

# Crowdsourcing Tags in Accounting and Finance: Review, Analysis, and Emerging Issues

Daniel E. O’Leary

*University of Southern California*

**ABSTRACT:** This paper examines the use of crowdsourcing informal and discretionary tags in accounting and finance. Crowd-provided tags employ short amounts of text to capture some characteristic(s) of documents, messages, financial information, and other objects. Increasingly, tagging is being used in systems for knowledge management, including facilitating search and categorization. The paper reviews previous research on tagging, summarizes accounting applications of crowdsourced tags, investigates problems associated with using crowdsourced tags in accounting, and generates a number of potential research issues. Empirical analysis of Delicious and Twitter data are used to illustrate some of the concepts associated with the emerging technology of crowdsourced tags in accounting and finance. Empirical analysis finds that users provide different tags for the same concept, potentially making it difficult to use tags for search purposes. In addition, users appear to generate redundant tags, with many tags simply capturing object title information.

**Keywords:** crowdsourcing; tagging; folksonomies; tagsonomies; errors in tags.

## INTRODUCTION

Increasingly tagging (using short chunks of text to describe an object) is being used to facilitate search and identification of a range of different objects, such as documents, messages, pictures, and financial information. Although there is a substantial literature associated with formal required tagging in accounting and finance that is required for regulatory purposes, such as, XBRL (e.g., [Bonsón, Cortijo, and Escobar 2008](#)), increasingly, there are alternative uses of tagging. Recently, tags have been built into accounting and human resource systems (e.g., Workday, see [Nittler 2012](#)).<sup>1</sup> However, to date tagging has only been integrated into a few accounting and finance

---

The author thanks the attendees at the 2013 American Accounting Association Annual Meeting for their comments. The author also thanks the referees for their comments on both of the previous versions of this paper. In particular, the author thanks one of the referees on the most recent version for extensive and in-depth comments.

An earlier version of this paper was presented at the 2013 American Accounting Association Annual Meeting in Anaheim, California.

*Published Online: June 2015*  
Corresponding author: Daniel E. O’Leary  
Email: [oleary@rcf.usc.edu](mailto:oleary@rcf.usc.edu)

---

<sup>1</sup> [Howlett \(2007\)](#).

systems. Accordingly, within accounting, such tagging is truly an emerging technology. Unfortunately, there is a very limited literature associated with *informal* tags (not from a standardized taxonomy) that are not required but instead provide *discretionary* information gathered from the *crowd* about accounting and financial objects.

In order to remedy this limitation in the literature, this research analyzes the use of tags based on accounting concepts. Unfortunately, users encounter a number of difficulties while tagging. As an example, an empirical analysis of tag data gathered from two settings, discussed below, finds that the most frequently used tag to represent a concept would find roughly 70 percent to 90 percent of the objects where that concept was used. However, 10 percent to 30 percent of the tags did not use the most common representations, so a search based on those less used versions would ignore 70 percent to 90 percent of the material. Further, the analysis finds that spelling errors and the lack of system identification of symbols (“/”) can further degrade the successful use of tags. In addition, since the tagging is discretionary, searching tags also ignore objects that were not tagged. Accordingly, this analysis suggests that relying exclusively on informal and discretionary tags may be inappropriate for those critical settings in which a higher precision is needed. Instead, such tags can supplement existing search information and potentially help evolve existing taxonomies and ontologies to capture emerging ideas.

### Crowdsourcing and Tags

Crowdsourcing and tags are two recent concepts that have received, at most, limited attention in the accounting and finance literatures, particularly in an “enterprise” setting. [Howe \(2006\)](#) first defined crowdsourcing in contrast to outsourcing. In particular, crowdsourcing refers to “the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers.”<sup>2</sup> When crowdsourcing is brought into organizations, the “community” involved is typically the organization in which the crowdsourcing is done, and may be referred to as “Enterprise Crowdsourcing.” Within the enterprise, the “crowd” may include portions of the organization not usually involved in the particular activity. For example, administrative workers may provide information that is used by client-facing personnel. As a result, in the case of accounting and finance, people not part of accounting and finance may be among the crowd providing tags, services, ideas, and content. As a result, a misunderstanding of accounting concepts could lead to some inappropriate tagging.

Tags are the part of the Web 2.0 (e.g., [O'Reilly 2005](#)) that focuses on users generating more content in web environments. In some settings those tags have become so frequently used (e.g., Twitter) that the accompanying term “Hashtag” (used to describe the symbol “#”), has become embedded in the culture to the extent that it is at the base of television sketches (e.g., [Fallon and Timberlake 2013](#)). In the case of Twitter, tags have evolved to include other uses beyond search and identification. For example, frequently used tags are used to delineate “trending topics.” However, tags apparently were first used in social media to capture information about pictures and information, such as articles or news stories (e.g., [Mathes 2004](#)).

Although there is limited empirical evidence, it appears that the use of tags is increasing over time. A survey by Pew Internet ([Rainie 2007](#)) found that roughly 28 percent of Internet users have tagged material/resources. More recently, [Glenn \(2012\)](#) reported that roughly 60 percent of mobile users tag their posts and pictures with location tags. These results suggest an increasing use of discretionary tags. Accordingly, the number of tags being developed appears to be increasing and has become quite substantial.

---

<sup>2</sup> Merriam-Webster (n.d.).

### Taxonomies versus Folksonomies (Tagsonomies)

Historically, a key aspect of knowledge management has been the central development and maintenance of ontologies and taxonomies used to organize and search knowledge. Typically, experts (e.g., “librarians” or “consultants”) develop and apply a consistent set of terms that define the ontologies and taxonomies. Experts control and distribute the vocabulary and rules for its use.

However, in an Internet world in which the rapidly expanding content is increasingly provided by users, an alternative approach comes from aggregating the knowledge in the crowd-based tags to provide insights into what developers and users of the materials have proposed as categorization of the specific objects. Since those tags are not formal or have not been proposed or managed by the librarians or other formal sources, these groups of tags have been proposed as creating “folksonomies.” Folksonomy, a combination of the words “folk” and “taxonomy” is a term used to capture the notion that users, not experts, can add “tags” to pictures, documents, messages, videos, and other objects to help others find information or to describe a particular object or situation (e.g., [Gruber 2007](#)). Further, the term folksonomy captures the notion that tags are informal and discretionary. A similar word “tagsonomy” also captures both the notion of informal tags and the use of a crowdsourced group to provide an informal approximation to a taxonomy of categories or at least provide some of the same capabilities, such as search and concept identification.<sup>3</sup>

### Scope of this Paper

Crowdsourced tags are an emerging technology in accounting practice. Firms such as Workday ([Nittler 2012](#)) have made crowdsourcing tags a possibility in accounting systems. In addition, social media is being introduced to firms through proprietary social media systems such as Yammer. Further, other firms have examined such tags for their knowledge management systems ([Burgelman and Blumenstein 2007](#)). Unfortunately, there is limited academic information about the use of discretionary and informal tags. As a result, this paper draws on empirical data from publicly available tagging systems, including Delicious and Twitter, in order to examine issues associated with crowdsourcing tagging.

In particular, the scope of this paper is with crowdsourced informal (not part of some formal taxonomy or ontology) and discretionary user-provided tags. These two key characteristics are summarized in Figure 1 with some examples illustrating association with the particular concepts in cells 1–4. The primary focus of this paper is on quadrant 4.

- Workday, a well-known accounting system, uses tags as the basis of a chart of accounts and/or organization model that reflects required use of a formal taxonomy/chart of accounts (1);
- Authors of articles in journals are typically required to provide keywords that describe their contribution, but frequently they do not draw those keywords from a formal taxonomy (2);
- Wordpress.com, a blogging site, allows users to choose descriptor terms from a formal taxonomy, but does not require the use of such terms (3); and
- Finally, Workday, Wordpress.com, and other software allow the user to input informal tags without gathering the terms from any set taxonomy (4).

These and other dimensions are discussed in further detail in the “Tag Characteristics” section.

### Tags in Accounting and Finance

Within accounting and finance, informal tags can be used in many ways (e.g., [Nittler 2012](#)). For example, tags can be used to capture:

<sup>3</sup> [Morville \(2005\)](#).

**FIGURE 1**  
**Required versus Not Required and Formal versus Informal**

Required	Workday Chart of Accounts	Author Provided Keywords
	1	2
Not Required	3	4
	Wordpress.com Categories	Workday Optional Event Tags
	Formal	Informal

- concepts embedded in knowledge management systems, in order to facilitate search and knowledge reuse;
- locations of flows of information, e.g., tag where information is captured, transformed, or queried;
- for whom a report, query, etc. was done;
- expense or revenue categorization;
- rationale for transactions, etc., e.g., transaction descriptions; and
- a range of other uses.

**Purposes of this Paper**

The primary purpose of this paper is to investigate discretionary tagging without using a formal taxonomy, with particular focus on accounting and financial applications. In particular, this paper analyzes relatively unique problems associated with tagging using accounting and finance concepts. A number of concerns with accounting concept tags are found based on an empirical analysis (including an apparent lack of knowledge of accounting in tag assignment, assigning accounting descriptors to situations where they are inappropriate, inconsistency of multi-word accounting concepts, spelling errors, and a range of other tagging errors).

In addition, this paper investigates tags generated for a particular knowledge resource. In one example, roughly 84 percent of the tags come from words directly related to the title. Accordingly, discretionary and informal tagging may not be fully leveraging object or the crowd's information.

Finally, this paper examines a number of extensions of tags in order to lay out emerging potential research issues. A number of those extensions are designed to mitigate some of the difficulties identified in the empirical analysis.

**Outline of this Paper**

This paper proceeds in the following manner. The "Introduction" discussed the problem, provided some motivation, and specified the purpose of this paper. The following section investigates some of the previous literature, focusing on eight key characteristics or dimensions of tags. The third section analyzes the use of discretionary and informal tags in three different

accounting settings. The fourth section examines some of the limitations of discretionary and informal tags based on an empirical analysis of both Delicious and Twitter data. The fifth and sixth sections analyze two potential research issues, tagging strategies and development of taxonomies and ontologies from tags, in detail, respectively. The next section investigates five other potential research issues. The “Conclusion” section briefly summarizes the paper, examines its contributions, and briefly reviews other potential related research issues.

## TAG CHARACTERISTICS

Researchers have argued that there is no common definition of social tagging and folksonomies (e.g., [Voss 2012](#)). As a result, it is premature to suggest that there is a unified theory of social tagging. However, although there has been limited attention given to crowdsourcing tags in accounting and finance, there is a general literature associated with tagging. In their literature reviews, [Marlow, Naaman, Boyd, and Davis \(2006\)](#) and [Gupta, Li, Yin, and Han \(2011\)](#) find that a number of different characteristics of tags have emerged that can facilitate additional analysis, including the following:

- Required versus Discretionary
- Formal versus Informal
- Time of Occurrence versus Alternative Time
- Content Initiator versus Content User (“tagging rights”)
- Manual versus Automated (“source of tagging resources”)
- Individual versus Group (“tag aggregation”)
- Internal versus External (Enterprise versus Internet) or Public versus Private
- Named or Anonymous Tags

### Required versus Discretionary

In contrast to required tagging (e.g., XBRL), discretionary tags are a part of a wide range of applications, generated for tagging documents, pictures, messages, and virtually any other object. Among the first two websites that enabled use of discretionary tag materials were Delicious.com and Flickr.com (e.g., [Mathes 2004](#); [Wetzker, Zimmermann, and Bauckhage 2008](#)). Although the site has changed hands since its original development, Delicious.com allowed multiple users to label the same materials/resources using whatever tags they felt were appropriate. User tags could be made public and used to direct browsers to tagged materials. Similarly, Flickr.com is a website that allows users to upload and tag their pictures, ultimately sharing those tags among different members of the crowd. Further, the use of discretionary tags has gathered substantial momentum from Web 2.0 applications, such as Twitter, Facebook, and other social media. An example set of Twitter tags is given in Figure 2 (#materials, #chemicals, and #industrialbiotech).

Unfortunately, there has been limited research into the usage of discretionary tags in accounting and financial applications. As a result, there are a number of potential concerns associated with discretionary tagging, including the extent to which such tags are generated and used, since they are not required. Further, in a discretionary environment, there is interest in whether the crowd continues to provide tags over time; if so, then why; and if not, then why not?

### Formal versus Informal

The use of a formal ontology or taxonomy indicates that all users employ tags from a limited set of pre-defined words. The extent of that formal ontology or taxonomy can vary from a skeletal framework to a detailed framework with multiple levels and usage rules. Further, formal approaches often employ some type of theory to structure the information. Unfortunately, use of a formal

**FIGURE 2**  
**Sample Twitter Message with Detailed Industry Tag Information**



Source: Screen capture of @IpreoCapMarkets from Twitter website.

approach can require some training to make sure that users understand the taxonomy or framework and its application.

The classic Web 2.0 approach is to use an informal framework that allows the crowd to easily apply their own tags. Generally, no formal theory or structure is required to provide a tag. The crowd can bring its knowledge to choose tags. Virtually anyone can generate tags that are meaningful to them for particular settings. Accordingly, tagging is promulgated as requiring no specific skills, but providing immediate benefit to the participants through an improved ability to search, rank, and index materials (e.g., [Hotho, Jaschke, Schmitz, and Stumme 2006](#)). Tagging can reflect changes in the way that materials/resources are viewed, facilitating a rapid evolution of indexing capabilities. Accordingly, researchers have argued that folksonomies provide “power to the people” ([Quintarelli 2005](#)). Further, folksonomies and tagging are said to cater to users who have a unique view of the world and would search with or for a low-frequency keyword ([Wu, Zubair, and Maly 2006](#)). As a result, an informal framework is flexible and adaptable.

Unfortunately, informal tagging has been criticized because of the low quality of the tags, and the overwhelming volume of tags (e.g., [Wu et al. 2006](#)). Further, as illustrated later in this paper, the crowd can develop a number of errors as part of generating informal tags. In addition, tags can have multiple meanings, or other words/tags can have the same meaning (e.g., [Golder and Huberman 2005](#)) that required disambiguation.

To date, there has been limited research into the use of informal accounting and finance tags. As a result, there are a number of potential concerns including the quality of the tags, the ability of informal tags to capture the nature of the content of the material being tagged, the consistency of the tags across users and time, and other issues.

### Time of Occurrence versus Alternative Time

The crowd can add tags at the time the content is generated (occurs) or they can be generated at some future time. In the case of Twitter, tags are generated at the time the message is developed and sent. Similarly, chart of accounts tags and organization models (e.g., Workday tags/Worktags) are typically generated at the time of the event being recorded ([Nittler 2012](#)).

However, in some settings, tags can be added at any other time. In the case of Twitter, tags can be added when a tweet is retweeted. In the case of Delicious, tags could be added at any time. In addition, in the case of Workday, as noted by Nittler (2012), “The ability to tag these businesses attributes to events—either at the time they occur or a later date if needed—is the building block in providing our Financial Management customers with a complete, multi-dimensional picture of their operations.”

Unfortunately, there has been limited research into the impact of using tags at times other than the event time. Non-event time tags could be analyzed to investigate a number of issues, such as the extent to which such tags lead to “revising history.” In particular, to what extent are tags that occur after event times consistent with tags that occur at event time. Are (non-event time) tags simply used to clarify event time tag information or are there other rationales?

### **Content Initiator versus Content User**

The crowd has multiple potential roles, both as tag developer or tag user. Tags can be generated by the message or content initiator (creator) or by some user of the content or message. For example, in the case of Twitter, tags are inserted into messages by the person doing the Tweet. However, if there were a database of Twitter tweets or in the case of a retweet, then users of the content potentially could add a tag. Since the creator and users (or even different users) may have different views of the content, those tags could differ substantially.

As a result, there are a number of concerns. Do creators and users provide similar tags? Do the tags generated by the users or creators help more with specific tasks such as search? Further, potentially allowing others to provide additional tags also can generate alternative meanings of the original content. For example, a retweeter can change the meaning of an original tweet. As a result, research is needed to determine if there are situations in accounting and finance when it is not appropriate for content users to add tags, or at least that the content user tags should be clearly differentiated from those of content initiator tags. Finally research needs to be initiated to determine which groups provide the “best” tags, for example, which tags provide the most information about the particular object being tagged?

### **Manual versus Automated**

Although the focus of this paper is on manual tags and their use, a number of tags can be generated automatically using automated approaches.<sup>4</sup> For example, using automated approaches can be used to capture the source, the date, and other characteristics of the material under consideration.

Since the focus of this paper is on the manual generation of tags by users as part of a crowdsourced generation of the tags, the concerns would include:

- “What do manual tags offer over and above automated tags?”
- “To what extent are manual tags correlated with automated tags?”
- “Can we build a system to automatically develop tags similar to manual tags?”

### **Individual versus Group or Different Groups**

Tags can be limited to a particular individual or to a group. In some environments, for example enterprises, it might be beneficial for comments to be captured according to organizational unit. As a result, there are a number of concerns including the extent to which different groups use different

<sup>4</sup> See: <https://support.google.com/analytics/answer/1733663?hl=en>

tags to describe the same materials. It could be particularly important to gather tags from the accounting and finance crowd, compared to nonaccounting and finance, because of basic knowledge differences. If different groups have their own perspective, then there may be a need to keep the tags separate and examine differences between the two groups. Further, differences in tags can be used to better understand points of view or concerns. As a result, monitoring tags can provide critical insight into differences of understanding between different groups.

### **Internal versus External (or Public versus Private)**

Tags can be limited to a particular organization (internal—or “enterprise”) or tags can be widely available on the Internet (external). If tags are limited to an organization, then a number of potential enterprise-based issues could affect tags and tagging. A more detailed analysis of some of the issues related to enterprise tags are discussed further below, as part of the extensions.

### **Named or Anonymous Tags**

If people can attach their “names” (whether real or fictitious) to tags, then tags can provide social and other functions. In particular, with names, the crowd is not anonymous but users can connect to each other. In addition, with names, the crowd can track who is making contributions, the quality of those contributions, and how often they are making those contributions.

Unfortunately, in accounting and auditing, associating names with tags may be inappropriate. Oftentimes financial information systems are designed to keep different users at arm’s length. Potentially names associated with tags could connect the “wrong” users together. In addition, in enterprises, names associated with tags can cause potential unintended consequences. For example, if the president of a company attaches a tag and the name is visible, then it is unlikely that anyone in the organization will disagree with that tag. Further, others may attach the same tag as a means of showing agreement and alignment with the rationale of the president. Accordingly, issues such as organization hierarchy (and friendship, etc.) can influence tag use when user names are attached to the tags. This enterprise issue is discussed further below.

## **EXAMPLES OF CROWDSOURCED TAGS IN ACCOUNTING AND FINANCE**

The eight characteristics of the previous section can be illustrated in three different uses of tags.

### **Generating Workday Tags in Accounting Information Systems**

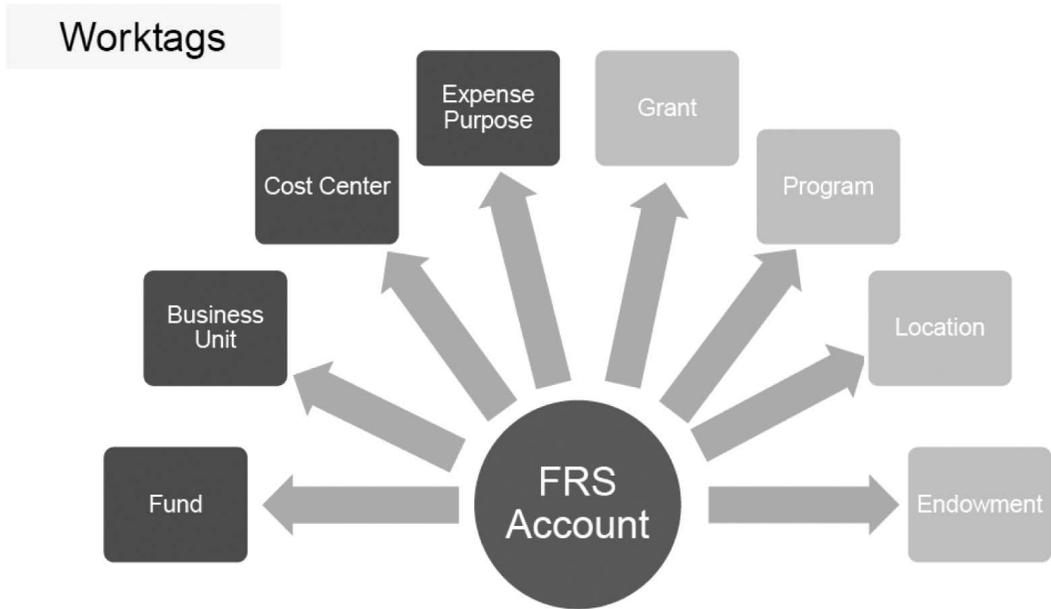
The cloud-based accounting and human resources software, Workday, allows the use of “Worktags.” The tags are used to “mark” different events so that information about those events can be summarized and reported. As noted by [Nittler \(2012\)](#):

Worktags are keywords assigned to business events, so our customers can aggregate, report, and analyze their business information within Workday. One example: An employee fills out a purchase order (an event), which can be assigned any number of tags, such as the name of the person making the purchase and the project it’s for. A tag could also be a department, customer, product, supplier, or subsidiary, while other types of events include a generated invoice, a submitted expense, or a payment received.

Worktags are not in a pre-specified hierarchy as is typically seen in a tree-based chart of accounts. Instead, information is captured directly about the particular tag. Accordingly, reports can be generated about the particular tag or set of tags of interest.

Tags can be assigned based on pre-established formal taxonomies or informally assigned by users at the time they record the event. For example, as seen in Figure 3, Brown University uses a

**FIGURE 3**  
**Formal Taxonomy of Workday Tags**



Available at: <http://brown.edu/about/administration/workday/sites/brown.edu.about.administration.workday/files/uploads/Welcome%20to%20Workday%20Financials.pdf> (last accessed on January 13, 2014).

formal taxonomy of “Workday Tags” as the basis of a chart of account information and organization model, which includes information such as “fund,” “business unit,” “cost center,” and other information, captured in the form of tags. University of Rochester (2013) also provides a summary of Workday tags for charts of accounts. Alternatively, Workday also allows users to put informal or user-specific tag information into the system.

Worktags also can be added to events at any point in time, either at the time of entry of the event data or at some later date. This aspect of tags illustrates their flexibility, but this aspect could represent a potential control issue. Unfortunately, it is easy to imagine “revisionist history” like issues for which events are tagged substantially after the fact to present evidence in support of or against some issue. For example, after a project fails, tags might be applied to indicate an association of different events with that project to make other projects appear profitable. As a result, some commentators have noted the concern of allowing the attachment of free text information.<sup>5</sup>

### Tagging Using Social Media: Categorizing Firms in Accounting Research

Investigating industry classification has been an active research area in accounting. For example, Bhojraj, Lee, and Oler (2003) compare four different industry classification schemes for a number of applications related to capital markets research. They find that the Global Industry Classifications Standard (GICS) classifications are significantly better than the other three at explaining stock return co-movements.

<sup>5</sup> Oracle.com (2007).

An alternative emerging approach is to analyze social media, such as Twitter messages for occurrence and co-occurrence of industry categories (e.g., hashtags) in Tweets or co-occurrence of firms in a Tweet. For example, in the Tweet above in Figure 2, a particular firm “BioAmber” is categorized according to three different segments, #materials, #chemicals, and #industrialbiotech, which do not occur together in the GICS map of firms, but are grouped in this Twitter message.

### Crowdsourced Tagging Article Information

Tags have been used in knowledge management in consulting firms. For example, according to Meister and Mark (2004), Hill & Knowlton started tagging so that they could try to turn tacit knowledge into explicit knowledge.

Knowledge management systems depend on users being able to easily search the various documents and other materials within the knowledge base. Burgelman and Blumenstein (2007) describe the use of tagging in a consulting firm designed to facilitate sharing information and connecting users with similar interests. In the system, all users would be able to add tags to virtually any document. The crowd could tag documents using any tags that it felt were appropriate. For example, reports might be tagged with a client name, project code, and topic. Ultimately, the goal was to “better share information across the firm and connect members with similar interests” Burgelman and Blumenstein (2007, 1). Further, as was also seen in Burgelman and Blumenstein (2007), one potential benefit of tags in business was the ability to “coin new phrases,” i.e., generate new knowledge or new ways of categorizing knowledge or capturing and categorizing innovative ideas. Both informal and formal taxonomies were planned; however, all tags would be made public. Management felt that by using formal and informal taxonomies the firm could create a “living” taxonomy. In order to facilitate generation of tags, the management team planned on having “tagging” parties, during which colleagues would get together and generate tags for different documents. Others, such as von Ahn and Dabbish (2004), also have approached tagging from the perspective of trying to make a game of it, “game-ifying” tagging.

### SEARCHING ACCOUNTING INFORMATION: LIMITATIONS OF TAGS

In the case of crowdsourced informal tags, as opposed to standardized tags such as XBRL, different people can use different tags to describe the same object, depending on their point of view or perspective (Bowen 2007). Accordingly, the quality of informal tags has been a concern. For example, Guy and Tonkin (2006) note “tags are often ambiguous, overly personalized and inexact.” In order to improve on that imprecision, some researchers provide approaches so that users’ tags can be improved. As an example, researchers have discussed educating the user, in terms of tagging best practices (e.g., Mejias 2005).

When users informally apply tags to accounting information, based on their discretion, there are a number of specific issues to consider. For example, a number of potential accounting tags consist of a group of words, such as, “accounts receivable.” Further, use of accounting words requires domain-specific knowledge of accounting. For example, in general there is interest in “accounts receivable,” not “account receivable.” In addition, there are many abbreviations that are specific to accounting information, such as, “A/R.” This section investigates these and other concerns about accounting tags, using data gathered from Delicious.com and Twitter, as summarized in Table 1.<sup>6</sup>

<sup>6</sup> Delicious.com data gathered December 2010, Twitter data gathered January 2014. Yahoo! sold Delicious in April 2011.

**TABLE 1**  
**Different Occurrences of “Accounts Receivable” Representations**

	Number in Delicious	Number in Twitter
Two Tag Data		
Both Accounts and Receivable	682	506
Both Accounts and Receivables	58	23
Both Account and Receivable	222	18
Both Account and Receivables	29	12
Totals—Two Tag Data	991	559
One Tag Data		
Accountreceivable	4	35
Accounts_Receivable	14	8
AccountsReceivable	32	1006
Accounts-Receivable	12	Not Feasible
Accounts_Recievable	1	0
A/R (Abbreviation)	140	Not Feasible
Accountreceivables	Not Gathered	8
Accountsreceivables	Not Gathered	69
Totals—One Tag Data	203	1126

### Methodology

Tags were gathered from Delicious.com and Twitter. Tags gathered from Twitter were gathered from “All Tweets” across “everything” using the Internet version of Twitter. Tag information was gathered using “#” and then the word. For the first four line items in Table 1 (in the Twitter column), two tags were used to gather information: #Accounts #Receivable.

Apparently the symbols “-” and “/” are not legitimate search symbols in Twitter, but “\_” is a legitimate search symbol. By themselves “-” and “/” yield “invalid search query” and when coupled with other terms, only the other terms are searched. Excluding certain terms, such as “/,” from search could be particularly difficult on disciplines such as accounting, finance, and engineering. For example, in Twitter (at the time of this writing) a search for “1/2” yields 12. Similarly, the approach for a classic “A/R” (accounts receivable) actually searches for AR.

Tag data were gathered from the original version of Delicious.com, but no data were gathered on the last two line items in Table 1. Interestingly, although originating the notions of tags, Delicious tag data did not include “#,” the current Twitter symbol used to denote a hashtag.

### Comparison of Findings on Delicious.com and Twitter.com

The counts between Delicious.com and Twitter.com are very similar, with one glaring difference. Twitter users apparently aggregate the two words “AccountsReceivable” without using symbols to link the words together. Considering the remaining seven items, for which data are available for both Twitter and Delicious data, the correlation is 0.9515 and is significant at better than 0.01, suggesting substantial similarity between the Delicious and Twitter data.

## Multiple Words

Multiple word tags provide a challenge to taggers. How do they accommodate the multiple words? For example, the term “Accounts Receivable” could be disambiguated as two words: “Accounts” and “Receivable.” Unfortunately, it is unlikely that every occurrence of those words individually was meant by the tagger to refer to “Accounts Receivable.” For example, in this Delicious.com analysis there were 38,358 uses of the tag “Accounts” and 777 of the tag “Receivable” and 682 occurrences of them when both tags were used together.

Taggers may make the two words into a single tag just running the two words together, for example, #accountsreceivable—the apparent choice in Twitter.<sup>7</sup> Alternatively, they can use some character to separate the words into distinguishable words, for example, using a dash or an underscore. There are a number of different approaches as seen in Table 1. Unfortunately, each of these choices is a different representation of the concepts and the existence of multiple digital representations can hinder finding information.

## Multiple Meanings/Abbreviations

Many words and abbreviations have multiple meanings. For example, an abbreviation for “Accounts Receivable” is “A/R.” As seen in Table 1 there are a number of documents that have been attributed the tag “A/R.” Unfortunately, that term has multiple meanings. For example, not only does A/R refer to “Accounts Receivable,” but also in Delicious.com it referred to “Age Regression.”

As noted above, Twitter does not seem to allow search of the term “#A/R,” thus excluding a number of potential accounting terms, for example, “A/P” for accounts payable. In particular, as noted above, the symbol “/” is not accommodated in a Twitter search for hashtagged words.

## Singular versus Plural

Using accounting words typically requires an understanding of accounting. For example, in general, accountants talk of “accounts payable” or “accounts receivable” where those terms are plural, rather than singular, unless they are talking about a particular account. Users may apply tags “misusing” a word. For example, as seen in Table 1, not only is there a tag “AccountsReceivable,” but also there is a tag “AccountReceivable” (no “s”). Typically, the term is set using a plural, but in four cases the tagger employed a singular version. Unfortunately, someone looking under the first tag is unlikely to also look under the second tag.

## Misspellings

As users implement tags, simple issues such as misspellings can have an impact on the quality of the tags. For example, Delicious.com contained the tag “Accounts\_Recievable” (with “Receivable” misspelled). Although an external analyst can never know the intention of the tagger with certainty, they can “guess” that there was a misspelling. For example, in the case of spelling, order-errors of “ie” and “ei” are common. In any case, for any other tagger to find the tag, they also would need to misspell the term.

In addition, #AccountsRecievable (hashtag with “receivable” misspelled) was found in Twitter once in 2009, three times in 2010, 11 times in 2011, 20 times in 2012, 18 times in 2013, and 33 times in 2014. It is clear that a misspelled hash tag would not be good publicity for a group whose

<sup>7</sup> Although multiple word tags are not unique to accounting and finance, as illustrated here, they provide an important potential concern.

business is processing accounts receivables. However, perhaps the Twitter tweeters are thinking that their potential clients cannot spell correctly either. It would seem that the more that a term is misspelled the more likely that it would be used and found, although that is an empirical question. However, if a term is frequently misspelled, then perhaps it would help to include the misspelling as a tag.

### **Mis-Tagging**

Further, in the case of user-applied tags, tags may be applied to objects that do not correspond to the tag. As an example, in more than one case (both in Delicious and Twitter), it was found that materials tagged with one of the variants of “accounts receivable” did not contain any information about accounts receivable or usage of the words. Accordingly, users guided to material with the tag may not actually find the information that they were searching for.

### **Context**

Context could be important, but unfortunately taggers may or may not supply the context. In the case of tags, context information is the “other” tags. For example, a user may be browsing for materials related to “accounts receivable,” with particular interest in “computer software” for “accounts receivable.” Unfortunately, not all users include all of the appropriate context information for such a search—the appropriate other tags are not included.

One approach to capturing information about context is to count the total number of tags applied. In Delicious.com, using the term “AccountsReceivable,” I found that the average number of tags was 4.66 and that the range was 1–16. Accordingly, the “contexts” applied by users of other tags vary substantially.

### **Mixing Words and Punctuation**

One of the more curious tag (errors) is when a tag word is combined with symbols or punctuation. As example, #accounting\_ and #\_accounting each had one entry in Twitter. As another example, the tag “accounting,” (i.e., “accountingcomma”) had 2,217 entries and “tax,” had 3,312 entries in Delicious.com. Similarly, the term “accounting;” (i.e., “accountingsemi-colon”) had 87 entries in Delicious. These examples likely are the equivalent of an input error, where the tagger mistakenly put a comma or semi-colon after entering the tag word. However, the unlikely combination must be used in order to get the information tagged.

### **Symbols**

In some cases symbols are used to tag materials. As an example, in Delicious.com “\*” had 61,433 entries, “#” had 13,618 entries, “^” had 481 entries, “!” had 34,331 entries, “-” had 239,912 entries and even “)” had 1,747 entries. Although “!” may indicate surprise, such a notion is clearly relative. Similarly, “?” may indicate questioning. However, in general, it is likely that there is no standard meaning associated with these symbols. Perhaps the symbols have been entered in error or maybe as a “marker” for the user to try to uniquely identify their tags. However, since such symbols can have over 200,000 items, there is clearly not much unique marking being done.

### **Impact of Medium**

In some cases the specific medium can influence the tags and the length of the tags. In the case of Twitter, since users are limited in the number of characters, financial abbreviations may not be

the common-usage abbreviations. For example, one hashtag that has been used in Twitter is “#Accountsrec.” Unfortunately, it is not clear that tag would be a common search term.

### Source of the Tagging Errors

In some cases taggers themselves will recognize that there are errors in the tags. In such instances where there is a consistent rationale for tagging errors, other taggers might point out the error source to others.

### Overcoming These Error Types

People can be very creative in generating different tags to represent the same concept, for example, “accounts receivable.” As a result, one approach would be to try to find the various tags used to capture a particular concept and include them in some intelligent system that would then use those multiple approaches to provide a single base concept. For example, such a system would take all of the examples in Table 1 and return the tags related to the concept of “accounts receivable.” However, to build such a system would require recognition of this issue and the alternative concept specifications.

### Summary

This section provides an empirical analysis of the accounting concept tag for “accounts receivable.” Application of the concept resulted in multiple approaches to represent those two words. In addition, the concept has a word that is difficult for some people to spell (receivable) and a word that is typically plural rather than singular (accounts). There also are abbreviations for the concept (A/R), some of which cannot be accommodated in different social media. When these and other concerns are put together, it is clear that not all object instances that appear to be intended to be tagged with this concept would be found with a single search. Instead, it appears that by choosing the most frequently used version, a user would find somewhere between roughly 70 percent (140/203) to 90 percent (1,006/1,126) of the (two tags) tagged material. However, if the user makes a spelling error or does not anticipate the most-often used version of the tag, then the odds of finding tagged material drop significantly. In addition, discretionary tagging creates the potential for untagged objects. As a result, it is questionable whether discretionary informal tags would be appropriate for critical applications.

## EMERGING RESEARCH ISSUE: TAGGING STRATEGIES

One potential area of future research is an analysis of discretionary tagging strategies. Analysis of this topic could take different approaches. For example, behavioral research could be used to investigate this topic. An alternative strategy is to empirically analyze tag data in order to infer tagging strategies. For example, in order to try to understand actual tagging strategies, resources can be examined. In so doing, the existence of apparent strategies and how they reuse tags promulgated by others can be analyzed. As an example, the financial article “What Would Happen if the U.S. Went Bankrupt?”<sup>8</sup> was analyzed. Information about the tags is summarized in Table 2.<sup>9</sup> The tag information represents tags from 79 different individuals.

<sup>8</sup> <http://www.bspcn.com/2011/06/07/what-would-happen-if-the-u-s-went-bankrupt/> (last accessed on December 1, 2010).

<sup>9</sup> Gathered from Delicious.com December 2010.

**TABLE 2**  
**Tags to Article “What Would Happen If the U.S. Went Bankrupt?”**

<u>Tag</u>	<u>Number</u>	<u>Percent</u>
Bankrupt	79	100.0
Bankruptcy	66	83.54
U.S._Bankruptcy	74	93.67
US_Bankrupt	36	45.57
Business	48	60.76
US	9	11.39
America	2	2.53
Finance	1	1.27
	315	

### Tag Reuse

One of the key issues in knowledge management is the notion of “reusing” knowledge (e.g., O’Leary 2014a). Analysis of tags suggests that there is evidence for tag reuse. In particular, it appears that as multiple taggers examine and tag an article that has already been tagged there is substantial tag reuse. In this example, although there was tag reuse, tags were not 100 percent redundant. For example, 93.67 percent of the tags on the article included the tag “U.S.\_Bankruptcy.” However, it is not clear whether information is added by simply repeating previous tags.

### General to Specific

Taggers exhibit behavior that suggests that they go from a concept, such as, “bankruptcy,” to specific instantiations of the concept, “U.S.\_Bankruptcy.” For the results in Table 2, only two taggers did not use both a representation of generic bankruptcy and the specific U.S. bankruptcy.

### Different Word for Same Thing—Synonyms

In Table 2, 41.5 percent of the taggers used both “U.S.\_Bankruptcy” and “US\_Bankruptcy.” In addition, 83.5 percent of taggers used both “Bankrupt” and “Bankruptcy.” This suggests that taggers exhibit behavior that indicates that they want others to be able to use their tags by using different words or representations for the same or similar concepts.

### Information Other than the Title

In this example, there is a very close relationship between the words in the title and the tags. “US” (“U.S.”) and “Bankrupt,” directly appear in the title. Five of the eight tags employ words from the title. Only the terms “Business,” “American” and “Finance” are really different. Accordingly, only 16 percent of the information in the tags is actually “different” information than the title. Further, “American” and “US” are highly correlated, while “Bankruptcy” would likely be viewed as an issue in “Business” and “Finance.” As a result, a potential research issue is to determine the extent to which tags are used beyond the title of the object.

### Building a System to Automatically Generate Tags

Examination of the title and the resulting tags suggests that it would be feasible to build a system that would analyze accounting and financial information and generate tags. Such a system

could employ domain knowledge or simply include key concepts in the title or abstract. For example, in the title “US” is the subject and thus it is likely that the subject of the object should be represented in the associated tags.

## EMERGING RESEARCH ISSUE: DEVELOPING TAXONOMIES AND ONTOLOGIES

Ontologies and taxonomies are important because they are used to facilitate search and reporting. In addition, ontology and taxonomy components reflect a conceptual understanding of the world that they describe. However, some researchers have argued that the formal development of taxonomies and ontologies does not scale well (Mathes 2004), as the amount of content available on the Internet and intranets has grown rapidly. In addition, expertise is expensive, and in some cases it is not clear which experts should be responsible for labeling, as emerging contents and existing technologies begin to merge and emerge. Furthermore, the velocity of change appears to be increasing as network effects involve a broader base of users, ultimately requiring rapid change in the ontologies and taxonomies. Finally, since in some domains, the knowledge is changing rapidly, there is interest in having those who actually do the work generate the appropriate taxonomies. Accordingly, researchers are increasingly interested in generating ontologies and taxonomies using tags (e.g., Limpens, Gandon, and Buffa 2008; Van Damme, Coenen, and Vandijck 2008; Barla and Bieliková 2009). Internally, the same needs and concerns have been recognized within companies. Thus, consulting firms have begun to explore the potential for evolving and generating ontologies and taxonomies using tags (Meister and Mark 2004; Burgelman and Blumenstein 2007).

### Evolving an Existing Taxonomy

The consulting firm Katzenbach (Burgelman and Blumenstein 2007) had planned to use tags to facilitate a “living” taxonomy. As new terms were used, tags would “eventually” be added to the formal taxonomy. That is, given a formal taxonomy, tags would be used to facilitate taxonomy evolution. Although their approach to evolution was not specified, evolution could take at least two different directions. First, if some set of tags begins to frequently appear, then it could be important to evolve the current system to include those tags. Second, as tags begin to decrease over time, it may become important for a taxonomy to begin to “forget” the information captured by those tags. However, there was no systematic approach proposed to analyze the tag data.

### Developing Hierarchical Relationships Using Tags

Taxonomies are distinguished by hierarchical relationships between items. As a result, if tags are to be hierarchically related, then a systematic approach is desired. It is assumed that there is a set of related tags that we wish to relate hierarchically. One tag-based approach is to choose those tags with the largest number of uses as higher in the hierarchy. For example, in Table 3, the number of Delicious tags given to “business” is much larger than any of the five functional areas listed below it, suggesting a “hierarchical” relationship between it and the other components.

### Using Entropy in Taxonomy Construction

Entropy, studied as part of information theory (e.g., Theil 1967), can be used to help build taxonomies using tag information. The formula for entropy is as follows: (Entropy)  $H = -\sum p_i * \ln(p_i)$ , where  $p_i$  is the relative frequency of the  $i$ th item. Lev (1968) and Theil (1967) refer to  $H$  as the “expected information” of this resulting discrete probability distribution.

The entropy function was first proposed by Shannon (1948) (e.g., Lev 1968). Subsequently, entropy has been used in a number of applications, in economics (e.g., Theil 1967), and accounting (e.g., Lev 1968; Lev and Theil 1978). However, although entropy plays a critical role in many

**TABLE 3**  
**Number of Tags**

<u>Discipline</u>	<u>Number of Delicious Tags</u>
Business	3,352,313
Accounting	73,076
Finance	1,041,089
Management	988,200
Marketing	2,374,375
Operations	14,227

machine-learning applications, e.g., [Bishop \(2006\)](#), and a number of researchers have used entropy-based approaches to investigate tagging ([Heymann, Ramage, and Garcia-Molina 2008](#); [Markines et al. 2009](#)), recently accounting and finance has largely ignored entropy.

There are some important characteristics of entropy. For example, if one of the  $p_i = 1$ , then  $H = 0$ , i.e., there is no information. Similarly, the maximum entropy for any set of  $k$  probabilities occurs with each  $p = 1/k$ .

To illustrate some of the potential uses of entropy with tags, consider the example in Table 3, rewritten below in Table 4 (Panel A) with tag frequency and corresponding entropy. Since entropy provides one measure of information, one approach would be to choose tagging systems that maximize the resulting combined entropy. Further, individual entropy measures can be used as

**TABLE 4**  
**Tag Frequency and Entropy**

**Panel A: Tag Frequency and Entropy**

<u>Categories</u>	<u>Tag</u>		
	<u>Frequency</u>	<u>p</u>	<u>Entropy</u>
Accounting	73,076	0.016	0.067
Finance	1,041,089	0.232	0.339
Management	988,200	0.220	0.333
Marketing	2,374,375	0.529	0.337
Operations	14,227	0.003	0.018
Total	4,490,967		1.094

**Panel B: Aggregated Tag Frequency and Entropy**

<u>Categories</u>	<u>Tag</u>		
	<u>Frequency</u>	<u>p</u>	<u>Entropy</u>
Finance and Accounting	1,114,165	0.24809	0.346
Management and Operations	1,002,427	0.22321	0.335
Marketing	2,374,375	0.52870	0.337
Total	4,490,967		1.018

cutoffs. Those tags below some cutoff could be merged into other categories. In addition, entropy can be used to formally assess the impact of information “loss” associated with operations associated with tags, such as aggregation of tags into taxonomy categories.

As an example, the maximum entropy of a five-item taxonomy ( $p_i = 0.20$ ) would be 1.609, which is roughly 47 percent larger than the entropy in Table 4 (Panel A). As another example, consider the entropy measure of information loss associated with aggregating tag information into different concepts in a taxonomy. In particular, compare the entropy in Table 4 (Panel A) and with Table 4 (Panel B) with the tags “Accounting” aggregated with “Finance and Accounting” and “Operations” with “Management and Operations.” This approach led to a decrease in entropy of 0.076.

Accordingly, entropy could be used to help construct taxonomies using tag counts. A taxonomy would require choosing between whether or not to include different items in the taxonomy. An analysis of the entropy of different configurations could be used to assess such an information loss. Other approaches derived from machine learning and previous accounting applications could be investigated.

### OTHER EMERGING RESEARCH ISSUES

There are a number of other emerging research issues with crowdsourcing tags in accounting and finance, including the following.

#### Why do People Tag Accounting and Financial Information?

There has been substantial analysis as to why people provide tags. Some researchers, (e.g., [Strohmaier, Körner, and Kern 2010](#)) have suggested there are two general reasons why people tag: to either categorize or describe resources. However, other researchers (e.g., [Ames and Naaman 2007](#)) appear to suggest that people tag for a range of reasons, some based on the particular domain. In addition, [Ames and Naaman \(2007\)](#) suggest that the incentives influence why the crowd tags. For example, in the case of photos, some tag to help others find the photo and to convey an opinion, as expressed in the photo. [Ames and Naaman \(2007\)](#) also generated a framework based on two dimensions: “sociality” (“Was the tag intended for others to use or for the originating individual?”) and “function” (“What was the tag’s intended use, e.g., organizational or communication?”). Unfortunately, there is limited research directly regarding why people might tag financial information or finance and accounting documents.

#### What Information Will People Tag?

It is an empirical question as to how to determine what accounting and financial information people actually tag in enterprise settings. From the analysis of Twitter and Delicious data it is apparent that people will discretionarily tag communications and knowledge contributions, such as articles. However, will people use discretionary tags on financial information, such as transactions and financial reports? If they do tag financial information, then what kinds of tags will they employ?

#### What Differences Are There with Enterprise Tagging?

At the time of this writing, there is substantial access to tag information in publicly available web and Internet Sources, such as Twitter, Flickr, and a range of other social media sites. However, there has not been access and analysis of tags used in enterprise sites: few (if any) enterprises have provided researchers their tags for detailed analysis. Access and analysis could be used to provide a better understanding of differences between tagging done in enterprises and tagging done in external sources (public and private) and the impact on the tags.

There are a number of potential research issues that might be generated from data derived from enterprise tagging systems, particularly in accounting and finance. For example, one of the concerns in accounting and finance is the usability of the financial information. One way to assess the actual understanding of users is to capture the tags that users place on different accounting and finance resources. Those tags could then be analyzed for both number and content in order to determine the views that users bring to accounting and financial resources. Further, the tags of different groups, for example, accounting and finance versus nonaccounting and finance could be compared to determine the difference in perspectives brought to the different resources.

In addition, enterprise tagging has a number of reasons for which the nature of tags is likely to be different than non-enterprise tagging, including organizational hierarchy and friendship. Organizational hierarchy provides an important basis for potential differences and hypotheses. Information and tags created by those higher in the organizational hierarchy, and identified as such, may be treated differently than those lower in the hierarchy. If the president of some organization provides information in some object, in general, then it is unlikely that negative comments will be made about that material. In addition, friendship within an organization is likely to influence the nature of tag comments. Information generated by friends in an organization is unlikely to generate negative comments by their colleagues. Other factors deriving from enterprise environments, such as competition between workers, also could drive additional differences between enterprise and non-enterprise environments.

### **How Can Tagging Facilitate the Big Data Lake?**

An emerging area for tagging is capturing information about databases that are embedded in the Big Data Lake (e.g., [O'Leary 2014b](#)). Firms are expected to put a number of different databases in the Big Data Lake, so that users can generate broad-based, real-time queries across original data. However, linkages between the different data sources and elements are not always clear. For example, data elements can have different names in different databases obscuring similarities between databases. As a result, as users investigate and determine the appropriateness of different linkages they can tag the variables in the different databases so that they capture relationships between data elements in the different databases. Such relationships can be helpful for those reusing the queries and for others also interested in analyzing similar or related relationships between the data in different databases in the enterprise's Big Data Lake.

### **How Frequently Are Tags Reused?**

One of the primary concerns of knowledge management is knowledge reuse (e.g., [O'Leary 2014a](#)). Similarly, if knowledge is to be indexed consistently over time, then a key concern is tagging that knowledge and the corresponding reuse of tags. This is likely a particularly important concern in accounting and finance as issues of comparison over time are paramount.

There is some evidence of reuse of tags in other domains. [Guy and Tonkin \(2006\)](#) found that only 10–15 percent of tags from Flickr and Delicious are single-use tags. Further, as time goes on, some of those single-use tags are likely to be reused. As a result, research finds that tags generally are used and reused. However, there is no direct evidence as to the extent of tag reuse in accounting and finance, other than the earlier discussions in this paper.

Finally, reuse and other forces may be pushing tag use toward formal taxonomies, as discussed above. For example, in the new version of Yammer, “Tags” have evolved into “Topics”: “Topics are a way to organize and categorize the content of your Yammer network. Anyone can add any topic to any thread, at any time—organizing your company’s knowledge on Yammer” ([Yammer Manual 2013](#)).

## SUMMARY, CONTRIBUTIONS, AND EXTENSIONS

This paper has investigated the use of informal and discretionary tags in accounting and finance to categorize knowledge about different objects including content or messages. In particular, this paper summarized eight of the key characteristics of tags, investigated some of the emergent difficulties of informal tags, reviewed some of the key accounting and finance discretionary tagging systems, and investigated a number of potential extensions and research issues.

### Contributions

Most of the previous research associated with using tags in accounting has to do with using required and formal XBRL tags. Unfortunately, there has been limited research of enterprise use of tags for accounting and finance, focusing on informal and discretionary crowdsourced tagging.

This paper analyzed some of the problems associated with discretionary and informal tags on accounting information. Based on the empirical analysis of an accounting concept "accounts receivable," it appears that informal tags identify the most frequently used version roughly 70 percent to 90 percent of the time. However, that does not include materials that were not tagged because the approach is discretionary. In addition, that assumes that the user is able to identify the most frequently used representation of the concept. Accordingly, it appears that discretionary and informal tagging may not provide the appropriate precision for some applications.

In addition, this paper analyzed the tags associated with a particular object. It was found that the tags often are repeated. Further, many of the tags could be predicted simply from the title, suggesting alternative automatic tags could be generated based on title information.

This paper also summarized some of the key characteristics associated with such tags, including whether the tags were formal or informal, required or discretionary, and six other dimensions. Further, this paper summarized some of the applications of tags to date in accounting and finance settings. Finally, this paper generated a number of emerging applications of tagging that could be addressed in future research.

### Extensions

This paper already has examined a number of emerging issues in detail; however, there are additional potential extensions. In addition to the emerging research issues discussed in this paper, there are a number of emerging research issues related to crowdsourcing tags in accounting and finance. At this point, there are perhaps more questions and uninvestigated areas than investigated concepts.

For example, this discussion raised a number of other potential research issues.

- Does game-ification of tagging influence the quality of the results?
- The above analysis in the section on "Searching Accounting Information: Limitations of Tags" indicated that there can be a number of problems associated with informal and discretionary tags. Accordingly, a key concern is "How can we build systems to try and disambiguate and correct tags generated by users?" (e.g., [Yeung, Gibbins, and Shadbolt 2007](#)).
- It is easy to imagine people generating millions and millions of tags over time. As a result, there may be a need to trace back tags over time (tagging responsibility systems). Another concern is "How do tags evolve over time? Is their evolution predictable?" ([Liu and Gruen 2008](#)).
- The world is non-stationary. Accordingly, it is not clear whether tags presented at some time  $t$  are still valid at some time  $t + k$ . For example, new technologies potentially could change tags put on a given object, say a news story about an emerging technology.

- An analysis of objects with multiple tags can help us analyze the relationship between those tags. For example, to what extent are tags applied independently of each other? Do users implicitly employ a hierarchical labeling approach?
- Why and how do people tag financial information? What are the incentives to tag accounting and financial information? Do incentives make a difference in the number and quality of tags?

In another line of research, system developers have built tag recommender systems that make recommendations, extending “page rank” (e.g., [Jaschke, Marinho, Hotho, Schmidt-Thieme, and Stumme 2007](#)). That research has resulted in a number of other systems, including “FolkRank” and “SocialPageRank.” More recent research has tried to capture context for the systems through other systems, such as “GroupMe!” ([Abel, Henze, and Krause 2010](#)).

In addition, researchers have begun to investigate other emerging issues, such as, “privacy threatening” implications (e.g., [Heidinger, Buchmann, Huber, Bohm, and Muller-Quade 2010](#)). For example, access to all tags that a particular user has accessed or generated may provide unique insights that ultimately can lead to a violation of his or her privacy.

Closely related to privacy are potential legal concerns associated with tags. If someone tags an object with a concern, then does it make response to that concern “more” of a legal obligation by that enterprise? For example, if a flaw is noted in a production process, and that processes is tagged recognizing that flaw, then does that mean the enterprise is legally responsible for issues related to that flaw?

Finally, as noted by [Bowen \(2007\)](#) “user tagging will lead to confusion because different people use different tags.” Thus, as noted by [Golder and Huberman \(2005\)](#), different taggers are likely to employ different levels of specificity. Accordingly, an important research issue is to what extent can this issue be mitigated through training, system support, or other approach? An approach referred to as “Lexitags” has been proposed (e.g., [Veres 2009](#)) to mitigate some of the potential ambiguity by presenting the user choices from WordNet, capturing synonymous word groups.

## REFERENCES

- Abel, F., N. Henze, and D. Krause. 2010. Optimizing search and ranking in folksonomy systems by exploiting context information. In *Web Information Systems and Technologies*, edited by J. Cordeiro, and J. Filipe, 113–127. Lisbon, Portugal: WEBIST, LNBIP.
- Ames, M., and M. Naaman. 2007. Why we tag: Motivations for annotation in mobile and online media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: ACM.
- Barla, M., and M. Bieliková. 2009. On deriving tagsonomies: Keyword relations coming from crowd. In *Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems*, edited by R. Kowalczyk, 309–320. Berlin-Heidelberg, Germany: Springer.
- Bhojraj, S., C. Lee, and D. K. Oler. 2003. What’s my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research* 41 (5): 745–774.
- Bishop, C. 2006. *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer.
- Bonsón, E., V. Cortijo, and T. Escobar. 2008. The role of XBRL in enhanced business reporting (EBR). *Journal of Emerging Technologies in Accounting* 5 (1): 161–173.
- Bowen, D. 2007. First person: “Folksonomy” takes power from expert librarians. *Financial Times* (November 7). Available at: <http://www.ft.com/cms/s/0/697db310-8c95-11dc-b887-0000779fd2ac.html#axzz3gunqjDti>
- Burgelman, R., and B. Blumenstein. 2007. Knowledge management at Katzenbach Partners LLC. Stanford Graduate School of Business, SM-162. *Harvard Business Review* (June).

- Fallon, J., and J. Timberlake. 2013. #Hashtag. Available at: <https://www.youtube.com/watch?v=57dzaMauXA>
- Glenn, D. 2012. Report: 62 percent of mobile user location-tag their posts and pictures. *SocialTimes*. Available at: [http://socialtimes.com/report-62-percent-of-mobile-users-location-tag-their-posts-and-pictures\\_b102128](http://socialtimes.com/report-62-percent-of-mobile-users-location-tag-their-posts-and-pictures_b102128)
- Golder, S., and B. Huberman. 2005. *The Structure of Collaborative Tagging Systems*. Available at: <http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf>
- Google.com. n.d. The benefits of auto-tagging. Available at: <https://support.google.com/analytics/answer/1733663?hl=en>
- Gruber, T. 2007. Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems* 3 (1): 1–11.
- Gupta, M., R. Li, Z. Yin, and J. Han. 2011. *An Overview of Social Tagging and Applications*. In *Social Network Data Analytics*, edited by C. C. Aggarwal. Berlin, Germany: Springer Science and Business Media.
- Guy, M., and E. Tonkin. 2006. Folksonomies. *D-Lib Magazine* 12 (1). Available at: <http://www.dlib.org/dlib/january06/guy/01guy.html>
- Heidinger, C., E. Buchmann, M. Huber, K. Bohm, and J. Muller-Quade. 2010. Privacy-aware folksonomies. Available at: [http://link.springer.com/chapter/10.1007%2F978-3-642-15464-5\\_17](http://link.springer.com/chapter/10.1007%2F978-3-642-15464-5_17)
- Heymann, P., D. Ramage, and H. Garcia-Molina. 2008. Social tag prediction. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY: ACM.
- Hotho, A., R. Jaschke, C. Schmitz, and G. Stumme. 2006. Information retrieval in folksonomies: Search and ranking. In *European Semantic Web Conference, Lecture Notes in Computer Science 4011*, edited by Y. Sure, and J. Domingue, 411–426. Berlin-Heidelberg, Germany: Springer-Verlag.
- Howe, J. 2006. The rise of crowdsourcing, *Wired* 14 (June). Available at: <http://archive.wired.com/wired/archive/14.06/crowds.html>
- Howlett, D. 2007. Workday financials: Accounting isn't real. *ZDNet.com* (August 20). Available at: <http://www.zdnet.com/article/workday-financials-accounting-isnt-real/>
- Jaschke, R., L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. 2007. Tag recommendations in folksonomies. In *Principles of Knowledge Discovery and Data Mining*, LNAI 4702, edited by J. N. Kok et al., 506–514. Berlin, Germany: Springer Verlag.
- Lev, B. 1968. The aggregation problem in financial statements: An informational approach. *Journal of Accounting Research* 6 (2): 247–261.
- Lev, B., and H. Theil. 1978. A maximum entropy approach to the choice of asset depreciation. *Journal of Accounting Research*: 286–293.
- Limpens, F., F. Gandon, and M. Buffa. 2008. Bridging ontologies and folksonomies to leverage knowledge sharing on the social web: A brief survey. *Automated Software Engineering Workshops, 23rd IEEE/ACM International Conference*.
- Liu, J., and D. M. Gruen. 2008. Between ontology and folksonomy: A study of collaborative implicit ontology evolution. In *Proceedings of the 13th International Conference on HeidingerIntelligent User Interfaces*, 361–364, Gran Canaria, Spain.
- Markines, B., C. Cattuto, F. Menczer, D. Benz, A. Hotho, and G. Stumme. 2009. Evaluating similarity measures for emergent semantics of social tagging. In *Proceedings of the 18th International Conference on World Wide Web*. New York, NY: ACM.
- Marlow, C., M. Naaman, D. Boyd, and M. Davis. 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. Odense, Denmark: *Hypertext 2006* (August 22–25). Available at: <http://pagesperso.lina.univ-nantes.fr/~prie-y/ens/06-07/M2R/articles/p31-marlow.pdf>
- Mathes, A. 2004. Folksonomies—Cooperative Classification and Communication through Shared Meta Data. Available at: <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Meister, D., and K. Mark. 2004. Hill & Knowlton: Knowledge management. Richard Ivey School of Business, 904E3. *Harvard Business Review* (February).

- Mejias, U. 2005. *Tag Literacy*. Blog, April 26. Available at: <http://blog.ulisesmejias.com/2005/04/26/tag-literacy/>
- Merriam-Webster. n.d. Definition of crowdsourcing. Available at: <http://www.merriam-webster.com/dictionary/crowdsourcing>
- Morville, P. 2005. Peter Morville: The tagsonomy interview. *You're It!* (blog). Available at: <http://tagsonomy.com/index.php/peter-morville-the-tagsonomy-interview/>
- Nittler, M. 2012. *Tales of the Cloud: The Story of Worktags*. Blog, November 5. Available at: [http://blogs.workday.com/Blog/story\\_of\\_worktags.html](http://blogs.workday.com/Blog/story_of_worktags.html)
- O'Leary, D. E. 2014a. Knowledge management: An empirical analysis of reuse and productivity. *Journal of Decision Systems* 23 (3): 249–265.
- O'Leary, D. E. 2014b. Embedding AI and crowdsourcing in the Big Data Lake. *IEEE Intelligent Systems* 29 (5).
- O'Reilly, T. 2005. What is Web 2.0? Design patterns and business models for the next generation of software. Blog, September 30. Available at: <http://www.oreilly.com/pub/a/web2/archive/what-is-web-2.0.html>
- Oracle.com. 2007. Workday vs. ERP, peach vs. nectarine? Oracle Financials Strategy (December 27). Available at: [https://blogs.oracle.com/financials/entry/workday\\_vs\\_erp\\_peach\\_vs\\_nectar](https://blogs.oracle.com/financials/entry/workday_vs_erp_peach_vs_nectar)
- Quintarelli, E. 2005. *Folksonomies: Power to the People*. Available at: <http://www.iskoi.org/doc/folksonomies.htm>
- Rainie, L. 2007. Tagging. *PewResearchCenter.org* (January 31). Available at: <http://pewresearch.org/pubs/402/tagging-play>
- Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* XXVII: 379–423.
- Strohmaier, M., C. Körner, and R. Kern. 2010. Why do users tag? Detecting users' motivation for tagging in social tagging systems. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. Menlo Park, CA: The AAAI Press.
- Theil, H. 1967. *Economics and Information Theory*. Chicago, IL, and Amsterdam, The Netherlands: Rand McNally and North Holland Publishing Co.
- University of Rochester. 2013. *UR Financials: FRS Crosswalk to the Workday Financial Data Model*. (April). Available at: [http://www.rochester.edu/adminfinance/urfinancials/wp-content/uploads/2013/05/UR\\_FRS-WorkdayFDM\\_FinancialPresentation-v2.pdf](http://www.rochester.edu/adminfinance/urfinancials/wp-content/uploads/2013/05/UR_FRS-WorkdayFDM_FinancialPresentation-v2.pdf)
- Van Damme, C., T. Coenen, and E. Vandijck. 2008. Turning a corporate folksonomy into a lightweight corporate ontology. In *Lecture Notes in Business Information Processing*, Vol. 7, 36–47. Berlin, Germany: Springer. Available at: [http://link.springer.com/chapter/10.1007%2F978-3-540-79396-0\\_4](http://link.springer.com/chapter/10.1007%2F978-3-540-79396-0_4)
- Veres, C. 2009. Crowdsourced semantics with semantic tagging: “Don't just tag it, Lexitag it!” *CEUR-WS.org*. Available at: <http://ceur-ws.org/Vol-1030/paper-09.pdf>
- von Ahn, L., and L. Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 319–326. New York, NY: ACM.
- Voss, J. 2012. *Linking Folksonomies to Knowledge Organization Systems*. *Metadata and Semantics Research*. 89–97. Berlin-Heidelberg, Germany: Springer. Available at: <http://www.jakobib.github.io/mtsr2012/paper.pdf> (last accessed on September 15, 2012).
- Wetzker, R., C. Zimmermann, and C. Bauckhage. 2008. Analyzing social bookmarking systems: A del.icio.us cookbook. In *Mining Social Data Workshop Proceedings, European Conference on Artificial Intelligence*.
- Wu, H., M. Zubair, and K. Maly. 2006. Harvesting social knowledge from folksonomies. Hypertext 2006, Odense, Denmark, August 22–25.
- Yammer. 2013. *Yammer Manual*. Available at: [http://www.dhs.vic.gov.au/\\_data/assets/pdf\\_file/0009/769950/Information-technology-Yammer-Guide-1-July-2013.pdf](http://www.dhs.vic.gov.au/_data/assets/pdf_file/0009/769950/Information-technology-Yammer-Guide-1-July-2013.pdf) (last accessed on July 15, 2013).
- Yeung, C., N. Gibbins, and N. Shadbolt. 2007. Tag meaning disambiguation through analysis of tripartite structure folksonomies. In *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops*, Silicon Valley, CA.