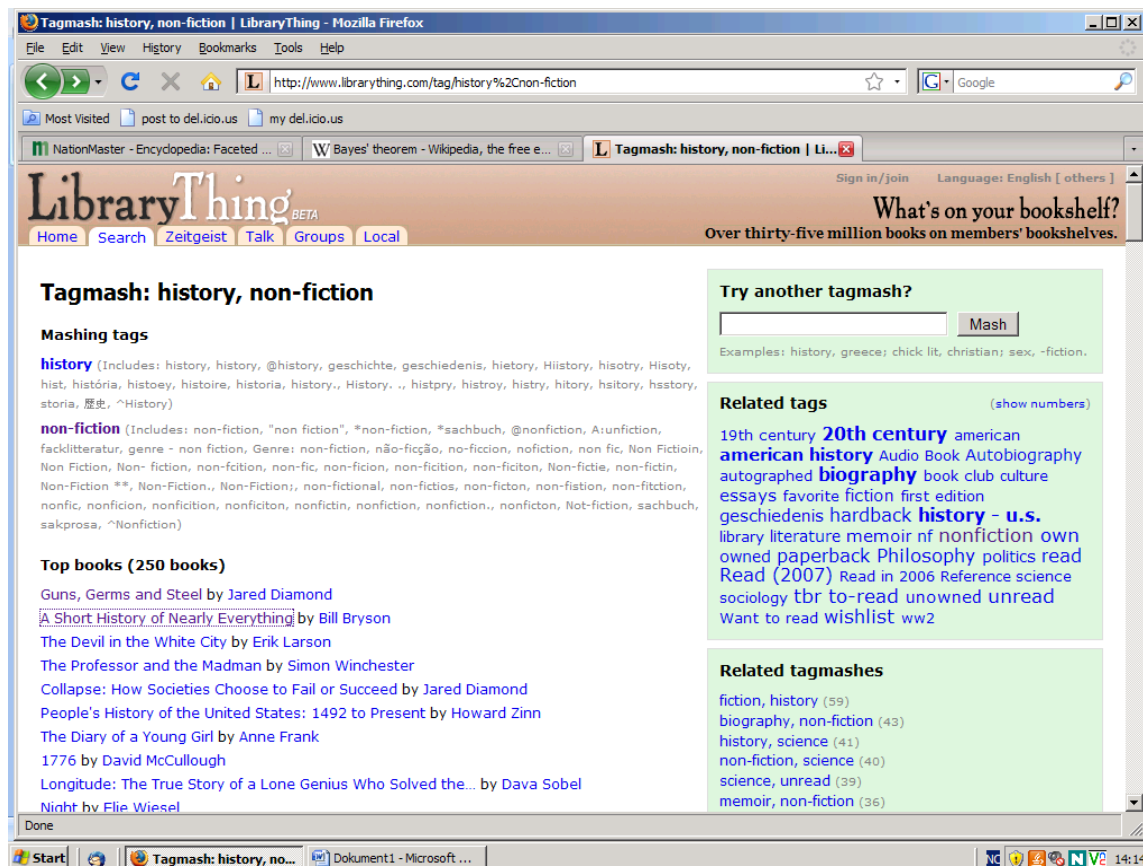
Figure 4: Screenshot of LibraryThing's Tagmash of History and Non-fiction tags

The dataset was further constricted to include only tags representing those books that had also been indexed with the subject heading 'history' by the Library of Congress. This was done in an attempt to constrict the dataset to tags representing books that belonged to a specific domain, namely the discipline History. Only 76 of the original 250 books (30.4%) were found to be given the LOC subject heading 'history' (see *Appendix 1* for list of titles of books used). These books had an average of 2070 LibraryThing members each[31]. Contextual information about precisely which users used each tag was not included in the dataset.

---

[31] The most popular book on the list was *Guns, Germs and Steel*, which had 10,071 members on January 14, 2009. The least popular on the list that day was *The Suspicions of Mr. Whicher*, with 514 members.

The final dataset consisted of 107,375[32] instances of 1,288 unique tags depicting 76 non-fiction books. These were cut and paste from the LibraryThing website to a Word document for further manipulation (see *Figure 5*). An id is affixed to the title of each book with an underscore and the number of instances of each tag (frequency) is shown in parentheses.

1_Guns, Germs and Steel / Jared Diamond
Agriculture(43) anthropology(581) archaeology(48) biology(79) Civilization(169) cultural studies(19) culture(128) culture diffusion(18) development(21) disease(35) ecology(55) economics(57) environment(62) epidemiology(19) ethnology(48) evolution(114) geography(149) germs(14) history(1,452) jared diamond(17) Natural History(22) nf(15) non-fiction(808) own(52) paperback(15) politics(38) popular science(24) prehistory(16) pulitzer prize(68) read(105) Science(454) social evolution(49) Social History(33) social science(42) societies(13) society(95) sociology(261) tbr(25) technology(70) unfinished(14) unread(102) war(29) WishList(23) World(36) World History(145)

3_The Devil in the White City/ Erik Larson
1893(12) 19th Century(44) 2005(13) 2006(20) 2007(18) America(23) American(25) American history(94) architect(12) architecture(161) Audiobook(14) Biography(32) book club(31) borrowed(13) chicago(525) Chicago history(17) Chicago World's Fair(41) columbian exposition(26) crime(150) Daniel Burnham(12) Ferris Wheel(18) fiction(84) historical(43) historical fiction(51) history(637) Illinois(21) library(13) murder(147) mystery(59) nf(13) non-fiction(626) Novel(17) own(29) read(98) serial killer(177) tbr(30) Thriller(19) to read(14) true crime(191) united states(18) unread(51) us history(20) world fair(13) world's columbian exposition(20) world's fair(201)

Figure 5: Excerpt of tags per book, showing 2 of 76 books

3.2.1 Systematization of data

In order to be able to look for facets and categories in the folksonomy as a whole, an overview of all the tags was necessary. To create this, each tag was manually cut and paste into an alphabetically ordered list of individual tags. In the list, contextual information about which book each individual tag belonged to and how frequently the tag was used on each book is retained (see *Figure 6*). In this way, although the tag was analyzed out of the context in which it was used, it was always possible to check the tag in the context of the book to which it belonged.

---

[32] This number includes 34 instances of 13 tags that were deemed by me to be too ambiguous to classify. These tags were taken out of the dataset and are not used when calculating percentages of the total dataset, making the total: 107,341 instances of 1,275 distinct tags. This will be discussed in more detail in section 4.4.3.

3_2005(13) 14_2005(11) 98_2005(5) 131_2005(4) 244_2005(5)

3_2006(20) 8_2006(6) 12_2006(5) 14_2006(13) 19_2006(3) 32_2006(4) 42_2006(5)
45_2006(5) 71_2006(3) 81_2006(2) 94_2006(7) 98_2006(5) 103_2006(4)
134_2006(3) 165_2006(11) 244_2006(5) 249_2006(2)

3_2007(18) 8_2007(10) 12_2007(5) 14_2007(12) 19_2007(3) 27_2007(5)
32_2007(4) 42_2007(6) 45_2007(11) 46_2007(4) 70_2007(3) 71_2007(5)
85_2007(4) 94_2007(7) 97_2007(9) 103_2007(13) 121_2007(3) 134_2007(7)
146_2007(19) 165_2007(11) 244_2007(3) 249_2007(3)

71_2008(3) 97_2008(2) 146_2008(5) 221_2008(3)

131_21st Century(2)

63_900(3)
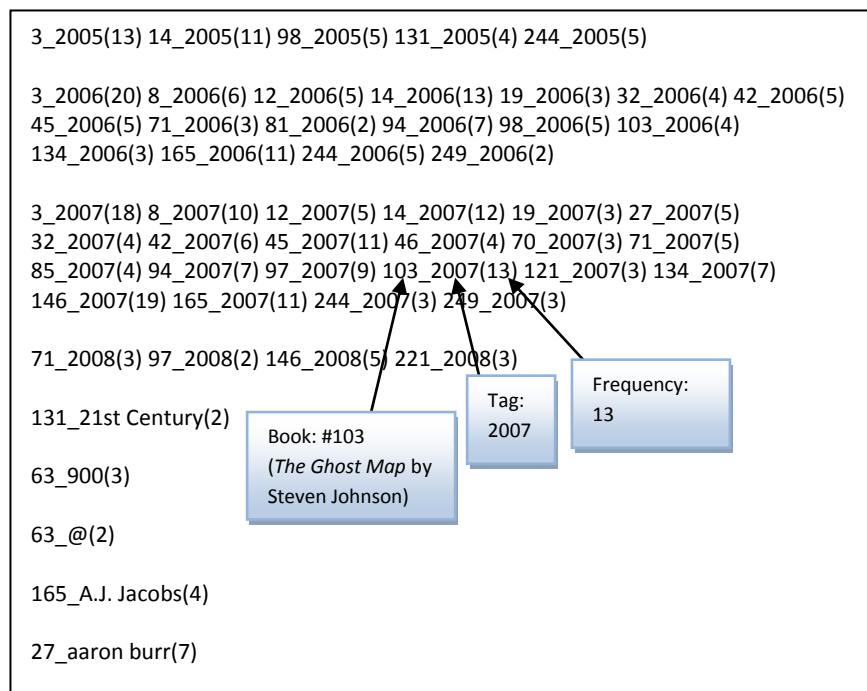
63_@(2)

165_A.J. Jacobs(4)

27_aaron burr(7)

Book: #103
(*The Ghost Map* by
Steven Johnson)

Tag:
2007

Frequency:
13

Figure 6: Excerpt of alphabetical list of tags, showing 9 of 1,275 unique tags

3.3   The postulational approach to a facet analysis of a folksonomy

The facet analysis techniques used in this research to facet analyze the list of tags in *Figure 6* are based on the faceted classification theories discussed in chapter 2. In the rest of this chapter, I will delineate the method I followed. Potential shortcomings of the method will be discussed in section 4.4.3, after a presentation and discussion of the results.

The method followed in this research was a non-linear and highly iterative process aimed at placing each tag in a mutually exclusive facet. Facet is defined as a category resulting from the application of a single principle of division. As discussed in section 2.3.8, the facet analysis of the folksonomy takes place before the classificationist has analyzed the universe to which the 76 books represented by the tags in the datastet belong. This means that facets differentiating aspects of the universe of 76 books have not yet been identified. The process involved in placing each tag in a facet is thus the reverse of the procedure normally performed by the classificationist; universes, classes, categories and facets are identified solely based on the analysis of the tags. The ensuing faceted classification ontology can therefore be said to be wholly inductive.

As discussed in section 2.4.4, the following postulates are proposed to serve as guidelines for the classifier throughout the process of facet analysis:

1. Look for conceptual categories to which all the facets in the universes to be classified belong. Use PMEST as a starting point.

2. Look for explicit or implicit basic facets. These represent classes in the universes to be classified.

3. All the explicit or implicit facets found will belong to one and only one of the conceptual categories found. By extension, each tag in the user-generated metadata will belong to one and only one facet.

In section 3.3.1, I will roughly explain the techniques used to facet analyze the dataset. In section 3.3.2, I will discuss techniques used to compensate for ambiguous and obscure tags. Although they are explained separately, it is important to reemphasize that the entire process was highly iterative; so when ambiguous or hard-to-interpret tags were encountered in the process described in section 3.3.1, they were evaluated with techniques described in section 3.3.2 and the initial analysis was readjusted as needed before the process continued.

3.3.1 Facet analysis process

Based on the above definition of facet and the proposed postulates, an algorithm was developed to use in the initial analysis of each tag (see *Figure 7*). The main reason for using the algorithm was to make the dataset more manageable by sorting the tags into smaller groupings. This would presumably facilitate in the identification of the facets, basic facets and the remaining conceptual categories by providing a systematic overview of the types of tags present in the dataset. The algorithm applies Ranganathan's Method of Residues, which is a technique intended to aid the classifier in figuring out the conceptual category to which identified facets belong. According to the Method of Residues, "if a certain manifestation is easily determined not to be one of 'Time', 'Space', or 'Energy', or 'Matter', it is taken to be a manifestation of the fundamental category 'Personality'" (Ranganathan, 1967, p. 401). Thus, it is posited that tags belonging to Time, Space and Energy categories exist in the dataset, and these are the first categories that are sorted out of

the dataset. Based on knowledge of the dataset, the algorithm presupposes that all of the tags represent the universe of books, which includes the universe of subjects.
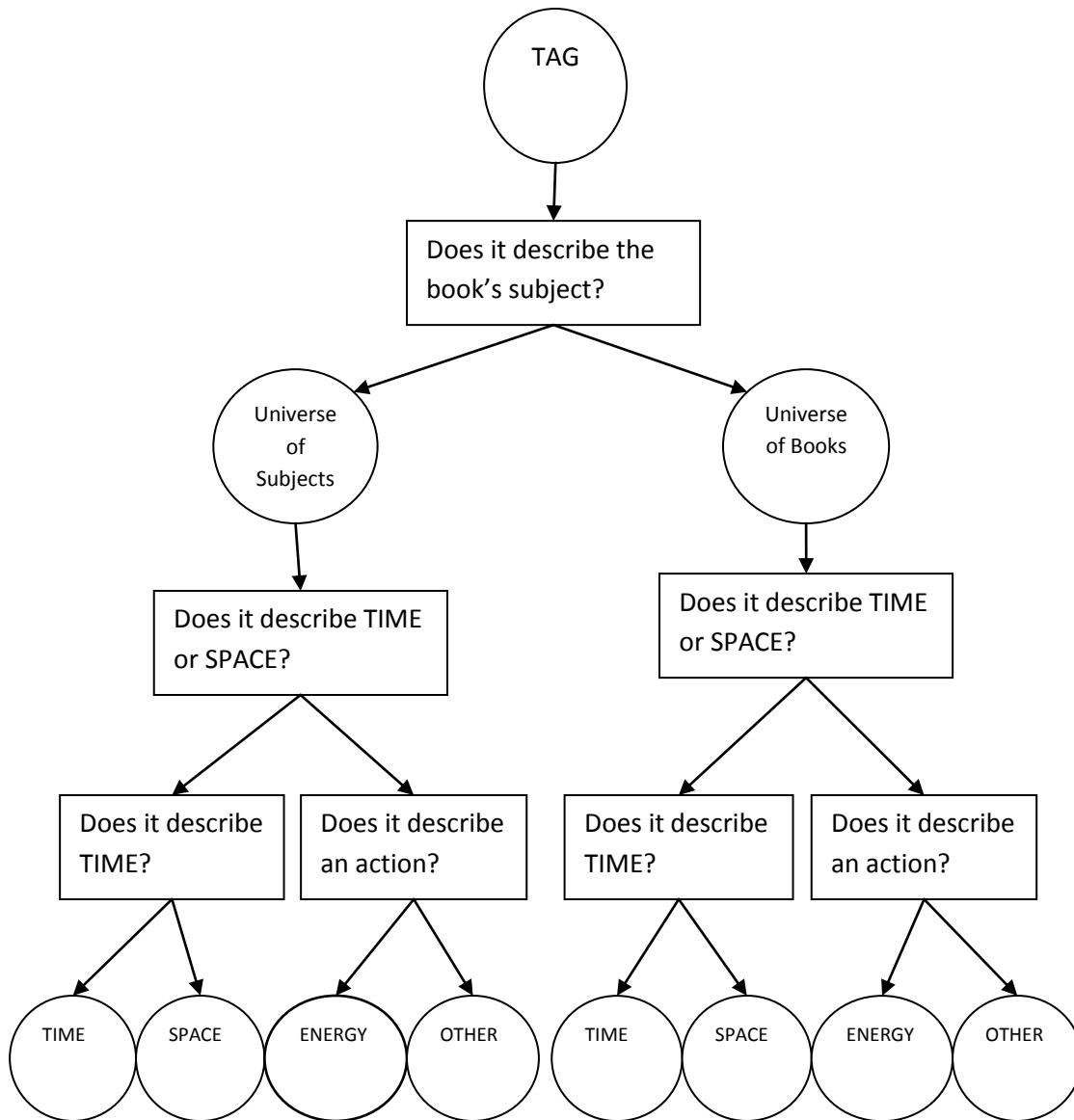
Figure 7: Initial algorithm used to analyze tags into conceptual categories

The first question in the algorithm pertains to whether the tag describes the book's subject. A very simple test was used to establish an initial feeling about whether the tag indicated aboutness: each distinct tag was imagined in the following sentence: "The book is about [*tag*]". For example, the sentence "The book is about *boy soldiers*" is meaningful, while the sentence "The

book is about *currently reading*" is not. If the sentence was not meaningful, the tag was interpreted as indicating an aspect other than the subject of the book. If the sentence was meaningful, the tag was checked for ambiguity (see section 3.3.2).

After the initial distinction between tags indicating subject and tags not indicating subject was made, attempts were made to identify general conceptual categories to which each tag belonged. Using Ranganathan's Fundamental Categories and his Method of Residues as a basis upon which to look for categories, tags were analyzed in both of the universes to see if they represented either Time or Space. Time and Space are considered to be universal categories and easily identifiable (ibid, p. 399). Attempts were then made to identify tags belonging to the Energy category. A simple linguistic analysis was used to identify which tags belonged to the Energy category: tags were placed in the Energy category if they consisted of verbs or action nouns indicating events or things that have happened.

The completion of the initial analysis of the tags with the above algorithm resulted in a rough division of the original dataset into eight categories: Time, Space, Energy and Other in both the universe of books and the universe of subjects. Each of these was then concurrently examined for facets and basic facets, and the search for more conceptual categories continued within the two "Other" categories. The identification of facets, basic facets and conceptual categories in this stage of the analysis was an ad hoc process in which tags were grouped together based on linguistic or operational similarities and then tested for the following criteria:

To ascertain that the grouping represented a facet, the following criteria had to be fulfilled:

- Facets are the results of a single principle of division
- A facet is a facet of a class, which is represented by a basic facet
- Every facet belongs to a conceptual category

To ascertain that the grouping represented a basic facet, the following criteria had to be fulfilled:
- Basic facets represent classes in the universe
- Classes are differentiated by facets

To ascertain that the grouping represented a conceptual category, it had to differentiate the entire universe and contain at least one facet.

The grouping together of tags was thus a highly iterative process in which the above criteria were checked, and adjustments and readjustments to the groupings were made accordingly. In this sense, the facet analytical process can be compared to puzzle-solving; the verified identification of facets often led to the identification of either implicit or explicit basic facets and the identification of conceptual categories often led to the identification of facets therein.

According to the rule based on the Postulate of Isolate Facet used in this research, each tag was placed in one and only one facet, and each facet was placed in one and only one category. This led to difficulties when compound tags were encountered. Into which facet should the compound tag placed? Compound tags were initially sorted out of the dataset. Upon completion of the facet analysis, if facets had already been identified for each aspect of the compound tags, then the compound tag was placed in the facet that was deemed to be least concrete. For example, *medieval europe* is a compound tag made up of *medieval ages*, which represents the facet "by Time" and *europe*, which represents the facet "by Place". After ascertaining that both of the facets had already been identified in the dataset, the tag was placed in the "by Time" facet. If a facet of a compound tag had not already been identified, the tag was placed in the new facet.

3.3.2 Compensatory techniques for ambiguity and obscurity

There were several techniques used throughout the analysis with the intention of compensating for the inherent ambiguity and obscurity involved in analyzing user-generated metadata.

Tags that could potentially indicate two or more aspects of books are here referred to as ambiguous tags. In most cases, ambiguous tags were easily resolved by examining them in their original context. To do this, summaries and reviews[33] of a few of the books that they represented were retrieved in order to gauge the forms and subject matters of the books. An example of an ambiguous tag is the tag *essays*. "This book is about *essays*" is a meaningful but semantically ambiguous sentence; a book can both be about essays and contain essays. This particular tag is used on 11 books in the dataset. The meaning of the tag was checked against summaries of a couple of these books, which revealed that both of the books contained essays. The tag was thus interpreted as belonging to the universe of books instead of the universe of subjects. A final

---

[33] The summaries and reviews were retrieved from Amazon.com and/or Wikipedia.

check was performed for all of the tags against all of the books upon completion of the facet analysis, as will be shown in section 3.3.3.

There were a number of tags that were still ambiguous even after double checking them against the books that they represented. In these cases, an interpretation of the tag was chosen based on what Ranganathan calls flair. Flair is described as "the limiting point between intellection and intuition" (Ranganathan, 1967, p. 550); it is here interpreted as being a gut feeling based on experience.

Tags that I was unable to interpret are here referred to as obscure tags. Upon encountering an obscure tag, a number of strategies were used in the attempt to unveil its meaning. First of all, the tag's page on LibraryThing was examined in order to gauge which other books the tag represented, how many users use it and which other tags are often used with it. In this way, tags like *box 3* became meaningful in the context of other tags used by the user (*box 1, box 2 …*). If the tag remained obscure after checking it against users and related tags, a Google search was performed on the tag. In this way, it became evident that *sa* is a common abbreviation for Shelf Awareness[34] and *tpb* an acronym for Trade Paperback[35].

Thirteen distinct tags in the dataset were deemed to be too ambiguous or too obscure to even guess at what they indicated. These were taken out of the dataset and will be discussed in more detail in section 4.4.3, along with a general discussion of the difficulties involved in interpreting user-generated metadata. Here, it is suffice to say that while each instance of a tag was placed in one and only one facet, the multiple instances of the tag could be interpreted to indicate different things in different books and thus be placed in different facets.

Upon completion of the facet analysis, a final verification of all of the tags was made by checking them in their original context, i.e. against the books they originally described in the dataset. To do this, tags were regrouped back to the book level with information about which facets they belonged to in order to verify that their placements within facets indeed made sense (see *Figure 8*

---

[34] *Shelf Awareness* is a "free e-mail newsletter dedicated to helping the people in stores, in libraries and on the Web buy, sell and lend books most wisely." http://www.shelf-awareness.com/ (The meaning of the tag *shelf awareness* was also discovered with a Google search).

[35] *Trade Paperback* is "a paperback book of a size similar to a typical hard-cover book, intended for sale in bookstores as distinguished from a cheaper and smaller paperback intended for sale on racks at drugstores, newsstands, etc." http://www.yourdictionary.com/trade-paperback

in chapter 4). In this way, each book was examined with faceted tags and it was possible to identify tags that may have been misplaced. At this stage, for example, it became clear that the ambiguous tag *reading* was most likely misplaced in three of the five books with which it was tagged. After the initial analysis, it had been interpreted as belonging to the Energy category in the universe of books, indicating an action that is performed by the user to the work of the book. It became clear during the verification of results, however, that while this most likely was a correct placement for the instances of the *reading* tag that represented *London: the Biography* and *From Beirut to Jerusalem*, the tag most likely indicates aboutness in the books *A History of Reading*, *The Book on the Bookshelf* and *The Know-It-All: One Man's Humble Quest to Become the Smartest Person in the World*.

Finally, the instances of each tag in each facet were enumerated for each book and placed in an Excel spreadsheet in order to calculate the size of each facet, category, class and universe.

4   Results: Presentation and Discussion

There are two different ways to present the results of this research, corresponding to the two different ways in which a facet analysis is traditionally applied. One can either present the results as facet analyzed books or as facet analyzed universes. At the book-level, focus is on each individual book. Here, the tags of each book are presented with the facets to which they belong and with the categories to which the identified facets belong (see *Figure 8*).

```
1. Guns, Germs and Steel / Jared Diamond [TOTAL: 5682]

PERSONALITY:
        BY SUBJECT: [TOTAL: 4438]
                BASIC FACET: Agriculture(43) anthropology(581) archaeology(48) biology(79) cultural
                studies(19) ecology(55) economics(57) ethnology(48) geography(149) history(1,452)
                Natural History(22) popular science(24) Science(454) Social History(33) social science(42)
                sociology(261) technology(70) World History(145) [TOTAL: 3582]
                PERSONALITY: [TOTAL: 481]
                        BY ENTITY: Civilization(169) culture(128) environment(62) societies(13)
                        society(95) germs(14) [TOTAL: 481]
                ENERGY: culture diffusion(18) development(21) disease(35) epidemiology(19)
                evolution(114) politics(38) social evolution(49) war(29) [TOTAL: 323]
                SPACE: World(36) [TOTAL: 36]
                TIME: prehistory(16) [TOTAL: 16]
MATTER:
        BY GENRE:
                BY TRUTH: nf(15) non-fiction(808) [TOTAL: 823]
        BY FORMAT: paperback(15) [TOTAL: 15]
ENERGY:
        BY ACTIVITY (WORK): read(105) tbr(25) unfinished(14) unread(102) [TOTAL: 246]
        BY ACTIVITY (PHYSICAL OBJECT): own(52) WishList(23) [TOTAL: 75]
AGENT:
        BY AUTHOR: jared diamond(17) [TOTAL: 17]
EXTERNAL RECEPTION:
        BY AWARD: pulitzer prize(68) [TOTAL: 68]
```

Figure 8: Book-level presentation of results of the facet analysis. Showing 1 of 76 books.

At the universe-level, focus is on the implicit universe to be classified corresponding to the entire facet analyzed folksonomy representing 76 books. Here, an ontological model of the universe is presented with all of the identified facets along with the conceptual categories to which they belong. Examples of tags are given with each facet, but they are rarely connected to the actual item they represent, serving more as examples of descriptors of all books. Due to the obvious

disadvantages pertaining to the amount of physical space that a book-level presentation requires, plus the fact that this type of presentation obscures the delineation of the aggregated results of the facet analysis, a universe-level presentation of the results has been chosen as the basis for discussion in the rest of this chapter (see *Table 1, Table 2* and *Figure 9*). For more examples of book-level results, see *Appendix 2*.

In section 4.1, I will present an overview of the results of the facet analysis and discuss the relationships discerned in the induced ontological model of the universes of books and subjects. In sections 4.2 and 4.3, I will explain the results in detail and discuss the basic facets, categories and facets that were identified in each universe. In section 4.4, I will discuss general observations and some of the difficulties and challenges met during the course of the facet analysis of the folksonomy.

4.1   Overview of Results

Over 107,000 instances of 1,275 unique tags representing 76 history books make up the folksonomy analyzed in this research. Subjecting them to a facet analysis resulted in the discernment of two distinct implicit universes: the universe of books and the universe of subjects contained within the universe of books. Basic facets, conceptual categories and facets were identified in the tags representing each of the universes (see *Table 1*, *Table 2*, and *Figure 9*). Basic facets were identified implicitly in the universe of books (books as physical objects and books as works) and explicitly in the universe of subjects (subjects as disciplines). These represent here the top-level classes in each of the universes. The conceptual categories discerned in both the universes are based on those postulated by Ranganathan: Personality, Matter, Energy, Space and Time. All of these were identified in the universe of books, while Personality, Energy, Space and Time were identified in the universe of subjects. An additional two categories were found that apply solely to the universe of books: Agent and External Reception. As will be shown, while it was fully possible to facet analyze the metadata representing the universe of books, results of the facet analysis of the metadata representing the universe of subjects remain incomplete.

In the *Table 1* and *Table 2*, each universe is presented with the conceptual categories and facets identified in each. The total number of tags representing each category and facet are shown in

parentheses next to the category or facet name as a percentage of the total number of tags in the dataset. For example, 75,858 of the 107,341 instances of tags in the dataset (70.67%) belong to facets in the Personality category in the universe of books; 75,713 of these belong to the facet "by Subject", 130 to the facet "by Type" and 15 to the facet "by Title". The basic facet in the universe of subjects was identified as being implicit, thus accounting for 0% of the dataset.

Universe	Category	Facet	Examples of tags
Universe of Books	Basic Facet (0%)	By aspect (0%)	Work (*implicit*), Physical Object (*implicit*)
	Personality (70.67%)	By subject (70.54%)	*See **Table 2: Universe of Subjects***
		By type (0.12%)	audiobook, library book
		By title (0.01%)	the histories
		By isbn (0.003%)	Isbn
	Matter (22.4%)	By genre (21.89%)	historical, mystery, non-fiction
		By binding (0.26%)	hardcover, paperbook
		By version (0.12%)	Translation
		By format (0.07%)	Audio, mp3
		By edition (0.06%)	first edition
		By series (0.01%)	Hinges of History
	Energy (4.95%)	By activity (work) (3.74%)	read, tbr, unread
		By activity (object) (1.17%)	borrowed, own, owned, wishlist
		By process (0.03%)	illustrated, made into movie, translated
	Agent (1.11%)	By author (0.76%)	gibbon, Albert Manguel, Australian author
		By publisher (0.28%)	folio, folio society, penguin classics
		By user (0.07%)	Book club, adult, teen, ya
	Space (0.17%)	By place (0.17%)	library, box 2, storage
	Time (0.36%)	By year written or published or by year read (0.36%)	100s, 1984, 2006, 2007
	External Reception (0.33%)	By source (0.15%)	Comedy Central, daily show, npr, This American Life
		By award (0.11%)	pulitzer prize, national book award
		By new expression (0.05%)	film, movie, Hinges of History
		By rating (0.02%)	favorite, staff pick

Table 1: Universe-level presentation of results of the facet analysis of the Universe of Books

The "by Subject" facet in the universe of books accounts for 70.54% of the dataset. It represents the universe of subjects and has been subjected to a facet analysis of its own (see *Table 2*). As will be explained in section 4.3, the results of the facet analysis of the universe of subjects are inconclusive. Percentages are therefore only given for the conceptual categories.

Universe	Category	Facet	Examples of tags
Universe of Subjects	Basic Facet (52.82%)	By discipline (52.82%)	biology, history, literature, religion
	Personality (16.16%)	By person	Aaron burr, Rasputin, sickert, us president, serial killer
		By group	American Indians, secret societies, marine corps, merovingians
		By entity	animals, Mayflower, theory, codes, map, television, culture, chrisianity, books
	Energy (12.72%)	By energy (find facets?)	cultural diffusion, crime, evolution, murder, politics
	Space (14.94%)	By place	america, boston, college, sea, the west, world
	Time (3.36%)	By time	19th century, 1990s, antiquity, dark ages, renaissance

Table 2: Universe-level presentation of results of the facet analysis of the Universe of Subjects

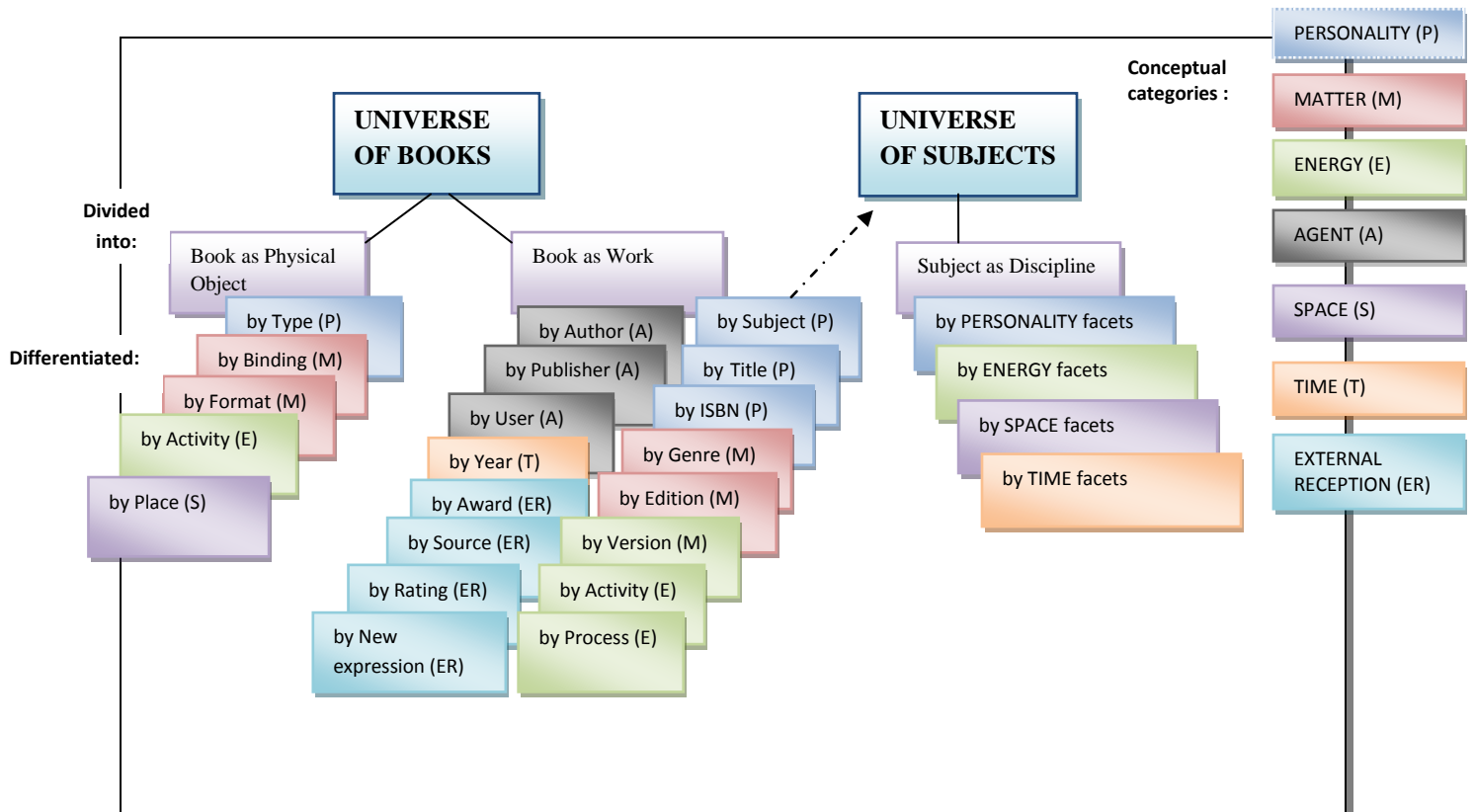### 4.1.1 Ontological relationships at the universe-level

Figure 9: Ontological model of the facet analyzed universes of books and subjects

The ontological model of the facet analyzed universes of books and subjects in *Figure 9* and the exemplified ontological relationships shown in *Figure 10* illustrate some of the relationships discerned in the folksonomy. Each facet evokes a specific "differentiated-by" relationship to the object of which it is a facet via the basic class to which the object belongs. For example, there is a "differentiated-by-author" relationship between the book as Work: *A History of Reading* and the tag: *Albert Manguel*, and a "differentiated-by-format" relationship between the book as Physical Object: *1776* and the tag: *hardcover*. The identification of facets in the tags space is thus significant because facets represent a new way of grouping tags. The most common grouping of tags is the tag cloud, which clusters tags together based on the frequency of tag co-occurences. Here, tags are grouped together based on shared common characteristics that distinguish them from other tags in the tagspace in relation to aspects of the entity they represent.

Unfortunately, intra-facet relationships are not explicit. In faceted classifications, intra-facet relationships are semantic relationships, like synonyms and hierarchical relationships (Broughton, 2006). Thus, the hierarchical relationships between *massachussetts*, *new england* and *united states*, which all belong to the "by Place" facet of the subject of the book, *Mayflower: A Story of Courage, Community, and War*, remain implicit, as do the synonyms *united states*, *us* and *usa* from the same facet. These can be inferred by those with knowledge of the domain, but they are not directly discernable in the model. This is consistent with Kwasnik's analysis of the role of classification in knowledge structures. She notes that one of the major disadvantages of faceted classifications lies in their lack of explicit intra-facet relationships, such that, "in terms of theorizing and model building, the faceted classification serves as a useful and multidimensional description but does not explicitly connect this description in an explanatory framework" (Kwasnik, 1999, p. 42). The consequences of this failure are seen most clearly in the lack of explicit intra-facet relationships between disciplines, discussed in section 4.3.2.
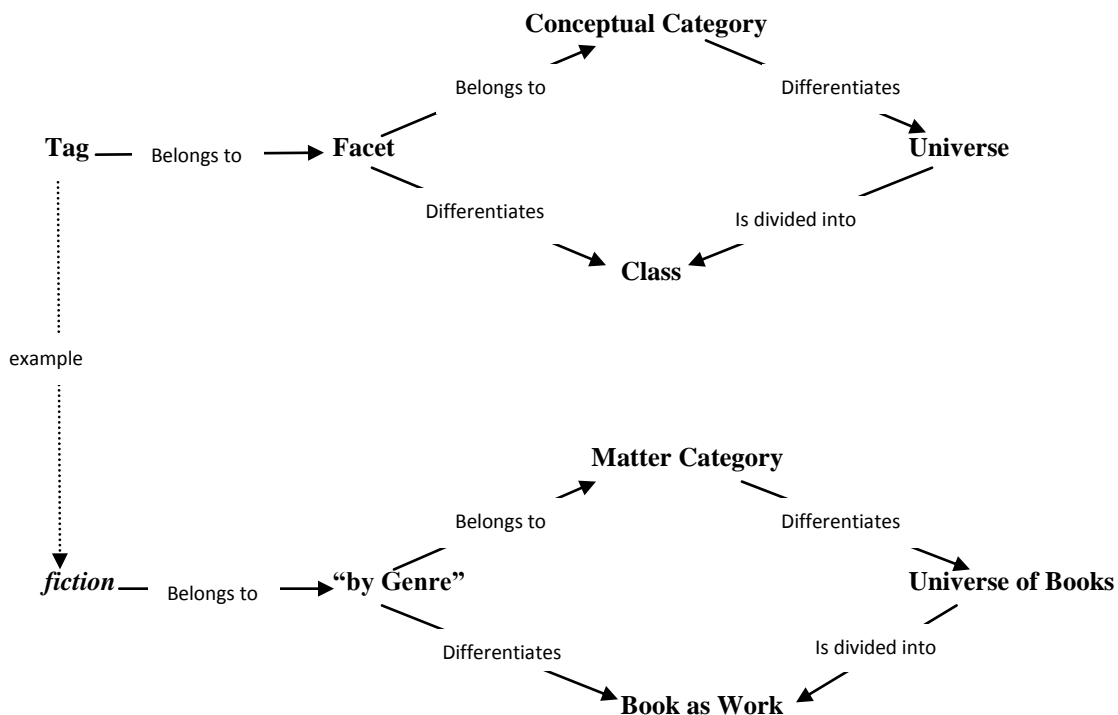
Figure 10: Ontological relationships between tag, facet, category, universe and class (top) with example (bottom)

Inter-facet relationships, on the other hand, are explicit in the model. Inter-facet relationships are syntactic relationships, "the number and variety [of which] seem unique to faceted classification" (Broughton, 2006). Since all tags belong to facets, they inherit both inter-facet and inter-category relationships. Examples of inter-facet relationships in the dataset would include [author]-[activity]-[subject] relationship, as in *orwell-writes about (implicit)-history*; and [user]-[activity]-[place] relationship, as in *book club-borrowed-library*. Additionally, since all facets belong to categories, they inherit inter-category relationships. Broughton (2009) notes that the use of "categories allows general rules to be proposed about the relationships in the domain (as opposed to the relationships between particular pairs of concepts)." Although the conceptual categories borrowed from Ranganathan are by definition diffuse, it has been found that the smaller the universe, the more specified the relationships can be defined. This is consistent with Vickery's work on special faceted classifications (Vickery, 1960). In the finite universe of books, the relationships identified between the categories are fairly specific: [Agent] performs [Energy] on [Personality], [Personality] is composed of [Matter], etc. The relationships are more general in the universe of subjects. Here, the category relationships are more diffuse.

4.2   The Universe of Books

The universe of books here is based on the analysis of tags representing 76 books. As such, it is a finite universe of 76 books. In this section, I will present the basic facets, facets and conceptual categories identified in the tags. These contribute to the exposition of a complete faceted classification ontology representing the universe of books.

4.2.1 Basic facets: Physical Object and Work

As mentioned, while there are no tags in the dataset that directly represent basic facets, the facet analysis revealed that there are at least two implicit basic facets in the universe of books. These basic facets represent the top-level classes in the universe of books; the first basic facet indicates books as physical objects and the second indicates books as works[36]. There are certain facets that

---

[36] The distinction between works and physical objects in the realm of books and other documents is hardly new. It is perhaps most prevalently described by the Functional Requirements for Bibliographic Records (FRBR) model, where the complex relationships between work, expression, manifestation and item (WEMI) are explained in detail.

only apply to one or the other of the basic facets, but conceptual categories have been found that hold for the entire universe.

4.2.2 Physical Object

Physical object is here defined by its relationship to work; the book as physical object encompasses everything about the book that doesn't have to do with the work contained within it. Broughton (2006) remarks, that the faceted classification of physical objects would "not be regarded as faceted classification by purists." According to the purist view, a faceted classification is subject-based, meaning it concerns the conceptual categorization and organization of topics of intellectual thought, whether these are contained in books, articles or other formats. In this light, a faceted classification of a physical object should be considered as a "relatively simple … classification with only one facet (that of the object to be classified) and the classification is restricted to the identification and labeling of arrays within that facet" (ibid). She exemplifies this with a purported faceted classification of the universe of jewelry. Here, two facets representing the two conceptual categories, Type and Material, are proposed. Broughton, however, shows that these are actually two subfacets of jewelry within one conceptual category, Type: jewelry type ("by function") and jewelry type ("by material").

The results of this research suggest that user-generated metadata representing physical objects do indeed possess facets belonging to separate conceptual categories. In fact, two facets relating to users' tasks were discerned, in addition to three facets relating to the properties of physical books. These belong to four distinct conceptual categories: Personality, Matter, Energy and Space. The identification of task-oriented tags is consistent with Kipp & Campbell (2007), who found "tags relating to time and task which suggest the presence of an extra dimension in classification and organization." While they propose that conventional two-dimensional classification systems are unable to facilitate these types of tags. Their unproblematic inclusion here, however, suggests that faceted classifications, which allow for multi-dimensional representations, are ideally suited for the task.

1.8% of all the analyzed tags in this research belong to facets of books as physical objects, presumably reflecting the fact that this aspect of books is of subordinate importance to users. Four conceptual categories are represented explicitly in this class, and an additional one is

represented implicitly. In the following sections, I will briefly discuss each conceptual category in the universe of books, delineating the four facets of Physical Object that were found and the one facet that was implied by the tags. In order to illustrate how all the conceptual categories relate to the Physical Object class, I will also suggest some potential facets that were not found in the dataset.

4.2.2.1 Personality facets of Physical Object

There were two tags in the analyzed dataset that can be said to represent the Personality of books as Physical Objects. These are *audiobook* and *library book* and they belong to the facet "by type". There are 128 instances of the tag *audiobook*, representing 18 books, and 2 instances of the tag *library book*, representing 1 book. Together, they make up 0.12% of the folksonomy.

The tag *library book* is interpreted as being a compound made up of the foci *book (paper)* and *library* (from the "by place" facet). *Book (paper)* is here presumed to be the implicit default tag in this facet; unless tagged otherwise (with, for example, *e-book* or *audiobook*), users are referring to paper books. This is hardly surprising since books by definition are generally considered to be paper books, audiobooks and e-books being seen as variants of paper books.

The placement of *audiobook* and *book (paper)* in the same facet implies that they are here interpreted as belonging to the same hierarchical level. In other words, it is assumed that when users tag a book as *audiobook*, they are referring to the same work as found in other types of books (as physical objects), but differentiating the physical type of book. This is, of course, a debatable contention. An equally valid interpretation of the tag could be that users are emphasizing a new expression of the work, different from other types of expressions of the work, like the one found in the *book (paper)*. This interpretation would have resulted in the tag's placement in the "by New Expression" facet, alongside the tags *movie* and *film* (see section 4.2.3.7).

4.2.2.2 Matter facets of Physical Object

Facets of Physical Object in the Matter category have in common that they include tags that describe the physical composition of books. 0.33% of the tags belong to facets of books as Physical Objects within the Matter category. The facet "by Binding" accounts for 0.26% of the dataset and includes the tags *hardcover, paperback, tpb* and *trade paperback*. The facet "by

Format" includes 0.07% of all the tags and is comprised of two tags: *audio* and *mp3*, the latter of which belongs to a subfacet of the "Format" facet.

4.2.2.3 Energy facets of Physical Object

It was during the analysis of the Energy facets in the universe of books that the distinction between books as Physical Objects and books as Works became clear. One of the facets belonging to the Energy category was clearly different from the other Energy facets: "by Activity". In this facet, we find things that people do to physical books. Examples of tags in this facet are: *borrowed*, *gift*, *keep*, *mooched*, *own*, *owned*, *signed*, *swapped*, *want* and *wishlist*, and they account for 1.17% of the tags in the folksonomy. The tag *own*, which accounts for 0.75% of all the tags, is used on 71 of the 76 books in the dataset. According to LibraryThing's Zeitgeist[37], it is the 18th most popular tag in the entire LibraryThing folksonomy.

With the exception of the tag *signed*, all of the tags in the facet are representative of the temporal tags described by Kipp & Campbell (2007), i.e., they describe Books as Physical Objects in terms of the users' intentions with them. As will be discussed shortly, this brings to light an interesting implicit facet in the Agent category, namely "by User".

4.2.2.4 Agent facets of Physical Object

It was found that "by User" is a major implicit facet in folksonomies, denoting the agent of what is done with books in the abovementioned "by Action" facet. The significance of a user-dimension in a faceted classification will be discussed in more detail in section 4.2.4.2.

4.2.2.5 Space facets of Physical Object

0.17% of the analyzed tags belong to the "by Place" facet in the Space category. Tags in this facet are interpreted to refer to the location of the physical books, like *box 2*, *box 6*, *f porch, home, library*, *loc upstairs north*, *office*, *school* and *storage*. Only 35 of the 76 books in the dataset were tagged with tags belonging to the "by Place" facet of Physical Object. The tag, *library*, accounts for 86.74% of the tags in this facet.

---

[37] http://www.librarything.com/zeitgeist (retrieved 11.22. 2009)

4.2.2.6 Time facets of Physical Object

All of the tags representing time were interpreted in this research as either belonging to the Time category in the Works class (in the universe of books) or the Time category in the universe of subjects. The tags representing Time aspects of the Works could, conceivably, also belong to facets of Physical Objects. In this case, the tags would have to be interpreted as indicating, for example, when the physical book was purchased or borrowed, rather than when the work was read or published.

4.2.2.7 External Reception facets of Physical Object

There were no tags in the dataset that belonged to facets in the External Reception category. A conceivable facet in this category might be books as Physical Objects "by New function", including tags like *decoration*, *furniture stabilizer* and *kindling*. At the time of this writing, however, there were no such tags in the entire LibraryThing folksonomy[38].

4.2.3 Work

98.2% of all the analyzed tags belong to facets of books as works. Work is here defined as it is in the FRBR model, as "a distinct intellectual or artistic creation" (IFLA, 1998). Here, we find traditional subject and genre-based facets and bibliographic facets, as well as nontraditional facets comprised of tags representing user-tasks.

4.2.3.1 Personality facets of Work

Comprising over 70% of the total tags, the personality category consists of tags belonging to the "by Subject" facet (70.54%), the "by Title" facet (0.01%) and the "by ISBN" facet (0.003%). As discussed earlier, the subject facet represents a universe of its own, namely the universe of subjects. Tags representing subject have been subjected to a facet analysis of its own in this research. For results from and a discussion of the facet analysis of the universe of subjects, see section 4.3.

There are 15 instances of two tags representing title: *The Histories* (9) and *Band of Brothers* (6). It is not unlikely that the latter of these tags actually indicates a new expression, namely the popular miniseries based on the book of the same title. The frequency of this tag occurrence,

---

[38] The tag *decoration* is used 847 times by 230 LibraryThing users, but a cursory glance at the titles of books tagged with it indicates that the tag belongs to facets from the universe of subjects.

however, is considered to be so low that its placement is of little consequence. The low occurrence and frequency of tags in the "by Title" facet may be attributed to a bibliographic skew, which will be discussed in more detail in section 4.2.3.4.

There are 3 instances of the tag *isbn* in the dataset. Although these are not connected to a number in the folksonomy, they are here interpreted as representing the title of a facet in which works are differentiated by an International Standard Book Number.

4.2.3.2 Matter facets of Work

There are three facets of Work that are here considered to belong to the Matter category. Work "by Genre" accounts for 21.89% of the total tags, making it the second largest facet identified in the folksonomy. Work "by Version", work "by Edition" and work "by Series" account for 0.12%, 0.06% and 0.01% of the tags respectively.

The "by Version" facet consists of the tag *translation*. Other potential tags in this facet are: *abridgement* and *original text*. Tags identified as belonging to the "by Edition" facet include *first edition* and *arc* (advanced readers' copies). There are eight instance of one tag belonging to the "by Series" facet: *Hinges of History*, indicating the series to which the book *How the Irish Saved Civilization* belongs.

Genre is a diffuse and thorny concept, both difficult to define and difficult to classify. The implications of this will be discussed further in section 4.4.3. Here, genre is regarded as the style or perspective that an author uses or that a user interprets that the author uses, that influences the composition or the reception of the work. In this facet are placed tags indicating traditional genres as well as adjectives interpreted as describing genre as it is defined here.

Nine subfacets were identified in the "by Genre" facet. These are based on an interpretation of what the main focus of the genre appears to be, i.e. which aspect of the work the genre is indicating. It is not inconceivable that a different dataset would reveal additional subfacets.

1. The "by Truth" subfacet of genre contains tags whose main function it is to communicate whether the book is truthful or not. Tags in this subgenre include: *alternate history*, *fiction*, *nf*, *non-fiction*, *true crime* and *true stories*.

2. The "by Plot" subfacet of genre contains tags whose main purpose it is to communicate what the book is about. For example, the genres *biography, science biography, Shakespeare biography, autobiography* and *memoir* all indicate that the book is about people, the latter two about the author him- or herself. The genres *travel writing* and *travelogue* both relate that the book is about travel, *culinary* and *food writing* indicate that it is about food and *nautical* and *sea stories* relates that the plot is about actions taking place on the sea. Likewise, *action* and *adventure* indicate that the plot contains action and adventure; *historical mystery*, *murder mystery* and *mystery* indicate that the plot contains a mystery; *crime* relates that the plot is about a crime. Perhaps less clear are the tags c*ultural*, *historical, popular history*, *political*, r*eligious*[39] and *social*, here interpreted to indicate genres where the plots are respectively about culture, history, politics, religion and society.

3. The genre subfacet "by Form" includes tags that indicate the document's form. Here, we find *diaries*, *essay*, *essays*, *journal*, *novel* and *short stories*.

4. Genre "by Time" includes two tags indicating that the style of the book can be discerned by when it was written: *classic* and *contemporary*.

5. The "by Elicited feeling" subfacet of genre contains tags whose main function it is to communicate the feeling the document intends to convey. For example, the genres *suspense* and *thriller* straightforwardly describe feelings. Although *humor* and *comedy* (and subfacets of comedy, like *satire*) are complex genres with a long tradition, they are here interpreted as indicating that the documents are laughter or amusement provoking, thus focusing on the intended feeling of amusement elicited. Likewise, the tag *funny* is here interpreted as depicting a genre that elicits laughter.

6. The genre subfacet "by Use" includes tags that indicate how the document is used by the user. Here, we find *reference*, *primary source* and *textbook*.

---

[39] In this dataset, the tag *religious* was used six times to describe *The History of God* and eight times to describe *The Holy Blood and The Holy Grail*. The tag was thus interpreted by me as indicating that the plot was about religion. Alternatively, *religious* could be used to describe a feeling elicited by the book, in which case it would belong to the Genre subfacet "by Elicited feeling".

7. Genre "by Method" contains tags whose main function it is to indicate the method used by the book's author, like *commentary*, *journalism*, *parody*, *reportage* and *survey*.

8. The genre subfacet "by Motive" includes tags that indicate motives ascribed to the author. The genres *revisionist history* and *radical history*, for example, both indicate that the author is presenting a new version of an accepted historical account. While *radical history* usually indicates political motives on the author's part, *revisionist history* can be interpreted in two ways: either as a serious scholarly pursuit to correct historical interpretations, or as a politically based attempt to change the way certain historical events are perceived. Here are also included tags indicating the author's apparent motives to push forward a political agenda: *left*, *liberal* and *progressive*.

9. The genre subfacet "by Style" includes tags that indicate the style used by the book's author. Here, we find the tag *academic*[40].

There are many acknowledged genre forms found in the dataset that are comprised of compound tags from two separate subfacets. These include *contemporary fiction, crime fiction*, *historical fiction*, *historical non-fiction*, *political commentary*, *political humor*, *political satire*, *social commentary* and *speculative biography*.

4.2.3.3 Energy facets of Work

There are two facets of Work identified in the dataset that belong to the Energy category: "by Activity" (3.74%) and "by Process" (0.03%). The "by Activity" facet includes task-oriented tags representing what users have done with the work of the book. Here, we find the following tags: *blogging*, *currently reading*, *half-read*, *partially read*, *read*, *read in 2007*, *reading*, *tbr*, *to be reviewed*, *to read*, *unfinished*, *unread*, *want to read*. The neologism *tbr* is a popular acronym for *to be read*.

Unread (1.43%) is the 12th most popular tag in the LibraryThing folksonomy and, after *history* and *non-fiction*, it is the most popular tag by book in this research, being used for all 76 books in

---

[40]In this dataset, the tag *academic* was used four times to describe one book: *The Structure of Scientific Revolutions*. It was thus interpreted by me as indicating the style in which the book was written. Alternatively, *academic* can be interpreted as indicating that a book is about academia, in which case it would be placed in the Genre subfacet "by Plot".

the dataset. *Read* (1.52%) is the 7th most popular tag in the entire LibraryThing folksonomy and it figures in 75 of the 76 books in the dataset[41].

Tags in the "by Process" facet denote transformations that have been made to the work, both during and after composition. Here, we find the tags: *abridged*, *illustrated*, *made into movie* and *translated*.

4.2.3.4 Agent facets of Work

In the Agents category, we find various facets of work with tags representing actors that are responsible for implicit and explicit tags found in the "by Activity" and the "by Process" facets. There are three distinct facets identified in the folksonomy: "by Author" (0.76%), "by Publisher" (0.28%) and "by User" (0.07%).

52 of the 76 books are represented by a tag from the "by Author" facet. These tags, however, are relatively low-frequency tags with an average of just 11 instances per book. The high occurrence and low frequency of tags in the "by Author" facet is very likely the result of a bibliographic skew in the dataset; although users may find it both useful and interesting to differentiate the books in their collections "by Author", there is no need to do so for retrieval purposes because this bibliographic information is already automatically presented on the book's page. As such, it is possible to search for books in the entire LibraryThing site and within one's own library using keywords indicating title, author, ISBN or tags.

A subfacet of the "by Author" facet is Author "by Nationality". In addition to the compound tag *australian author*, this subfacet consists of those tags that indicate a nationality that obviously does not refer to the subject matter of the book. An example of this is found in the tag *british*, which was used nine times to represent the book *The History of the Decline and Fall of the Roman Empire* by the British author Edward Gibbon.

The "by Publisher" facet includes tags representing various publishers. There were eight different publishers identified in the dataset; tags representing the Folio Society, a publisher of fine books and collectibles, account for over half of the occurrences and instances of the tag in this facet.

---

[41] *The Voyage of the Beagle* by Charles Darwin is the only book in the dataset that was not tagged *read*.

In the "by User" facet, we find the tag *book club*, as well as tags in the subfacet "by Age group", like *adult*, *children's non-fiction*, *teen* and *ya* (young adult). What is interesting here is that, while the tags in the subfacet "by Age group" represent user groups for whom the tagger feels the book is intended, the tag *book club* presumably represents a group of which the tagger him- or herself is a member. In this way, the user is explicitly represented in the classification, albeit indirectly.

Like the "by User" facet of Physical Object, the user is also represented implicitly in the "by User" facet of Work, by virtue of the task-oriented tags in the "by Activity" facet.

4.2.3.5 Space facets of Work

There are no tags in the dataset that are here interpreted as belonging to a Space facet of Work. Examples of these types of tags would be *Spain* and *London*, where it is clear that what was meant with the tag designated where the work was, for example, *written* or *published*. It cannot be precluded that some of the tags in the Space category in the universe of subjects actually belong here.

4.2.3.6 Time facets of Work

The Time facet of Work ("by Year") includes tags that are here interpreted as indicating either the year the work was read by the user or the year the work was published. These make up 0.36% of the total tags in the dataset. As noted earlier, another possibility is that the tags indicate the year the book as a Physical Object was purchased, but they are not interpreted as such in this research. It was deemed impossible to determine the exact intention with each tag in the "by Year" facet. For example, while the tag *100s* representing the book *The Twelve Caesars* clearly indicates the time the work was written, it is impossible to tell whether the tags *2001* and *2006*, both representing the book *The Map that Changed the World*, indicate years the users read the work or the years the particular edition the user has was published.

The abovementioned bibliographic skew slants the interpretation options slightly in favor of tags indicating "by Year (read)". Year of publication is already published along with the bibliographic information that is automatically given on each book's page, decreasing the need for each user to provide that information through tagging.

4.2.3.7 External Reception facets of Work

0.34% of the tags in the folksonomy have been placed in various facets that collectively belong to a category entitled External Reception. This is a fairly problematic category. Here, we find tags in facets that have been interpreted to indicate where the book has been mentioned ("by Source"), awards the book has won ("by Award"), how the book has been rated ("by Rating") and new expressions of the book ("by New Expression").

0.15% of the analyzed tags belong to the "by Source" facet: *Comedy Central*, *daily show*, *npr*, *sa*, *shelf awareness* and *This American Life*. It cannot be precluded that the user is indicating some other type of relationship to the work with the use of these tags. For example, the tags *Comedy Central* and *daily show*, both used on the book *America (The Book): A Citizen's Guide to Democracy Inaction*, may have been chosen because the user associates them with the author, Jon Stewart. The "by Award" facet accounts for 0.11% of the tags and is, by comparison, fairly unproblematic with straightforward tags like *Pulitzer Prize*, *National Book Award* and *nobel prize*.

Only 0.05% of all the tags indicate a type of rating. These include the tags *favorite*, *favorites*, *recommended* and *staff pick*. The paucity of tags in this facet was initially surprising. It can, however, be explained by the fact LibraryThing has a popular feature in which users can rate their books with 1-5 stars, and users are able to browse their collections by the number of stars each book has received by them.

Finally, we also find a facet in the External Reception category in which tags indicate new expressions of the works: "by New Expression" (0.05%). Here are the tags *film* and *movie*. New expressions actually represent items with associated relationships to the works, and as such, would not traditionally belong to the universe being classified. In this model, they are included as a facet of the Work in the External Reception category to indicate that associations that users have between the book and various remakes of the book.

4.3 The Universe of Subjects

The universe of subjects exists as a Personality facet of Work in the Universe of Books: "by Subject". Tags representing the universe of subjects account for slightly over 70.54%; results of the facet analysis of this universe are delineated below.

4.3.1 Basic facets: Disciplines or Concepts?

As has been discussed, there are at least two fundamentally different ways to approach a universal faceted classification of the universe of subjects. The first is the more traditional way, represented by the Colon Classification and the 2nd Edition of the Bliss Bibliographic Classification, in which the universe of subjects is initially divided into classes based on disciplines. Here, basic facets correspond to disciplines (and major subdisciplines), and the remaining facets are facets of these disciplines. The second approach to universal faceted classification is represented by the CRG's work on a New General Classification in the 1960s and 1970s and by Gnoli etal's current work on the Integrative Levels Classification Project. In this approach, there is no initial division of the universe of subjects into disciplines, but rather into classes based on the integrative levels of phenomena. Here, each phenomenon is considered in its own right and is synthesizable with other phenomena in the classification scheme. Facets are thus considered as "the relations typical of a phenomenon with other phenomena" (Gnoli, 2006). In this light, basic facet is synonymous with the phenomena considered in each compound subject. In order to identify basic facets in the universe of subjects, it was necessary to examine which approach best reflects the implicit classification represented by the analyzed folksonomy.

4.3.2 Disciplines

There was one determining factor that suggested that the representation of the universe of subjects in the folksonomy was based on an initial division into disciplines: 37.35% of all the tags in the analyzed dataset indicate a discipline[42]. Although this percentage is likely distorted by the fact that the initial selection of tagged books was made using the tags *history* and *non-fiction*, a cursory facet analysis of the top 20 non-fiction books[43] in LibraryThing reveals that disciplines still account for over 26% of the total tags. This is consistent with the number of tags indicating disciplines within the list of most related tags[44] to *non-fiction*; 11 of 40 tags (27%) indicate a discipline[45]. A closer examination of the tags reveals that the use of disciplines to represent basic

---

[42] Tags indicating discipline account for slightly over half of the tags representing the Universe of Subjects.

[43] http://www.librarything.com/tag/non-fiction (Retrieved November 10, 2009)

[44] Ibid.

[45] As expected, this percentage is lower when fiction books are included; only 10 of the top 75 tags (13.33%) in the entire LibraryThing folksonomy indicate a discipline, and only 5 of the 40 tags (12.5%) related to *fiction* indicate a discipline, 3 of which belong to the discipline *literature*. http://www.librarything.com/zeitgeist and http://www.librarything.com/tag/fiction (Retrieved November, 20, 2009).

facets (and thus implicitly, the main subdivision of the universe of subjects) is far more problematic than first assumed.
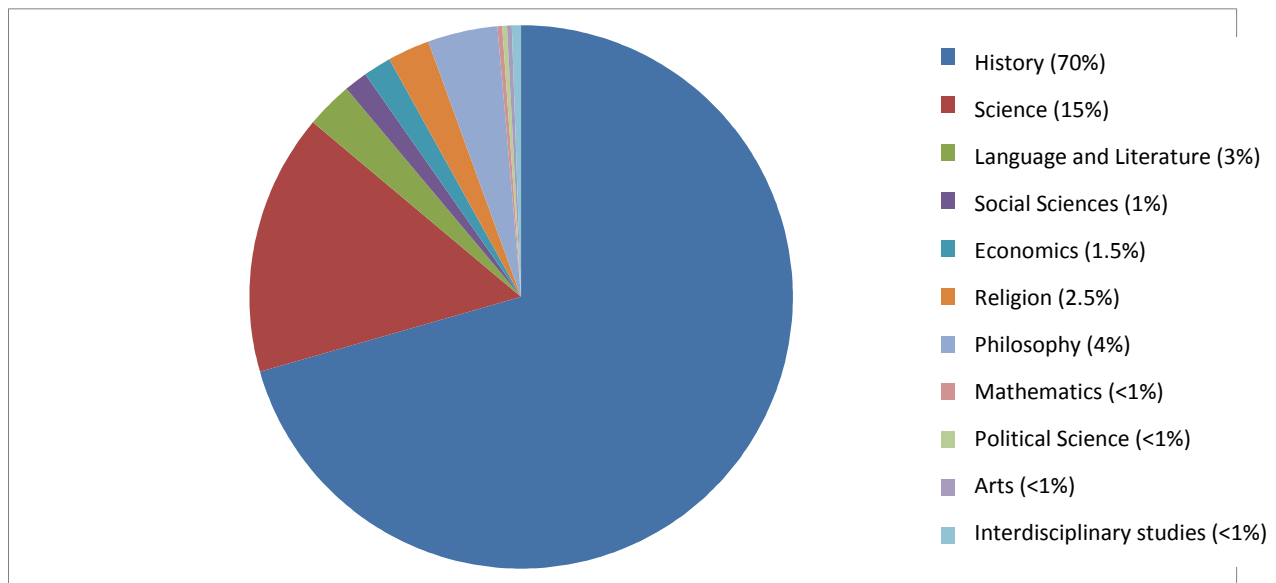

Figure 11: Distribution of tags indicating disciplines in the dataset

Figure 11 shows the distribution of the 11 disciplines identified in the dataset. Tags indicating the discipline History account for 70% of all the discipline tags. While the tags *history* (21287 instances), *900*[46] (3) and *historiography* (52) make up 75.5% of all the tags indicating history, the remaining tags represent a wide variety of facets of history. Here, we find compound tags indicating history "by facet", like:

> History "by Time": *14th century history*, *18th century history*, *20th century history*, *ancient history, classical history*, *colonial history*, *early modern history*, *medieval history*, *modern history*, *pre-columbian history*, *pre-contemporary history*

> History "by Place": *african history*, *american history*, *asian history*, *australian history*, *belgian history*, *british history*, *california history*, *chicago history*, *chinese history*, *commodity history*, *english history*, *european history*, *french history*, *german history*, *history—us*, *israeli history*, *japanese history*, *london history*, *maritime history*, *middle*

---

[46] 900 is the class number for history, geography and biography in the Dewey Decimal System.

eastern history, *russian history*, *spanish history*, *texas history*, *us history*, *western history*, *world history*

History "by Time" and "by Place": *ancient greek history*, *ancient roman history*

History "by Energy": *history—wwii*, *history of reading*

History "by Group" or "by Person": *family history*, *jewish history*, *indian history*, *military history*, *native american history*, *naval history*, *royal history*, *tudor history*, *women's history*

History "by Entity": *book history*, *church history*, *commodity history, culinary history*, *cultural history¸ food history*, *history of ideas*, *history of life*, *history of medicine*, *history of sexuality*, *intellectual history*, *library history, medical history*, *social history*, *urban history*

Additionally, we find the following tags indicating phase relationships between history and another discipline: *history of science*, *history of technology*, *religious history*, *historical linguistics*, *language history*, *literary history*, *political history*, *history of english*, *history of geology*, *history of philosophy* and *history of biology*. Phase relationships are defined by Ranganathan as being the "the assembling together of two or more Subjects," thus resulting in a Complex Subject (Ranganathan, 1967 p.358). He writes that there are at least five different types of phase relationships between two or more subjects: General Relation, Bias, Comparison, Difference and Influencing. Although the above compounds most likely illustrate phase relationships indicating the influence of one discipline on another[47], it is difficult to determine this with any certainty.

The uncertainty regarding the types of relationships indicated by the above-listed compound tags is exacerbated by the remaining 30% of tags in the basic facet that indicate disciplines other than History: Science (15%), Philosophy (4%), Language and literature (3%), Religion (2.5%), Economics (1.5%), Social Sciences (1%), Mathematics (<1%), Political Science (<1%), Arts

---

[47] Influencing phase relationships between disciplines indicate that one of the disciplines is seen from the perspective of the other. As such, the phase relationship in *history of science* indicates that the core discipline (science) is influenced by history.

(<1%) and Interdisciplinary Studies (<1%). As discussed earlier, we were unable to expose intra-facet relationships in the folksonomy using facet analysis techniques. It was thus impossible to tell how each discipline tag related to the other discipline tags.

Given the high occurrence of history tags in the basic facet (70%), it is tempting to propose that history is the core discipline and that the other disciplines are presented in terms of their relationships to history. If this indeed were the case, the remaining tags representing the universe of subjects in the dataset would all belong to facets of history. Unfortunately, it appears to be impossible to determine this with any certainty through the process of facet analysis. Indeed, the influencing phase relationships suggested by the compound tags above suggest otherwise. Furthermore, some of the books are tagged with a higher percentage of other disciplines than History. *History of Western Philosophy*, for example, is represented with 70% Philosophy tags and 30% History tags within the basic facet.

The sheer abundance of different disciplines represented in some of the books brings to mind the León manifesto. The León manifesto is associated with the Integrative Levels Classification Project discussed in section 2.2.2.2, which has as its primary goal to develop a universal faceted classification scheme that is non-disciplinary based. The León manifesto states that "the current trend towards an increasing interdisciplinarity of knowledge calls for essentially new knowledge organization systems (KOS), based on a substantive revision of the principles underlying the traditional discipline-based KOS" (Gnoli & Szostak, 2007). In *Guns, Germs and Steel*, for example, the following tags indicate discipline:

> History (46.85%): *archaeology*, *history*, *social history*, *world history*
>
> Science: (39.42%): *anthropology*, *biology*, *ecology*, *ethnology*, *geography*, *natural history*, *popular science*, *science*
>
> Social Sciences (8.46%): *social science*, *sociology*
>
> Economics (4.75%): *agriculture*, *economics*, *technology*
>
> Interdisciplinary Studies (0.53%): *cultural studies*

Here, it is clearly very difficult to interpret the precise nature of the relationships between the disciplines and between the subdisciplines within each discipline. It is conceivable that different users use the same tags to indicate different relationships between the disciplines. It would appear

as though users interpret the book as being so interdisciplinary that its placement in a discipline-based classification is impossible.

Finally, there are many examples of books in the dataset where *history* accounts for all or a vast majority of the discipline tags. In these cases, it would seem clear that the basic facet of the book is History and that the other facets are thus facets of History. It is impossible, however, to determine whether users intend for all the other subject tags in these cases to belong to facets of History and not a smaller or an unnamed discipline, if indeed to any discipline at all. In *The Meaning of Everything*, for example, tags indicating History account for 79% of the discipline tags; the rest indicate Language and Literature. It would appear, however, that over 90% of the tags[48] in the Personality category representing the book would presumably belong to facets of disciplines within Language and Literature.

Due to all of the abovementioned difficulties, it was deemed impossible to determine precisely what the identified facets in the universe of subjects were facets of. The four fundamental categories (PEST) identified in the dataset, however, were distinguishable as being applicable throughout the entire universe. In the next four sections, tags belonging to facets of the four fundamental categories will be delineated. They will be presented in reverse order[49] (TSEP), reflecting the degree of facility in which they were identified and further analyzed. For the sake of simplicity, I will hereafter refer to them as facets of Subject, subject being a vague descriptor for a specific but undetermined discipline. Discipline is understood to have been problematically chosen as basic facet due solely to the preponderance of tags indicating discipline in the dataset. The discipline/concept dichotomy will be discussed throughout the following sections to illustrate the difficulties encountered in identifying facets.

4.3.2.1 Time facets of Subjects

3.3% of the tags representing the universe of subjects belong to the "by Time" facet of Subject. This accounts for 2.37% of the total tags in the dataset. It was relatively easy to distinguish which tags belonged to this category. This is consistent with Ranganathan's observation:

---

[48] These tags include *dictionary*, *language*, *lexicographers*, *oed* (Oxford English Dictionary) and *words*. In the Colon Classification, for example, these would all be classified within the discipline Linguistics.

[49] The order will thus be from most abstract category to most concrete category.

"Perhaps the fundamental category 'Time' gives the least difficulty in its identification" (Ranganathan, 1967, p.399).

In the dataset, we find tags belonging to the facet "by Time" within the following two subfacets:

"by Period name":

> *ancient*, *antiquity*, *byzantine*, *classical*, *colonial*, *colonial period*, *dark ages*, *depression era*, *early modern*, *early republic*, *edwardian*, *elizabethan*, *late antiquity*, *medieval*, *middle ages*, *modern*, *pre-columbian*, *prehistory*, *regency*, *renaissance*, *victorian*, *victorian era*

"by Years":

> *5th century BC*, *12th century*, *14th century*, *1421*, *15th century*, *16th century*, *17th century*, *1774-1793*, *1776*, *18th century*, *1893*, *19th century*, *1900*, *1906*, *1918*, *1930s*, *1970s*, *1990s*, *20th century*, *21st century*

Many of the tags in the subfacets of Time manifest themselves as compound tags:

"by Period name" and "by Place":

> *ancient greece*, *ancient rome*, *classical greece*, *colonial america*, *england-regency*, *english renaissance*, *imperial russia*, *medieval europe*, *victorian england*, *victorian london*

"by Years" and "by Place":

> *england 19th century*

Tags in the subfacet "by Period name" account for slightly more of the "by Time" facet (61%) than do tags in the subfacet "by Years".

4.3.2.2 Space facets of Subjects
Subject "by Place" in the Space category was another easy facet to distinguish. It comprises 14.94% of the tags representing the universe of subjects and 10.53% of the total tags. Here, we find the subfacets:

"by Geopolitical unit":

> *africa*, *america*, *american south*, *american west*, *americas*, *ancient greece*, *argentina*, *aquitaine*, *asia*, *athens*, *atlantic*, *australia*, *australië*, *austria-hungary*, *barcelona*, *beirut*,

belgian congo, belgien, boston, brasil, britain, british isles¸california¸catalonia, chicago¸chile¸china¸congo¸ecuador¸egypt, eire, england, europe, france, galapagos¸galveston, georgia, germany¸ gloucester, great britain, greece, illinois, indonesia, ireland, israel, italy, iwo jima, japan, java, kansas, kongo, krakatoa, latin america, lebanon, londen, london, maine, malawi, maryland, massachusetts, mesoamerica, middle east, midwest, mogadishu, mozambique, nanjing, nanking, new england, new orleans, new south wales, newfoundland, norfolk island, normandy, north america, north carolina, oceania, oklahoma, oxford, pacific, palestine, paris, pennsylvania, persia, plymouth, rennes-le-château, rhodesia, rome, russia, san francisco, san fransisco, savannah, sierra leone, somalia, somolia, spain, south africa, south america, south carolina, southeast asia, sparta, sydney, tasmania, texas, the americas, tx, u.s., uk, united states, us, usa, versailles, virginia, world, wounded knee, zaire, zambia, zimbabwe

The following tags indicating Place "by Geopolitical unit" in adjective form were also placed here:

african, american, american english, american indian, ancient greek, asian, asian american, australian, british, celtic, chinese, english, european, french, german, greek, indian, indonesian, international, irish, japanese, latin, native american, roman, russian

This is acknowledged to be a problematic solution; it is likely that some of the adjectives indicate ethnicity or language. Their detachment from the noun or pronoun they are intended to describe makes it difficult, however, to ascertain their correct placement. The difficulty of adjectival tags will be discussed in more detail in section 4.4.3.

Other subfacets of Place include:

"by Function":

bibliothek, cod fisheries, college, fisheries, libraries, library school, monastery, monastic libraries, museums, penal colony, school

"by Physiographic features":

ocean, san andreas fault, sea, volcanoes

"by Orientation":

longitude, midwest, southern, the south, the west, west, western

"by Density of population":

cities, *city*, *settlement*

4.3.2.3 Energy facets of Subjects

That facets belonging to the Space and Time categories should be easily identifiable is hardly surprising[50]. Ranganathan points out that their respective identifications are "in accordance with what we commonly understand by [the terms Space and Time]" (Ranganathan, 1967, p. 399). On the other hand, he writes, "the identification of the fundamental category 'Energy' is a little more difficult than that of 'Space' or 'Time'" (ibid, p. 400). This is consistent with the observations made in this research.

Tags interpreted as belonging to Energy facets of Subject account for 12.72% of the tags representing the universe of subjects and 8.98% of the total dataset. These include tags indicating "action of one kind or another. The action may be among and by all kinds of entities—inanimate, animate, conceptual, intellectual, and intuitive" (ibid, p. 400). Here are included tags that are verbs, as well as nouns representing events, processes and actions.

I was unable to determine specific facets in the Energy category, but have created cursory groupings of the tags based on disciplines and concepts with which they may be associated. It is important to emphasize that they are not facets. They are somewhat problematic groupings, as will be discussed below.

Policy:

> *american civil war*, *american revolution*, *american revolutionary war*, *american politics*, *apartheid*, *assassination*, *arab-israeli conflict*, *atrocities*, *battle of the bulge*, *battles and campaigns*, *civil war*, *colonization*, *combat*, *conspiracy*, *crusades*, *d-day*, *depression*, *espionage*, *european theater*, *french revolution*, *genocide*, *globalization*, *holocaust*, *hundred years war*, *indian wars*, *israel-arab conflicts*, *king philip's war*, *massacre*, *mideast politics*¸ *modern warfare*, *pacific campaign*, *pacific theatre*, *peloponnesian war*, *persian wars*, *politics*, *race relations*, *rape of nanking*, *revolution*, *revolutionary war*, *russian civil war*, *russian revolution*, *segregation*, *slavery*, *spanish civil war*, *terrorism*, *torture*, *war*, *war crimes*, *warfare*, *wars*, *wwi*, *wwii*

Meteorology:

---

[50] Indeed, it has been speculated that these are the only two categories that are common to all subjects (Spiteri, 1998) and the Method of Residues is based on their easy identification.

disaster, drought, droughts, dust bowl, dustbowl, earthquakes, eruption, flood, galveston hurricane, hurricanes, natural disasters, plate techtonics, plate tectonics, san francisco earthquake, storm, storms¸ tsunami, weather, weather service

Mathematical geography:

cartography, map collecting, map theft, map thefts, mapmaking, mapping

Education:

american education, education, enlightenment, intelligence testing, IQ testing, learning, Learning and scholarship, pedagogy, teaching, testing

People/Society:

adultery, alcoholism, childhood, coming of age, crime, growing up, labor, murder, rape, rehabilitation, sanitation, scandal, sex, survival, trade, transportation, urban planning, violence

Disease[51]:

1918 influenza, black death, cholera, disease, epidemic, infectious diseases, influenza, pandemic, plague, spanish flu

Ecology:

culture diffusion, evolution, human evolution, natural selection, science evolution, scientific revolution, social evolution

Books:

book collecting, bookbinding, books and reading, leitura, lesen, publishing, reading, shelving, storage, storing_books, writing

Travel:

circumnavigation, discovery, exploration, navigation, sailing, seafaring, shipwrecks, voyages

General:

---

[51] Ranganathan writes that facets associated with Morphology, Physiology, Disease, Ecology and Hygiene had been mistakenly placed within the Energy category for over thirty years. These should rather be regarded as properties and thus be placed in the Matter category (Ranganathan, 1967, pp.400-401). In the facet analysis performed in this research, I have interpreted the tags in the dataset that indicate aspects of disease and ecology as representing processes, or things that happen, in the same way that aspects of policy do. They are thus here interpreted as rightfully belonging to the Energy category.

change, *collecting*, *development*, *discourse*, *measurement*, *paradigm change*, *paradigm shift*, *quest*

Other:

code breaking, *computing*, *cooking*, *craniometry*, *criminal investigation*, *cryptanalysis*, *detective work*, *encryption*, *fishing*, *forensics*, *overfishing*, *photography*¸ *reenacting*, *reenactments*, *theft*

The first problem associated with the above groupings is that some of them are here interpreted to reflect facets of disciplines while others are interpreted to reflect facets of concepts. The first four groupings, for example, reflect existent facets of disciplines in the Colon Classification: Policy is a facet of History; Meteorology and Mathematical geography are both facets of Geography[52]; and Disease is a facet of Medicine[53]. Education could have been interpreted as a discipline itself (as it is in the Colon Classification and in the BC2), but is here seen as a facet belonging to the discipline Social Sciences. Likewise, it is likely that the tags associated with People/Society also belong to various facets of Social Sciences. Some of the tags associated with Ecology are likely facets of Biology, while others are likely facets of Social Sciences.

The tags in the grouping Books, on the other hand, all indicate various Energy facets of the concept Book, regardless from which discipline the concept Book is regarded. The same can be said of the concept Travel. Indeed, several of the groupings interpreted as reflecting discipline-based facets could equally validly be interpreted as being facets of concepts. In this light, the tags in the groupings reflecting, for example, Disease or Meteorology, would all be interpreted as belonging to various facets of the concept Disease or the concept Weather, regardless from which disciplines they were viewed.

Tags in the General grouping are here interpreted as being too general to be able to associate with any one discipline or concept. Tags in the Other grouping could have been be further categorized into either discipline-based or concept-based groupings, but due to the miniscule representation of tags in this group, a further categorization was not undertaken.

---

[52] Tags indicating *geography* have in this research been interpreted as subdisciplines of the Science discipline.
[53] Tags indicating *medicine* have in this research been interpreted as subdisciplines of the Science discipline.

Due to all of the abovementioned difficulties, further attempts to identify facets in the Energy category were abandoned.

4.3.2.4 Personality facets of Subjects

The same problems discovered in the facet analysis of tags in Energy facets of Subject were present in the analysis of tags in the Personality facets. Like the Energy category, the facet analysis of tags in this category is also incomplete.

16.16% of the tags representing the universe of subjects belong to facets of Subject in the Personality category. These account for 11.4% of the tags in the entire dataset. Ranganathan remarks that this category "presents the greatest difficulty. It is too elusive. It is ineffable" (Ranganathan, 1967, p. 401).

Here, we find the universal facets "by Person" and "by Group", which includes tags grouped together by single characteristics indicating either a person or a group. These facets are considered to be universal because they can be applied to any discipline.

"by Person":

> "by Name": *aaron burr, abraham lincoln, adams, alexander hamilton, alexandra, anne boleyn, anne of cleves, aquinas, aristotle, austen, benjamin franklin, berkeley, caesar, catherine of aragon, catherine parr, charles dickens, christ, clinton, columbus, conrad, crippen, daniel burnham, darwin, descartes, dickens, eleanor, eleanor of aquitaine, enguerrand de coucy, garfield, george washington, guglielmo marconi, hamilton, henry ii, henry viii, henry whitehead, hitler, ishmael beah, jack the ripper, james augustus henry, murray, james madison, james murray, jane seymour, jefferson, jesus, jesus christ, jim williams, john adams, john harrison, john snow, kant, katherine howard, king of france, king leopold, lincoln, locke, louis xv, louis xvi, machiavelli, marconi, marie antoinette, mary magdalene, mckinley, nicholas, nicholas ii, nietzsche, plato, rasputin, romanov, shakespeare, sickert, socrates, st. patrick, stanley, thomas jefferson, tsar nicholas, tsarina alexandra, walter sickert, washington, william bradford, william smith, zheng he*

> "by Occupation/Role": *architect, assassins, authors, boy soldier, bibliophile, child soldier, child soldiers, civil war reenactors, confederates, convicts, courtesans, exile, expats, explorers, fishermen, foodie, founders, founding fathers, inventors¸ lexicographers, mistresses, monarchs, monks, nerd, paratroopers, pilgrims, presidents,*

queen, *queens*, *reenactors*, *roman emperors*, *royal mistresses*, *royals*, *scientists*, *serial killer*, *serial murder*, *thieves*, *tsar*, *us presidents*

(other subfacets here include: "by Familial relations", "by Ethnicity", "by Age group" and "by Gender")

"by Group":

"By 'Willingness to use force'": *military*, *army*, *marine corps*, *navy*, *us army*, *usmc*, *101st airborne*, *airborne*, *delta force*, *easy company*, *infantry*, *knights templar*, *rangers*, *special forces*, *special operations*, *special ops*

"By Ancestral relations": *british royalty*, *european royalty*, *royalty*, *family*, *royal family*, *russian family*, *russian royal family*, *angevin*, *merovingian*, *ming dynasty*, *plantagenet*, *tudor*

(others: "By political organization")

The remaining tags were grouped together into a general "by Entity" facet. It was deemed too difficult to figure out precisely which disciplines the following groupings were facets of, if indeed they were intended to be facets of any particular discipline (see discussion of Energy facets in section 4.3.2.3). They are presented below in general groupings suggesting which disciplines or concepts they may differentiate when grouped into facets.

"by Entity":

Religion:

atheism, *catholicism*, *celtic christianity*, *christianity*, *islam*, *judaica*, *judaism*, *monasticism*, *monotheism*, *mysticism*, *spirituality*, *vodou*, *world religions*

Ideology/Movement:

social movements, *anarchism*, *communism*, *darwinism*, *eugenics*, *fascism*, *feminism*, *fundamentalism*, *imperialism*, *marxism*, *nazism*, *racism*, *radicalism*, *socialism*

Science:

animals, *big bang*, *biological determinism*, *bodies*, *fossils*, *germs*, *dna*, *environment*, *nature*, *strata*, *virus*, *viruses*, *species*, *quantum theory*, *coal*, *rubber*, *cod*, *codfish*, *fish*, *food*, *salt*, *drugs*

Travel:

75

beagle, *boats*, *discoveries*, *hms beagle*, *map*, *maps*, *mayflower*, *ships*,

Theories, either universal or belonging to specific disciplines:

> *bell curve*, *knowledge*, *paradigm*, *conspiracy theory*, *ideas*, *queer theory*, *social theory*, *theory*

Language and Literature:

> *book*, *books*, *rare books*, *bibliography*, *cookbook*, *recipes*, *dictionary*, *encyclopedia*, *bible*, *book of kells*, *britannica*, *brittanica*, *encyclopedia britannica*, *heart of darkness*, *oed*; *bookshelf*, *bookshelves*; *language*, *languages*, *americanisms*, *slang*, *grammar*, *words*

Other:

> *chronometers*, *clocks*; *ciphers*, *code*, *codes*; *ferris wheel*, *flag*; *god*, *grail*, *invention*, *inventions*, *computers*, *radio*, *television*; *culture*, *customs*, *chivalry*, *etiquette*, *manners*, *court life*, *social life*, *social life and customs*, *antiquities*, *americana*, *victoriana*; *collection*, *death*, *intelligence*, *iq*, *Literacy*, *design*, *Gender*, *Class*, *race*, *nobility*, *bibliophilia*, *Bibliomania*, *homosexuality*, *Sexuality*, *black hawk*

4.4   General discussions

The rest of this chapter is devoted to brief general discussions of the research performed here. In sections 4.4.1 and 4.4.2, I will discuss general observations about the facet analyses of tags representing the universe of books and the universe of subjects. Here, I will also discuss whether the ontological models of the results substantiate or violate the Canons of Classification discussed in sections 2.3.1-2.3.7. In section 4.4.3, I will discuss general problems associated with the facet analysis of user-generated metadata and potential shortcomings of the methodology.

4.4.1   General discussion of the Universe of Books

The facet analysis of the universe of books can be likened to a facet analysis of a very specific domain in a special faceted classification. As such, it was relatively easy to expose the underlying faceted classification ontology because there were only two aspects of the universe that were differentiated by facets: Books as Physical Objects and Books as Works. This distinction was seen most clearly in the facet analysis of tags in the Energy category. Tags like *borrowed*, *own* and *owned* were interpreted as clear indications of the user's relationship to the physical book, while tags like *read*, *half-finished* and *to_be_read* indicate the user's relationship to the work represented in the book.

The analysis revealed that descriptions of books as physical objects account for less than 2% of all of the tags in the dataset, slightly over half of which belong to the "by Activity" facet in the Energy category. This seems to be a fairly clear indication that LibraryThing users are most concerned with the Work aspect represented in books.

One of the most interesting results of the facet analysis of tags representing the universe of books was the identification of task-oriented tags and the successful placement of these in facets in the classification. Even though the tags in these facets account for less than 5% of the total tag set, they are important because they show that faceted classifications can be used to integrate the user-dimension into a classification of digital objects.

4.4.1.1 The Universe of Books and the Canons of Classification

It was shown that the postulational approach to facet analysis could be used to manually expose a faceted classification ontology representing the universe of books in the folksonomy chosen for this research. The question remains whether the induced faceted classification substantiates or violates Ranganathan's Canons of Classification. Observations pertaining to how the facet "by Subject" relates to the Canons of Classification will be discussed separately, in section 4.4.2.1.

The results of the facet analysis seem to corroborate both the Canon of Differentiation and the Canon of Concomitance. These canons state that each facet should "give rise at least to two classes or ranked isolates" (Ranganathan, 1967, p. 145) and that no two facets should "give rise to the same array of subjects or of isolate ideas" (ibid, p. 153). The fact that the induced faceted classification obeys these two canons is a direct result of the way the methodology was formed in this research. The very definition of facets in the ontological model used in this research is that they differentiate classes in the universe. The methodology followed here demanded that tags be grouped together to identify facets that did precisely this: differentiate classes in the universe. Concurrently, the bottom-up identification of facets based on groupings of differentiating tags diminished the likelihood that two different facets would be identified based on the same differentiating characteristics.

The interpretation of the Canon of Exclusiveness discussed above is likewise substantiated. Facets were identified by grouping together tags that could be interpreted as being the result of

the application of a single principle of division. This automatically fulfills the Canon of Exclusiveness. Furthermore, based on an extension of the postulate of isolate facets, the methodology demanded that tags be placed in one and only one facet. Therefore, compound tags were split up and placed in just one facet. In this way, although facets existed for both aspects of the compound tag, the three instances of the tag *children's nonfiction* in the dataset were enumerated in the subfacet "by Age group" in the "by User" facet of Work and not in the "by Truth" subfacet of "by Genre".

If, however, one were to interpret the Canon of Mutual Exclusivity as pertaining to the relationships between the tags in each facet, i.e. that no entity can be described by more than one focus in each facet, then the induced faceted classification appears to be in violation of the canon. This becomes clear in several facets. The "by Elicited feeling" subfacet of Genre, for example, includes tags like *thriller* and *funny*, which are by no means mutually exclusive. Likewise, the subfacet "by Plot" of the same facet includes tags like *biography*, *political* and *travel writing*. It is possible that these problems could be resolved by continuing the facet analysis until all of the tags belonged to mutually exclusive subfacets, but it is questionable whether this is necessary. Furthermore, there are several examples of violation of the Canon in facets that likely cannot be further analyzed. The facet "by Award", for example, includes types of awards each work may have received, like *pulitzer prize* and *national book award*. The facet "by Source" includes where the users may have heard of the book and includes tags like *shelf awareness*, *this American life* and *the daily show*. It is neither inconceivable that a book receives more than one prize nor that it is mentioned in multiple arenas.

If we interpret the Canon of Exhaustiveness as meaning that a facet is exhausted when every conceivable focus within it is presented, then this canon is obviously violated by the induced faceted classification. Here, both the identification of facets and the composition of facets are dependent on the existent tags in the dataset. Tags indicating author, publisher and time of publication, for example, were added for some books but not for all of them. Thus, these facets were not exhausted. If, however, the canon is interpreted to mean that the classification should be hospitable enough to allow for the interpolation and extrapolation of new facets in each class, then the induced ontology presumably substantiates the canon. This was seen in the various examples of potential facets in the ontology; see, for example, section 4.2.2.7.

A faceted classification based on user-generated metadata appears to substantiate the Canon of Relevance, but this is a highly problematic substantiation. According to the Canon of Relevance, facets should "be relevant to the purpose of the classification" (ibid, p.146). Although an analysis of users' motivations for tagging is beyond the scope of this research, it can be assumed that the purposes of tagging are multifarious. This is consistent with the "short answer" to what tags are, presented at LibraryThing:

> "Tags are a simple way to categorize books according to how you think of them, not how some library official does. Anything can be a tag—just type words or phrases, separated by commas. Thus one person will tag the The DaVinci Code 'novels' while another tags it 'trashy, religion, mary,' and still another only 'summer home.' Tags are particularly useful for searching and sorting—when you need a list of all your novels or all the books at the summer home" (LibraryThing, 2009).

In this sense, the purpose of tagging is that each individual user chooses tags that are most relevant to him or her. It has been suggested that the aggregation of user tags reveals a consensus of meaning about tagged resources, illustrated in the idea of emergent semantics (Marchetti et al, 2007). While this seems likely to be the case with tags indicating aboutness and other ascertainable aspects of tagged resources, it becomes very problematic when considering task-oriented tags, like those found in facets indicating user activities and feelings. Indeed, it seems unlikely that the aggregation of user-relevance found in facets containing task-oriented tags results in any form of social relevance. Thus, while the aggregation may be able to give us an indication of which facets are most relevant on the user-level, it comes at the cost to relevance at the social-level.

The incongruity in user- and social relevance manifests itself in violations of both the Canon of Ascertainability and the Canon of Permanence. Although the subfacet of Genre "by Elicited feeling", for example, appears to be a relevant facet for individual users, it is very difficult to ascertain whether a book actually is *funny* or a *thriller*. Likewise, although the facet "by Activity" seems to be relevant for many individual users in order to keep track of which books they have read, *unread* and *half-finished* are most likely highly impermanent.

4.4.2 General discussion of the Universe of Subjects

The universe of subjects is the very object of universal library faceted classifications. Unsurprisingly, many of the difficulties that arose during the course of the facet analysis of tags representing the universe of subjects were consistent with theoretical challenges that have been discussed in great detail over the past decades concerning potential structures of universal faceted classifications. As we have seen, there are at least two different ways to model a universal faceted classification: one can either divide the universe of subjects into smaller units of knowledge, such as disciplines and such as has been done traditionally, or one can base it on the concepts in the universe and divide it by some other means, as was attempted in the CRG's quest for a New General Classification and as is being explored in the Integrative Levels Classification Project.

The biggest problem associated with exposing a faceted classification ontology from the tags representing the universe of subjects was that it was problematic to determine precisely what each facet was differentiating. While it was possible to identify implicit basic facets in the universe of books, the universe of subjects is too big and it was therefore necessary to attempt to model the ontology on preexisting faceted classification models. Due to the high occurrence of tags indicating disciplines, it was assumed that users acknowledge disciplines to be a natural initial division of the universe of subjects. Intra-facet relationships, however, are not exposed during the course of facet analysis and it was therefore neither possible to reveal what types of phase relationships the disciplines had with one another nor to which disciplines the remaining tags indicating subjects differentiated. As was shown, some of the groupings seemed clearly to include facets of disciplines, while others seemed to include facets of concepts. Further attempts to expose facets were discontinued upon the realization that they could not be affixed with any certainty to classes in the universe of subjects.

Four universal facets were identified: "by Time", "by Space", "by Person" and "by Group". Like conceptual categories, these facets appear to be applicable to every discipline. The remaining tags were identified as belonging either to the Energy category or to the Personality category, but they were not placed in facets. It can therefore be said that the facet analysis of the tags representing the universe of subjects resulted in a faceted classification ontology based on conceptual

categories rather than on facets. In this way, the ontology lacks the expressiveness it would have had with the identification of facets.

4.4.2.1 The Universe of Subjects and the Canons of Classification

In this section, the conceptual categories (PEST) identified in the universe of subjects will be treated like facets of the entire universe in order to be able to discuss whether they substantiate the Canons of Classification. Unfortunately, this entails ignoring the entire basic facet "by Discipline", which makes up over half of the tags representing the universe of subjects.

The incipient results of the facet analysis of the tags representing the universe of subjects indicate that the induced faceted classification substantiates most of the Canons of Classification. The validations of the Canon of Differentiation and the Canon of Concomitance were both seen as entirely unproblematic: each category contained at least two tags and no two categories gave rise to the same array of tags.

The abovementioned discussions concerning the various interpretations of the Canon of Exhaustiveness and the Canon of Exclusiveness can also be applied to the results of the categories in the universe of subjects. It could be argued that the sample of tags representing just 76 books is far too small a sample to achieve exhaustiveness in an induced ontology; Like Spiteri (1998), I am unsure whether the Canon of Exhaustiveness ever can be substantiated in a growing universe. There were many tags in the dataset representing compound subjects from the category Time and the category Space, like *medieval europe* and *ancient greece*. These were all enumerated in the Time category in compliance with the Canon of Exclusiveness.

Overall, however, the Canon of Exclusiveness was violated by at least the Energy category and the Personality category in the universe of subjects. This is a natural result of the fact that there was no precondition that the categories be composed as a result of the application of a single principle of division; indeed, it is assumed that the categories contain many different facets.

There were no apparent conflicts between the Canon of Relevance and the Canons of Ascertainability and Permanence in the universe of subjects. The latter two were both substantiated by the induced faceted classification ontology: all of the tags in the four categories identified in the dataset were considered both ascertainable and permanent. This is likely due to

the fact that none of the tags indicating subject were interpreted as being task-oriented tags; the purpose of tagging the subject matter of a book is likely less multifarious than the purposes of tagging books in general.

4.4.3   Discussion of potential problems

In this section, I will discuss potential problems with this research. First, I will look at the inherent vulnerability related to the interpretation of user-generated metadata. Secondly, I will discuss potential shortcomings of the methodology.

4.4.3.1 Inherent vulnerability of data source

The research performed in this thesis is inherently speculative; it is based on my interpretations of the intended meanings in tags that have been disconnected from their user-contexts. It is acknowledged that absolute certainty of meaning can never be obtained by the methods used here, so compensatory strategies were sought.

I have tried to compensate for the vulnerability involved in interpreting the tags with a number of strategies. First of all, I have preserved the initial resource context of each tag; i.e. I have been able to double check to which book each tag was initially affixed in order to gleam context from knowledge about the book. Secondly, each tag in LibraryThing has its own "tag-page" where information about which books it represents and which users regularly uses it. Although individual user-context was not available in the dataset, it was thus possible to check obscure and ambiguous tags against the collections of users that used the tags regularly. In this way, context was obtained through the interpretation of how the tag was used with other tags by specific users. Thirdly, external sources like Google, Amazon and Wikipedia were regularly consulted in order to gauge the meaning of obscure tags. Finally, it has been assumed throughout this research that the more frequently a tag is used on a particular entity, the more likely the meaning of the tag is stable. This assumption is based on research confirming the Power Law distribution of tags (Halpin et al., 2007) and implies that tags used most frequently on an entity represent a consensus amongst users concerning the meaning of the entity. By focusing on the "short head" of the Power Law in the form of the most popular tags of the most popular books tagged with *history* and *non-fiction*, I have attempted to diminish the number of ambiguous, random and misleading tags in the dataset.

Despite the compensatory techniques mentioned above, it is acknowledged that some of the tags may have been misinterpreted during the course of the analysis. Tags in the universe of books that were interpreted as indicating years, for example, were grouped together to indicate year published. As mentioned, they may just as easily indicate year read or year purchased. It is nonetheless believed that eventual misinterpretations have not affected the overall resulting faceted classification model enough to be of any major consequence.

There were 34 instances of 13 tags that were deemed too ambiguous or too obscure for me to be able to place them in a facet:

> @, *hb*, *hc*, *Hillel*, *m*, *main*, *npl 2008*, *other*, *q*, *qh*, *sts*, *z*.

Many of these (*hb*, *hc*, *m*, *q*, *qh* and *z*) presumably indicate private category notations for users, but I was unable to interpret with any certainty what these symbols represented. *Hillel* could refer to a Jewish student campus or to a famous Jewish leader; it was difficult for me to interpret what it indicated in terms of the book to which it was affixed: *Sex with Kings: 500 Years of Adultery, Power, Rivalry, and Revenge*. The tag *other* could have been used in every facet in the induced faceted classification in order to substantiate the Canon of Exhaustivity, but was instead taken out of the dataset because its meaning was too difficult to ascertain.

Finally, it was found that adjectival tags in the dataset were very difficult to interpret. The interpretation of these for placements in facets is acknowledged as being problematic. In addition to the problematic adjectival tags indicating place of subject (discussed in section 4.3.2.2) or of author (discussed in section 4.2.3.4), and the adjectives interpreted here to indicate genre (section 4.2.3.2), the following adjectives were placed in the "by Entity" facet of Subject:

> *maritime*, *naval*, *native*, *revolutionary*, *gay*, *puritan*, *nazi*, *medical*, *molecular*, *queer*, *christian* and *jewish*

These make up 0.17% of the entire dataset and are assumed to characterize entities (like ideas, theories, etc) or people in the universe of subjects. Their exact placement, however, was too uncertain and it is debatable whether these too should have been removed from the dataset.

4.4.3.2 Shortcomings of the methodology

There are several potential shortcomings with the methodology used in this research. First of all, the initial algorithm applied to the tags rested on a number of assumptions that may not be valid. For example, the first step in the algorithm the assumption that each tag either did or did not indicate aboutness. Those that were interpreted as not being about the book's subject matter were automatically sorted together in the universe of books. This led to the somewhat problematic placement of tags like *film* and *movie* in the "by New Expression" facet of the book as Work. Although it seems unproblematic to place this facet in the conceptual category External reception, it is acknowledged that there may have been a better solution to exposing the relationship between the book, the work and the new expression. One of the consequences of the current placement is that it would be difficult to combine the induced faceted classification in this research with a faceted classification representing the universe of films, despite the presumption that they would have many facets in common.

Another problematic area in the methodology involved the identification and grouping of disciplines. Disciplines can be said to be a somewhat arbitrary organizational unit of knowledge; their boundaries are socially constructed in an attempt to divide areas of academic pursuits. As such, the boundaries and compositions of disciplines are interpreted differently by different actors. Even within traditional library classifications, the initial division of the universe of subjects into disciplines is different from classification to classification. My use of the BC2 to check boundaries was a somewhat arbitrary choice, made solely because the BC2 is one of the only universal faceted classifications in existence. It is highly conceivable that users had other types of library classification disciplines in mind when they tagged subject matter by discipline; for example that they were referring to disciplines as they are used by the Dewey Decimal Classification or by the Book Industry Standards and Communications (BISAC) subject headings, or that they were referring to disciplines as they remember them from school. In this light, the failed attempt to identify basic facets in disciplines to represent classes in the universe of subjects was problematic from the beginning.

The same problems can be applied to the arbitrary definition of genre used in this research. Here, too, it was very difficult to determine with any certainty that a tag was used to indicate "the style or perspective that an author uses or that a user interprets that the author uses, that influences the

composition or the reception of the work" (my definition, section 4.2.3.2). Furthermore, the facet analysis of the "by Genre" facet revealed nine distinct subfacets based on different aspects of the work on which the tag focused. The quantity and diversity of subfacets in the "by Genre" facet may indicate that users use a number of different strategies to describe the composition of a work, and it is possible that these could have been modeled differently. Additionally, several of the tags that were interpreted as indicating genre, like *historical*, *political* and *comedy*, are used as BISAC subject headings. This reinforces the inherent difficulty involved in dividing the universe of subjects and the universe of books (as works) into classes. More research on the use of genre in user-generated metadata is highly recommended.

Finally, it can be mentioned that, while tags representing 76 books seemed to be sufficient to induce an ontology of the finite universe of books in the tag space, it is likely that they were not sufficient to induce an ontology of the universe of subjects. Indeed, it could be argued that 76 books are really too miniscule a sample to analyze even just books that belonged to the History discipline, assuming it could be established that every tag belonged to facets of History. It is not inconceivable that, given a larger dataset and less time constraints the facet analysis of the universe of subjects could have given clearer results. The problem of identifying basic facets in the universe of subjects, however, would likely remain.

5   Conclusions and recommendations for further study

Weinberger (2006) remarked that, while classification systems are like trees, folksonomies are more like piles of leaves. This research has shown that facet analysis techniques can be used to rake tags into facets and to bag facets into conceptual categories. These have manifested themselves here as two induced faceted classification ontologies: one representing the universe of books and one representing the universe of subjects. The ontology representing the universe of books is considered to be complete. The ontology representing the universe of subjects, however, is incomplete. This was found primarily to be due to difficulties encountered in unearthing just *what* the facets in the universe of subjects were of.

Four research questions were posed at the beginning of this research. The first question sought to find out how the postulational approach to facet analysis could be applied to a folksonomy. To answer this, an in-depth study of Ranganathan's theories of faceted classification was undertaken, including an examination of how his postulates have been received and interpreted by the Classification Research Group. This was presented in chapter 2 and was used as the basis upon which to form the methodology presented in chapter 3.

The second question concerned challenges and problem areas encountered in a facet analysis of this type of data. There were two types of problems identified. All of the tags were removed from their original user-context, necessitating my own speculative interpretation of each tag. Implications of this were discussed in section 4.4.3.1. It was also deemed impossible to identify the intended relationships between tags representing disciplines in the universe of subjects and between the remaining tags indicating subject and these disciplines. This was not found to be a problem in the analysis of the universe of books, likely because the universe of books is a much smaller and confined universe; it was therefore easier to interpret which aspect of the universe each tag was indicating.

The identified facets, conceptual categories and relationships were delineated and characterized in depth in sections 4.2-4.3, answering the third research question. The fourth question was answered in sections 4.4.1.1 and 4.4.2.1, where the induced ontologies were examined in light of the Canons of Classification in order to discuss whether these were substantiated or violated. The main difference found between the two ontologies in terms of the Canons of Classification was in

their substantiation of the Canon of Relevance. The identification of facets comprised of task-oriented tags in the universe of books led to a problematic understanding of relevance, which in turn manifested itself in violations of the Canon of Ascertainability and the Canon of Permanence. No task-oriented tags were identified in the universe of subjects.

5.1 Implications of results

Despite the fact that the nature of this research was primarily theoretical, there are a number of practical implications of the results[54]. First of all, it has been shown that facet analysis techniques may be used for the identification and characterization of facets and conceptual categories in tags representing books. This indicates new knowledge: the most popular facets and categories are not the same as the most popular tags. In this light, it is possible that the "wisdom of the crowd" may be invoked to alleviate one of the major challenges involved in the creation of faceted structures, namely the selection of facets.

In 1960, Vickery noted that, "theoretically, an unlimited number of facets could represent the various perspectives contained in each topic" (Vickery, 1960, p. 20). Kwasnik (1999, p. 44) cites the difficulty of choosing the right facets as one of the leading problems in the creation of faceted classifications. The challenge of selecting facets is particularly relevant when constructing faceted classifications for websites, where usually only a select number of facets in a predefined order are displayed on the user interface. As the author of a recent user interface engineering article on designing for faceted search remarked: "Not all facets are created equal: some access points are more important than others depending on what users are doing and where they are in the site" (Lemieux, 2009). Unearthing which facets are most popular in folksonomies may provide valuable clues as to which facets are most important from an aggregated users' perspective.

The facets and categories discerned in this research indicate the following ranking in popularity:

1. by Discipline [basic facet of subjects] (ca 37%)
2. by Genre [facet of book as work] (ca 22%)

---

[54] There has been an implicit assumption that broad folksonomies represent the "wisdom of the crowd" throughout this research. Although the substantiation of this hypothesis is inconsequential for the results of the facet analysis, it is a determining factor for some of the practical implications of the results.

3. by facets in the Personality category [category of subjects] (ca 16%)
4. by Place [facet of subjects] (ca 15%)
5. by facets in the Energy category [category of subjects] (ca 13%)
6. by Activity (work) [facet of book as work] (ca 4%)
7. by Time [facet of subjects] (ca 3%)
8. by Activity (physical object) [facet of book as physical object] (ca 1%)

Even when taking into consideration the possible distortions of the results due to bibliographic information being readily available at LibraryThing and the disproportionate number of *history* and *non-fiction* tags in the dataset, it is quite clear here that facets indicating the universe of subjects are more popular than facets indicating the universe of books. Of the three facets listed above that indicate the universe of books, genre is clearly the most popular, followed by the two task-oriented facets indicating what users do to books as works and books as physical objects.

Conversely, the selection of facets for use on websites is often limited to the metadata available at the time of facet construction. The results of this research imply that facets could be culled from user-generated metadata, either in the form of tags or otherwise[55]. In an OPAC, this could perhaps manifest itself as facets indicating awards associated with books in the collection.

This research has also shown the possibility of including a user dimension in a faceted classification. This presents many possibilities for designers of faceted structures on websites to allow for user interaction without bothering the basic structure of the classification system. It implies, however, that steps be taken to take into account the overall purpose of the classification so that relevance for all users is preserved. One way this may be accomplished is through efforts to personalize the faceted classification, such that users are only presented with task-oriented facets that are relevant to them.

One can, for example, imagine its use in a faceted online public access catalog where a user could log on and mark objects in the catalog based on specified criteria, like whether or not the user has read or enjoyed them. Logged on users could then be presented with a facet on the search page indicating task-oriented tags in addition to all the other facets of documents normally presented there. This personalized version of the catalog would allow for the user to narrow search results

---

[55] LibraryThing, for example, also collects user-generated metadata about each book called "Common Knowledge," including information about the book's characters, epithets and awards. This information is available via an API. http://www.librarything.com/commonknowledge/ and http://www.librarything.com/api

by, for example, books that haven't yet been read by the user ("by User Activity") that are by such and such author ("by Author") and about such and such subject ("by Subject").

In the presentation of the results of the facet analysis of tags representing the universe of subjects, it was seen that inherent difficulties were encountered involving in the identification of what the facets were of. This implies that, in order to successfully expose a faceted classification ontology in a tag space in which some of the tags indicate aboutness, some prior form of initial division of the universe of subjects appears necessary. This is consistent with Schmitz' (2006) examination of Flickr tags, which found that the use of "domain-specific upper model ontologies" is necessary for the inducement of a faceted ontology in the tags.

There are several ways this may be accomplished. One is a tagging system that requires users to choose a core category in which to place each resource before tagging it. This requirement, however, seems counterintuitive to the freedom and ease of tagging. Furthermore, it would not answer to the contention made by the León manifesto that the distinction between disciplines is becoming less and less rigid (Gnoli & Szostak, 2007). Schmitz (2006) examines the automatic correlation of gazetteers and common taxonomies to tags.

5.2   Prospects for future research

Based on the results of this research, there are a number of interesting prospects for future research. First of all, the results of the facet analysis should be examined in light of the theory of "emergent semantics" in folksonomy research. This theory, briefly discussed in section 1.1.2, has been alluded to in several discussions of the results. Due to time constraints, an in-depth study of the implications of using facets composed of aggregate tags has been considered beyond the scope of this research. Of particular interest is the intersection between "emergent semantics", task-oriented tags and the concepts of user and social relevance.

More research is recommended based on the examination of different ways with which to initially divide the universe of subjects in order to facilitate the facet analysis of tags. In this light, the Open Shelves Classification (OSC) project at LibraryThing is interesting. The OSC aims to create top-level, statistically-tested classes in the universe of subjects through a bottom-up collaboration of LibraryThing users (Public Library Association, 2009). An analysis of the use of

disciplines in tags compared to the OSC top-level classes is recommended, in particular vis-á-vis differences in fiction and non-fiction books.

It would undoubtedly be interesting to explore the relationship between tags and genre in more depth. Nine subfacets of genre were found in this research. Can the facet analysis of user-generated tags help us to define genre from a sociocognitive perspective?

Fu et al. (2009) have developed a model for the prediction of tag choices based on a cognitive study of the imitation effect in tagging. It was suggestion here that LibraryThing folksonomies are largely comprised of "uninfluenced" tags, meaning that users choose tags removed from prior tags of the resource in question. Further research on the validity of this suggestion and an eventual comparison of folksonomies based on "imitated" tags and those based on "uninfluenced" tags would be very interesting.

Finally, the methodology applied manually to the dataset in this research was a highly laborious and time-consuming effort. Further research into automating this process is recommended. Interested readers are referred to Stoica et al. (2007) for their work on the automatic extraction of faceted hierarchical metadata in texts and to Marchetti et al. (2007) for their work on the extraction of tags expressing semantic relationships.

Bibliography

Abrera, J. (1974). Traditional Classification: Characteristics, Uses and Problems. *Drexel Library Quarterly*, 10 (4), 21-36.

Adkisson, H. (2003). *Use of Faceted Classification*. Retrieved March 24, 2009, from http://www.webdesignpractices.com/navigation/facets.html

Austin, D. (1969). Prospects for a New General Classificaion. *Journal of Librarianship*, 1 (3), 149-169.

Austin, D. (1972). The CRG research into a freely faceted scheme. In: Maltby, A (Ed.), *Classification in the 1970s: A second look* (pp.158-194). London: Clive Bingley Ltd.

Austin, D. (1984). *PRECIS. A manual of concept analysis and subject indexing. (2nd ed.)*. London: The British Library.

Beaudoin, J. (2007). Flickr image tagging: patterns made visible. *Bulletin of the American Society for Information Science and Technology,* 34 (1), 26-29.

Beghtol, C. (2005). The Facet Concept as a Universal Principle of Subdivision. In: Raghavan, K.S. & Prasad, K.N. (Eds.) *Knowledge Organization, Information Systems and Other Essays. Professor A. Neelameghan Festschrift* (pp. 53-70). New Delhi: Ess Ess Publications.

Beghtol, C. (2008). From the Universe of Knowledge to the Universe of Concepts: The Structural Revolution in Classification for Information Retrieval. *Axiomathes*, 18 (2), 131-144.

Broughton, V. & McIlwaine, I.C. (2000). The Classification Research Group—Then and Now. *Knowledge Organization*, 27 (4), 195-199.

Broughton, V. (2001). Faceted classification as a basis for knowledge organization in a digital environment; the Bliss Bibliographic Classification as a model for vocabulary management and the creation of multi-dimensional knowledge structures. *The New Review of Hypermedia and Multimedia*, 7 (1), 67-102.

Broughton, V. & Slavic, A. (2002-2003). *Towards a knowledge structure for high performance subject access and retrieval within managed digital collections.* Retrieved January 2, 2009, from http://www.ucl.ac.uk/fatks/about.htm

Broughton, V. (2006). The need for a faceted classification as the basis of all methods of

information retrieval. *Aslib Proceedings: New Information Perspectives*, 58 (1/2), 49-72.

Broughton, V. (2009). *Facet analysis as the theoretical basis of vocabulary tool construction: Fundamental Purposes in Knowledge Organization*. PowerPoint Lecture presented at the ISCO UK Conference, 2009. Retrieved November 22, 2009, from http://www.iskouk.org/conf2009/presentations/broughton_ISKOUK2009_presentation.pdf

Classification Research Group. (1955). The need for a faceted classification as the basis of all methods of information retrieval. *The Library Association record*, 57 (7), 262-268.

Dahlberg, I. (1992). Cognitive Paradigms in Knowledge Organization. *International Classification*, 19(3), 125 and 145.

Denton, W. (2003). *How to Make a Faceted Classification and Put It On the Web.* Retrieved November 3, 2008, from http://www.miskatonic.org/library/facet-web-howto.html

Dix, A., Levialdi, S. & Malizia, A. (2006). *Semantic Halo for Collaboration Tagging Systems*. Workshop on the Social Navigation and Community-Based Adaptation Technologies In Conjunction with Adaptive Hypermedia and Adaptive Web-Based Systems (AH'06), Dublin, Ireland. Retrieved November 2, 2009, from http://www.sis.pitt.edu/~paws/SNC_BAT06/crc/malizia.pdf

Ellis, D. & Vasconcelos, A. (1999). Ranganathan and the net: Using facet analysis to search and organize the World Wide Web. *Aslib Proceedings*, 51 (1), 3-10.

English, J., Hearst, M., Sinha, R., Swearingen, K., & Yee, K.-P. (2002). Hierarchical faceted metadata in site search interfaces. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*, pages 628-639. ACM Press.

Farradane, J. (1961). Fundamental Fallacies and New Needs in Classification. In: Foskett D.J., Palmer B.I. (Eds.), *The Sayers memorial volume* (pp. 136-150). London: The Library Association.

Foskett, D.J. (1961). Classification and integrative levels. In: Foskett D.J., Palmer B.I. (Eds.), *The Sayers memorial volume* (pp. 136-150). London: The Library Association.

Foskett, D.J. (1978). The theory of integrative levels and its relevance to the design of information systems. *Aslib Proceedings*, 30 (6), 202-208.

Foskett, D.J. (2003). Facet Analysis. In: Drake, M (Ed.), *Encyclopedia of Library and Information Science* (pp. 1063-1067). New York: Marcel Dekker.

Fu, W-T., Kannampallil, T. & Kang, R. (2009). A Semantic Imitation Model of Social Tag Choices. *Proceedings of the 2009 International Conference on Computational Science and Engineering*, 4, 66-73.

Giess, M.D., Wild, P.J. & McMahon, C.A. (2008). The generation of faceted classification schemes for use in the organisation of engineering design documents. *International Journal of Information Management*, 28 (5), pp. 379-390.

Gilchrist, A. (2003). Thesauri, taxonomies and ontologies—an etymological note. *Journal of Documentation*, 59 (1), 7-18.

Gnoli, C. (2005). BC2 classes for phenomena: An application of the theory of integrative levels. *The Bliss Classification Bulletin*, *No. 47, 2005*.

Gnoli, C. (2006). The meaning of facets in nondisciplinary classifications. In: Budin, G., Swertz, C. & Mitgutsch, K. (Eds.). *Knowledge organization for a global learning society: proceedings 9th international ISKO conference, 4-7 July 2006, Vienna,* pp.11-18. Ergon: Würzburg.

Gnoli, C. & Szostak, R. (2007). *The León manifesto*. Retrieved March 13, 2009, from http://www.iskoi.org/ilc/leon.htm

Gnoli, C. (2008). Categories and Facets in Integrative Levels. *Axiomathes: an international journal in ontology and cognitive systems*, 18 (2), 177-192.

Golder, S. & Huberman, B. (2005). *The Structure of Collaborative Tagging Systems*. Retrieved September 15, 2008, from http://arxiv.org/ftp/cs/papers/0508/0508082.pdf

Halpin, H., Robu, V. & Shepherd, H. (2007). The Complex Dynamics of Collaborative Tagging. In *WWW 2007, May 8-12, 2007, Banff, Alberta, Canada*. Retrieved November 10, 2009, from http://www2007.org/papers/paper635.pdf

Hammond, T., Hannay, T., Lund, B. & Scott, J. (2005). Social Bookmarking Tools (I): A General Review. *D-Lib Magazine,* 11(4). Retrieved August 3, 2009, from http://www.dlib.org/dlib/april05/hammond/04hammond.html

Hearst, M. (2006). Design recommendations for hierarchical faceted search interfaces. *ACM*

SIGIR Workshop on Faceted Search.

Hjørland, B. (2001). Towards a Theory of Aboutness, Subject, Topicality, Theme, Domain, Field, Content … and Relevance. *Journal of the American Society for Information Science and Technology,* 52 (9), 774-778.

Hong, M. (2006). *The Dandelion bibliography of facet analysis*. Retrieved September 3, 2009, from http://www.iskoi.org/ilc/dsic/

IDC. (2009 May). *As the Economy Contracts, the Digital Universe Expands* [Multimedia White Paper]. Retrieved November 20, 2009, from http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm

IFLA. (1998). *Functional Requirements for Bibliographic Records: Final Report—1998*. Retrieved June 2, 2009, from http://archive.ifla.org/VII/s13/frbr/frbr1.htm#3.2

Ingwersen, P. & Wormell, I. (1992). Ranganathan in the perspective of advanced information retrieval. *Libri: international library review*, 42 (3), 184-201.

Integrative Level Classification. (2004). Retrieved September 30, 2009, from http://www.iskoi.org/ilc/

Kalbach, J. (2007). *Designing Web Navigation*. Beijing: O'Reilly.

Kipp, M. & Campbell, D. (2006). Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. In *Proceedings of the American Society for Information Science and Technology*, Austin, Texas.

Kome, S. (2005). *Hierarchical Subject Relationships in Folksonomies*. Unpublished master's thesis, University of North Carolina, Chapel Hill.

Kwasnik, B. (1999). The Role of Classification in Knowledge Representation and Discovery. *Library Trends*, 48(1), 22-47.

La Barre, K. & Cochrane, P. A. (2005). Facet Analysis as a Knoweldge Management Tool on the Internet. In: Raghavan, K.S. & Prasad, K.N. (Eds.) *Knowledge Organization, Information Systems and Other Essays. Professor A. Neelameghan Festschrift* (pp. 53-70). New Delhi: Ess Ess Publications.

La Barre, K. (2006). The use of faceted analytic-sythetic theory as revealed in the practice of

website construction and design. PhD dissertation, Indiana University, United States. Retrieved September 30, 2008, from https://netfiles.uiuc.edu/klabarre/www/LaBarre_FAST.pdf

Lambiotte, R. & Ausloos, M. (2005). Collaborative tagging as a tripartite network. *arXive: Computer Science*. Retrieved October 30, 2009, from http://arxiv.org/PS_cache/cs/pdf/0512/0512090v2.pdf

Lemieux, S. (2009, March 1). Designing for faceted search. *KMWorld magazine*. Retrieved April 20, 2009, from http://www.kmworld.com/Articles/Editorial/Feature/Designing-for-faceted-search-52781.aspx

Lee, C. (2007). Taking Context Seriously: A Framework for Contextual Information in Digital Libraries. In *UNC SILS Technical Report 2007-04*.

Library of Congress. (2007, April). The Library's Digital Evolution: Librarian of Congress Testifies Before Congress. *Information Bulletin*. Retrieved September 2, 2009, from http://www.loc.gov/loc/lcib/0704/digital.html

LibraryThing. (2009). Retrieved November 29, 2009, from http://www.librarything.com

Louie, A. (2003, March). *Using Faceted Classification to Provide Structure for Information Architecture*. Poster presented at the American Society of Information Science & Technology (ASIS&T) Information Architecture Summit, Portland, Oregon. Retrieved October 20, 2009 from http://www.aaronlouie.com/Portfolio/Intellectual/FacetsIA/index.html

Maltby, A. (Ed.). (1975). *Sayers' Manual of Classification for Librarians (5th ed.)*. London: André Deutsch.

Marchetti, A., Tesconi, M., Ronzano, F., Rosella, M. & Minutoli, S. (2007). SemKey: A Semantic Collaborative Tagging System. In *WWW 2007, May 8-12, 2007, Banff, Alberta, Canada*. Retrieved November 10, 2009, from http://www.ibiblio.org/www_tagging/2007/paper_45.pdf

Mathes, A. (2004). *Folksonomies- Cooperative Classification and Communication Through Shared Metadata*. Retrieved January 12, 2009, from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

Mika, P. (2007). Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, 5 (1), 5-15.

Miksa, S.D. (2002). Pigeonholes and Punchcards: Identifying the Division Between Library Classification Research and Information Retrieval Research, 1952-1970. PhD Dissertation, Florida State University, United States. Retrieved March 15, 2009, from http://courses.unt.edu/smiksa/documents/Miksa_Dissertation_2002.pdf

Mills, J. (1960). *A Modern Outline of Library Classification*. London: Chapman & Hall.

Mills, J. (1972). Bibliographic Classification. In: Maltby, A (Ed.), Classification in the 1970s: A second look (pp.25-50). London: Clive Bingley Ltd.

Mills, J. & Broughton, V. (Eds.). (1977). *Bliss Bibliographic Classification (2nd ed.)*. (Introduction and Auxiliary Schedules). London: Butterworths.

Mills, J. (2004). Faceted Classification and logical division in information retrieval. *Library Trends.* Retrieved January 30, 2009, from http://www.thefreelibrary.com/_/print/PrintArticle.aspx?id=118236065

Morrison, P.J. (2008). Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing and Management: an International Journal,* 44 (4), 1562-1579.

Morville, P. & Rosenfield, L. (2007). *Information Architecture for the World Wide Web (3rd ed.)*. California: O'Reilly.

Public Library Association. (2009). New Classification System for Public Libraries? *PLA Blog: The official blog of the Public Library Association*. Retrieved November 4, 2009, from http://plablog.org/2009/01/new-classification-system-for-public-libraries.html

Quintarelli, E. (2005). *Folksonomies: Power to the People*. Paper presented at the ISKO Italy-UniMIB meeting, June 24, 2005, Milan (IT). Retrieved March 22, 2009, from http://www.iskoi.org/doc/folksonomies.htm

Quintarelli, E., Resmini, A. & Rosati, L. (2006). *Facetag: Integrating Bottom-up and Top-down Classification in a Social Tagging System*. Paper presented at the EuroIA Conference 2006, Berlin (DE).

Ranganathan, S.R. (1960). *Colon Classification (6th ed.)*. London: Asia Publishing House.

Ranganathan, S.R. (1961). Library Classification on the March. In: Foskett D.J., Palmer B.I.
     (Eds.), *The Sayers memorial volume* (pp. 136-150). London: The Library Association.

Ranganathan, S.R. (1962). *Elements of Library Classification*. Bombay: Asia Publishing House.

Ranganathan, S.R. (1965). *The Colon Classification*. In: Artandi S. (Ed.), *Rutgers Series on
     Systems for the Intellectual Organization of Information*. (Vol. IV). New Jersey: Graduate
     School of Library Service Rutgers.

Ranganathan, S.R. (1967). *Prolegomena to Library Classification (3rd ed.)*. Bombay: Asia
      Publishing House.

Ranganathan, S.R. (1997). Subject Heading and Facet Analysis. In: Neelameghan, S. (Ed.), *S. R.
      Ranganathan's Postulates and Normative Principles: Applications in Specialized
     Databases Design, Indexing and Retrieval* (pp. 127-146). Bangalore: Ess Ess
     Publications.

Rosenfield, L. (2005). *Folksonomies? How about Metadata Ecologies?* Retrieved September 3,
     2009, from http://www.louisrosenfeld.com/home/bloug_archive/000330.html

Petersen, E. (2006). Beneath the Metadata: Some Philosophical Problems with Folksonomies. *D-
     Lib Magazine*, 12(11). Retrieved September 3, 2009, from
     http://www.dlib.org/dlib/november06/peterson/11peterson.html

Rabourn, T. (2003, August 10). Faceted Movable Type. *Pixelcharmer*. Retrieved November 27,
     2009 from http://www.pixelcharmer.com/fieldnotes/2003/faceted-movable-type/

Redmond-Neal, A. (n.d.). *Facets Help Move from Search to Found* [White paper]. Retrieved
     November 27, 2009, from
     http://www.dataharmony.com/library/whitePapers/facets_help_move_from_search_to_fo
     und.html

Schmitz, P. (2006). Inducing ontology from flickr tags. In *WWW 2006, May 22–26, 2006,
     Edinburgh, UK*. IW3C2.

Schwartz, C. (2008). Thesauri and facets and tags, oh my! A look at three decades in subject
     analysis. *Library Trends,* 56 (4), 830-842.

Shirky, C. (2005). *Ontology is Overrated: Categories, Links, and Tags*. Retrieved January 12,

2009, from http://www.shirky.com/writings/ontology_overrated.html

Sinha, R. (2005, September 27). A Cognitive Analysis of Tagging. *Rashmi's blog*. Retrieved May 10, 2009, from http://rashmisinha.com/2005/09/27/a-cognitive-analysis-of-tagging/

Smadja, F. (2005, December 5). *Searching the TagSpace with RawSugar*. PowerPoint Lecture presented at the IBM IR Seminar. Retrieved November 22, 2009, from http://www.haifa.ibm.com/Workshops/ir2005/papers/Searching_the_TagSpace_with_RawSugar2005.pdf

Smith, T. (2007). *Cataloging and You*: *Measuring the Efficacy of a Folksonomy for Subject Analysis*. Master's thesis, School of Information Resources and Library Science and Learning Technologies Center, University of Arizona.

Spalding, T. (2007a, February 20). When tags work and when they don't: Amazon and LibraryThing. *Thingology blog*. Retrieved September 30, 2008, from http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php

Spalding, T. (2007b, July 24). TagMash: Book tagging grows up. *Thingology blog*. Retrieved November 10, 2009, from http://www.librarything.com/thingology/2007/07/tagmash-book-tagging-grows-up.php

Spiteri, L. (1998). A Simplified Model for Facet Analysis. *Canadian Journal of Information and Library Science,* 23, 1-30. Retrieved December 5, 2008, from http://iainstitute.org/en/learn/research/a_simplified_model_for_facet_analysis.php

Steckel, M. (2002). *Ranganathan for IAs: An Introduction to the Thought of S.R. Ranganathan for Information Architects*. Retrieved September 3, 2009, from http://www.boxesandarrows.com/view/ranganathan_for_ias

Sterken, V. (2008). *Classified: Analysis of user generated metadata in the LibraryThing folksonomy*. Unpublished Master Thesis, Master in Busisness Information and Service Management, The University of Brussels.

Stoica, E., Hearst, M. & Richardson, M. (2007). Automating Creation of Hierarchical Faceted Metadata Structures. In *Proceedings of NAACL-HLT, Rochester NY, April 2007*.

Trant, J. (2009). Studying Social Tagging and Folksonomy: A Review and Framework. *Journal*

of Digital Information, 10(1). Retrieved September 20, 2009, from
http://journals.tdl.org/jodi/article/view/269

Uddin, M. & Janecek, P. (2007). Faceted classification in web information architecture: A
framework for using semantic web tools. *The Electronic Library*, 25(2), 219-233.

Vander Wal, T. (2007, February 2). Folksonomy coinage and definition. *Folksonomy ::
Vanderwal.net*. Retrieved September 1, 2009, from
http://www.vanderwal.net/folksonomy.html

Vickery, B.C. (1960). *Faceted Classification: A guide to construction and use of special
schemes*. London: Aslib.

Vickery, B.C. (1966). *Faceted Classification Schemes*. New Brunswick, N.J.: The Graduate
School of Library Services, Rutgers.

Vickery, B.C. (1975). *Classification and Indexing in Science (3rd. ed.)*. London: Butterworths.

Vickery, B.C. (2008?). *The structure of subject classifications for document retrieval.* Retrieved
March 9, 2009, from http://www.lucis.me.uk/classification.htm

Weaver, M. (2007). Contextual metadata: Faceted schemas in virtual library communities.
Library Hi Tech, 25 (4), 579-594.

Weinberger, D. (2006).*Taxonomies and Tags: From Trees to Piles of Leaves*. Retrieved April 28,
2009, from http://www.hyperorg.com/blogger/misc/taxonomies_and_tags.html

Wilson, T. (2006). *The strict faceted classification model*. Retrieved October 3, 2008, from
http://www.facetmap.com/pub/strict_faceted_classification.pdf

Xiao, Y. (1994). Faceted Classification: A Consideration of its Features as a Paradigm of
Knowledge Organization. *Knowledge Organization*, 21(2), 64-68.

Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). Faceted metadata for image search and
browsing. In *CHI '03: Proceedings of the conference on Human factors in computing
systems*, pages 401-408. ACM Press.

APPENDICES

Appendix 1: List of books[56] represented by tags in the dataset

1. 1_Guns, Germs and Steel / Jared Diamond
2. 3_The Devil in the White City / Erik Larson
3. 4_The Professor and the Madman / Simon Winchester
4. 6_People's History of the United States: 1492 to Present / Howard Zinn
5. 8_1776 / David McCullough
6. 9_Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time / Dava Sobel
7. 12_Lies My Teacher Told Me: Everything Your American History Textbook Got Wrong / James W. Loewen
8. 14_Assassination Vacation / Sarah Vowell
9. 16_America (The Book): A Citizen's Guide to Democracy Inaction / Jon Stewart
10. 17_Bury My Heart at Wounded Knee: An Indian History of the American West / Dee Brown
11. 19_Salt; A World History / Mark Kurlansky
12. 21_The Histories / Herodotus
13. 22_How the Irish Saved Civilization / Thomas Cahill
14. 23_A History of God / Karen Armstrong
15. 25_A Distant Mirror; The Calamitous 14th Century / Barbara W. Tuchman
16. 26_The Code Book / Simon Singh
17. 27_Founding Brothers / Joseph J. Ellis
18. 29_Krakatoa / Simon Winchester
19. 32_The Mother Tongue / Bill Bryson
20. 39_Midnight in the Garden of Good and Evil / John Berendt
21. 42_1491: New Revelations of the Americas Before Columbus / Charles C. Mann
22. 44_Band of Brothers : E Company, 506th Regiment, 101st Airborne from Normandy to Hitler's Eagle's Nest / Stephen E. Ambrose
23. 45_Mayflower: A Story of Courage, Community, and War / Nathaniel Philbrick
24. 46_The Meaning of Everything / Simon Winchester
25. 52_The History of the Peloponnesian War: Revised Edition / Thucydides
26. 53_The Perfect Storm / Sebastian Junger
27. 54_Black Hawk Down: A Story of Modern War / Mark Bowden
28. 56_Rise And Fall Of The Third Reich / William L. Shirer
29. 58_The Discoverers / Daniel J. Boorstin
30. 63_Don't Know Much About History: Everything You Need to Know about American… / Kenneth C. Davis

---

[56] All of the books are presented with an ID representing their ranking in the initial list of 250 books: ID_Title / Author

Appendix 2: A random sample of results presented at the book-level

1_Guns, Germs and Steel / Jared Diamond

PERSONALITY:
      BY SUBJECT:
            BASIC FACET: Agriculture(43) anthropology(581) archaeology(48) biology(79) cultural studies(19) ecology(55) economics(57) ethnology(48) geography(149) history(1,452) Natural History(22) popular science(24) Science(454) Social History(33) social science(42) sociology(261) technology(70) World History(145)
            PERSONALITY:
                  BY ENTITY: Civilization(169) culture(128) environment(62) societies(13) society(95) germs(14)]
            ENERGY: culture diffusion(18) development(21) disease(35) epidemiology(19) evolution(114) politics(38) social evolution(49) war(29)
            SPACE: World(36)
            TIME: prehistory(16)
MATTER:
      BY GENRE:
            BY TRUTH ASPECT: nf(15) non-fiction(808)
      BY FORMAT: paperback(15)
ENERGY:
      BY ACTIVITY (WORK): read(105) tbr(25) unfinished(14) unread(102)
      BY ACTIVITY (PHYSICAL OBJECT): own(52) WishList(23)
AGENT:
      BY AUTHOR: jared diamond(17)
EXTERNAL RECEPTION:
      BY AWARD: pulitzer prize(68)

9_Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time / Dava Sobel

PERSONALITY:
      BY SUBJECT:
            BASIC FACET: astronomy(23) British History(13) engineering(9) European History(7) geography(50) history(501) history of science(83) history of technology(11) maritime history(13) mathematics(11) naval history(7) popular science(24) science(385) technology(26)
            PERSONALITY:
                  BY ENTITY: chronometers(10) clocks(34) invention(10) maps(16) maritime(25) naval(10) time(37)
                  BY PERSON: John Harrison(14)
            ENERGY: cartography(15) discovery(10) exploration(34) horology(13) mapping(8) measurement(7) navigation(94) sailing(16) Seefahrt(7)
            SPACE: britain(13) england(20) longitude(58) sea(7)
            TIME: 18th century(25)
MATTER:
      BY GENRE:
            BY TRUTH ASPECT: fiction(9) non-fiction(254)
            BY PLOT: biography(143) historical(11) nautical(26)
ENERGY:
      BY ACTIVITY (WORK): read(37) tbr(10) unread(16)
      BY ACTIVITY (PHYSICAL OBJECT): own(6)

19_Salt; A World History/ Mark Kurlansky

PERSONALITY:
    BY SUBJECT:
        BASIC FACET: Anthropology(7) Chemistry(9) commodity history(4) culinary history(6) cultural history(6) cultural studies(4) economic history(7) Economics(8) food history(20) Geology(8) history(385) history of science(4) microhistory(6) natural history(4) popular science(3) science(17) social history(12) sociology(5) Technology(4) World history(37)
        PERSONALITY:
            BY ENTITY: cookery(3)culture(6) food(144) nature(3) salt(93)
            BY PERSON: foodie(4)
        ENERGY: Cooking(16) politics(5) trade(9)
        SPACE: global(3) world(10)
MATTER:
    BY GENRE:
        BY TRUTH ASPECT: nf(5) non-fiction(170)
        BY PLOT: Culinary(3) Food Writing(7) Popular History(3)
        BY USE: Reference(4)
    BY FORMAT: Audiobook(6)
ENERGY:
    BY ACTIVITY (WORK): read(20) tbr(11) unread(28)
    BY ACTIVITY (PHYSICAL OBJECT): own(6) wishlist(7)
AGENT:
    BY USER (subfacet by age): children's nonfiction(3)
SPACE:
    BY PLACE: library(5)
TIME:
    BY YEAR PUBLISHED: 2006(3) 2007(3)

27_Founding Brothers/ Joseph J. Ellis

PERSONALITY:
    BY SUBJECT:
        BASIC FACET: american history(217) history(459) us history(55)
        PERSONALITY:
            BY ENTITY: constitution(11) revolutionary(5)
            BY PERSON: aaron burr(7) adams(6) alexander hamilton(12) Benjamin Franklin(6) founders(8) Founding Fathers(46) george washington(12) hamilton(6) james madison(6) jefferson(6) John Adams(16) Presidents(21) Thomas Jefferson(15) washington(6) [TOTAL: 173]
            BY GROUP: government(10)
        ENERGY: american revolution(108) politics(27) revolution(30) revolutionary war(54)
        SPACE: America(33) american(36) colonial america(6) united states(20) us(10) usa(10)
        TIME: 18th century(26) Colonial(10) colonial america(6) early republic(8)
MATTER:
    BY GENRE:
        BY TRUTH ASPECT: nf(5) non-fiction(166)
        BY PLOT: biography(139) historical(7)
ENERGY:
    BY ACTIVITY (WORK): read(10) tbr(11) unread(28)
    BY ACTIVITY (PHYSICAL OBJECT): own(15) owned(5)
TIME:
    BY YEAR PUBLISHED: 2007(5)
EXTERNAL RECEPTION:
    BY AWARD: Pulitzer prize(19)

44_Band of Brothers : E Company, 506th Regiment, 101st Airborne from Normandy to Hitler's Eagle's Nest/ Stephen E. Ambrose

PERSONALITY:
    BY SUBJECT:
        BASIC FACET: American history(40) European History(4) history(304) military history(94) US history(7)
        PERSONALITY:
            BY GROUP: 101st airborne(15) Airborne(12) Army(8) Easy Company(7) infantry(3) military(85)
            Paratroopers(8) US Army(12)
        ENERGY: battle of the bulge(4) d-day(17) European Theater(7) war(78) Warfare(8) wwii(378)
        SPACE: america(10) American(8) Europe(20) france(9) germany(6) normandy(10) usa(6)
        TIME: 20th century(10) modern(3)
    BY TITLE: Band of Brothers(6) [TOTAL: 6]
MATTER:
    BY GENRE:
        BY TRUTH ASPECT: fiction(7) historical fiction(4) non-fiction(139)
        BY PLOT: biography(11)
        BY FORM: novel(5)
    BY FORMAT: paperback(6)
ENERGY:
    BY ACTIVITY (WORK): Read(18) tbr(6) unread(8)
    BY ACTIVITY (PHYSICAL OBJECT): own(7)
AGENT:
    BY AUTHOR: Ambrose(6) stephen ambrose(6)
EXTERNAL RECEPTION:
    BY NEW EXPRESSION: Movie(4)

56_Rise And Fall Of The Third Reich/ William L. Shirer

PERSONALITY:
    BY SUBJECT:
        BASIC FACET: 20th century history(7) european history(38) german history(49) history(522) history - WWII(4) military
        history(31) modern history(6) political history(3) world history(9) [TOTAL: 669]
        PERSONALITY:
            BY PERSON: hitler(52)
            BY GROUP: military(13) nazi party(4) nazis(33)
            BY ENTITY: fascism(12) Nationalsozialismus(5) nazi(42) nazism(34)
        ENERGY: genocide(4) Holocaust(49) politics(14) war(40) warfare(5) wwii(365) [TOTAL: 477]
        SPACE: europe(33) Germany(174) nazi germany(29) third reich(40) [TOTAL: 276] german(13)
        TIME: 20th century(27) [TOTAL: 27]
MATTER:
    BY GENRE:
        BY TRUTH ASPECT: non-fiction(123)
        BY FORMAT: Hardcover(6) paperback(3)
ENERGY:
    BY ACTIVITY (WORK): read(15) tbr(4) unread(21)
    BY ACTIVITY (PHYSICAL OBJECT): own(9)
AGENT:
    BY AUTHOR: Shirer(4)
    BY PUBLISHER: Folio Society(8)
EXTERNAL RECEPTION:
    BY AWARD: National Book Award(4)

73_Portrait Of A Killer: Jack The Ripper -- Case Closed/ Patricia Cornwell

PERSONALITY:
>BY SUBJECT:
>>BASIC FACET: art(6) english history(5) history(81)
>>PERSONALITY:
>>>BY PERSON: jack the ripper(93) serial killer(30) Sickert(3) Walter Sickert(5)
>>>BY GROUP: scotland_yard(5)
>>ENERGY: crime(90) criminal investigation(4) forensics(15) murder(19) serial murder(3)
>>SPACE: England(23) London(16) uk(4) Victorian London(3)
>>TIME: 19th century(10) victorian(4) Victorian London(3)

MATTER:
>BY GENRE:
>>BY TRUTH ASPECT: crime fiction(4) Fiction(20) nf(4) non-fiction(160) true crime(138)
>>BY PLOT: biography(22) historical(10) Historical Mystery(4) murder mystery(4) mystery(49)
>>BY FEELING: suspense(7) thriller(7)
>>BY TIME-DEFINED STYLE: contemporary(3)
>BY FORMAT: Hardcover(5) paperback(6)
>BY EDITION: First Edition(5)

ENERGY:
>BY ACTIVITY (WORK): read(16) tbr(13) unread(21)
>BY ACTIVITY (PHYSICAL OBJECT): Own(8) owned(5) Signed(3)

AGENT:
>BY AUTHOR: cornwell(9) Patricia Cornwell(11)

85_King Leopold's Ghost/ Adam Hochschild

PERSONALITY:
>BY SUBJECT:
>>BASIC FACET: african history(34) african studies(3) belgian history(4) colonial history(4) european history(11) history(255) History-Africa(4) world history(5)
>>PERSONALITY:
>>>BY ENTITY: Heart of Darkness(3) human rights(11) imperialism(24) race(4) rubber(5)
>>>BY PERSON: conrad(3) King Leopold(4) Stanley(3) [TOTAL: 10]
>>ENERGY: colonialism(68) colonization(3) Exploration(3) genocide(18) politics(6) slavery(23)
>>SPACE: africa(190) African(5)belgian congo(8) Belgien(5) Belgium(60) Congo(92) empire(4) Europe(7) Kongo(7) zaire(11)
>>TIME: 19th century(11) 20th Century(6) colonial(4)

MATTER:
>BY GENRE:
>>BY TRUTH ASPECT: non-fiction(108)
>>BY PLOT: biography (16) historical(4)

ENERGY:
>BY ACTIVITY (WORK): read(12) Read in 2007(3) tbr(4) to read(6) unread(13)
>BY ACTIVITY (PHYSICAL OBJECT): owned(3)

TIME:
>BY YEAR PUBLISHED: 2007(4)

97_The Worst Hard Time/ Timothy Egan

PERSONALITY:
        BY SUBJECT:
                BASIC FACET: agriculture(7) American History(38) economics(5) history(135) us history(11)
                PERSONALITY
                        BY ENTITY: environment(5) nature(2) poverty(3)
                ENERGY: depression(20) disaster(3) drought(6) droughts(3) Dust Bowl(74) dustbowl(11) great depression(49)
                SPACE: america(5) American(4)American West(3) great plains(12) Kansas(3) midwest(2) oklahoma(13) Texas(8) the west(2) united states(6) usa(4)
                TIME: 1930s(15) 20th century(6) depression era(4)
MATTER:
        BY GENRE:
                BY TRUTH ASPECT: historical non-fiction(2) non-fiction(98)
                BY PLOT: Historical(3)
        BY FORMAT: Audio(3)
ENERGY:
        BY ACTIVITY (WORK): read(12) tbr(9) to read(5) unread(9)
        BY ACTIVITY (PHYSICAL OBJECT): borrowed(3) own(3) wishlist(5)
AGENT:
        BY USER: Book Group(4)
TIME:
        BY YEAR PUBLISHED: 2007(9) 2008(2)
EXTERNAL RECEPTION:
        BY AWARD: National Book Award(9)

111_The Island of Lost Maps: A True Story of Cartographic Crime/ Miles Harvey

PERSONALITY:
        BY SUBJECT:
                BASIC FACET: Geography(12) history(91) history of science(3)
                PERSONALITY:
                        BY ENTITY: books(5) books about books(7) map(3) Maps(92) rare books(4)
                        BY PERSON: thieves(2)
                ENERGY: cartography(81) collecting(3) crime(53) exploration(6) Map Collecting(3) map theft(3) Map Thefts(4) theft(15)
                SPACE: American(5) archives(8)Libraries(32) museums(2)
                TIME: 20th century(3)
MATTER:
        BY GENRE:
                BY TRUTH ASPECT: fiction(8) nf(3) non-fiction(106) true crime(20) true stories(2)
                BY PLOT: adventure(3) biography(8) historical(2) mystery(4) travel(9)
        BY FORMAT: paperback(2) Trade Paperback(2)
        BY EDITION: first edition(2)
ENERGY:
        BY ACTIVITY (WORK): read(8) tbr(5) to read(2) unread(16)
        BY ACTIVITY (PHYSICAL OBJECT): Borrowed(2) signed(2) wishlist(2)
SPACE:
        BY PLACE: f porch(2) library(4)
TIME:
        BY YEAR PUBLISHED: 2000(2)

134_Marie Antoinette: The Journey/ Antonia Fraser

PERSONALITY:
        BY SUBJECT:
                BASIC FACET: European History(11) French history(39) history(151) women's history(4)
                PERSONALITY:
                        BY ENTITY: court life(2)
                        BY PERSON: consort of Louis XVI(3) King of France(3) louis xv(3) Louis XVI(9) Marie Antoinette(46) queen(3) Queens(9) Women(8)
                        BY GROUP: Family(4) monarchy(5) royalty(31)
                ENERGY: french revolution(41) War(2)
                SPACE: Europe(5) France(88) french(12) paris(4) versailles(5)
                TIME: 1774-1793(2) 18th century(20)
MATTER:
        BY GENRE:
                BY TRUTH ASPECT: Fiction(3) Historical Fiction(5) non-fiction(106)
                BY PLOT: biography(229) historical(16) Historical Biography(8) memoir(3)
        BY FORMAT: paperback(2)
ENERGY:
        BY ACTIVITY (WORK): currently reading(3) read(10) tbr(6) To Read(4) unread(19)
        BY ACTIVITY (PHYSICAL OBJECT): Own(5) owned(3)
AGENT:
        BY AUTHOR: Antonia Fraser(5)
SPACE:
        BY PLACE: Library(3)
TIME:
        BY YEAR PUBLISHED: 2006(3) 2007(7)
EXTERNAL RECEPTION:
        BY NEW EXPRESSION: movie(3)

160_Eleanor of Aquitaine: A Life/ Alison Weir

PERSONALITY:
        BY SUBJECT:
                BASIC FACET: british history(22) english history(18) European history(10) french history(7) history(217) medieval history(29) women's history(6)
                PERSONALITY:
                        BY PERSON: Eleanor(8) eleanor of aquitaine(36) famous women(3) Henry II(7) queen(8) queens(9) women(15)
                        BY GROUP: Angevin(2) british monarchy(3) british royalty(2) Monarchy(5) plantagenet(9) royal family(2) royalty(29)
                ENERGY: crusades(6) politics(3)
                SPACE: Aquitaine(3) britain(9) england(57) Europe(6) france(47) british(6) European(4) french(3)
                TIME: 12th century(17) medieval(54) Middle Ages(22)
MATTER:
        BY GENRE:
                BY TRUTH ASPECT: fiction(4) historical fiction(7) Fiction(3) nf(3) non-fiction(80)
                BY PLOT: bio(5) biography(206) historical(8) historical biography(8) Historical Non-Fiction(3) nf(3) non-fiction(80)
ENERGY:
        BY ACTIVITY (WORK): read(7) tbr(11) unread(12)
        BY ACTIVITY (PHYSICAL OBJECT): own(8)
SPACE:
        BY PLACE: Library(3)

206_The Rape of Nanking: The Forgotten Holocaust of World War II/ Iris Shun-Ru Chang

PERSONALITY:
        BY SUBJECT:
                BASIC FACET: Asian History(5) Asian studies(3) chinese history(19) history(184) japanese history(8) military history(12) world history(6)
                PERSONALITY:
                        BY PERSON: Women(2)
                        BY GROUP: military(7)
                        ENERGY: atrocities(6) battles and campaigns(2) genocide(10) Holocaust(9) massacre(4) pacific theatre(3) politics(3) Rape(4) Rape of Nanking(4) Torture(3) war(26) war crimes(3) wwii(134)
                SPACE: asia(11) china(129) japan(79) nanjing(5) nanking(15) Pacific(2) [TOTAL: 241] asian(4) Asian American(2) chinese(4) japanese(2)
                TIME: 20th century(6)
MATTER:
        BY GENRE:
                BY TRUTH ASPECT: nf(2) non-fiction(76)
                BY PLOT: historical(3)
ENERGY:
        BY ACTIVITY (WORK): read(9) tbr(7) to read(2) unread(13)
        BY ACTIVITY (PHYSICAL OBJECT): own(3)
AGENT:
        BY AUTHOR: chang(2)
SPACE:
        BY PLACE: storage(2)
EXTERNAL RECEPTION:
        BY RATING: recommended(2)

244_Don't Let's Go to the Dogs Tonight/ Alexandra Fuller

PERSONALITY:
        BY SUBJECT:
                BASIC FACET: history(6)
                PERSONALITY:
                        BY PERSON: expats(3) sisters(5) women(5)
                        BY GROUP: family(10)
                        BY ENTITY: race(3) racism(7)
                  ENERGY: African childhood(3) alcoholism(6) apartheid(8) childhood(32) civil war(8) colonialism(9) coming of age(6) growing up(3) war(9)
                SPACE: africa(205) african(5)malawi(12) Mozambique(4) rhodesia(29) south africa(3) zambia(14) zimbabwe(68)
                TIME: 1970s(3) 20th Century(6)
MATTER:
        BY GENRE:
                BY TRUTH ASPECT: fiction(12) nf(4) non-fiction(98)
                BY PLOT: adventure(3) autobiography(45) biography(48) biography/autobiography(5) childhood memoir(4) memoir(229) travel(4)
        BY FORMAT: audiobook(3)
ENERGY:
        BY ACTIVITY (WORK): read(10) tbr(12) to read(4) unread(13)
        BY ACTIVITY (PHYSICAL OBJECT): Own(13)
AGENT:
        BY USER: book club(15)
TIME:
        BY YEAR PUBLISHED: 2005(5) 2006(5) 2007(3)
EXTERNAL RECEPTION:
        BY RATING: favorite(4)