Term paper examples

Table of contents

These are examples of good term papers. That does not mean that I agree with every word or every formatting detail. There must be some flexibility

You chose which examples you want to look at or read.

EMBASE

Paper grade A

A thorough analysis and critique showing understanding of course concepts. More in-depth analysis of the index language would have improved the paper further.

Learning objectives demonstrated:

2.0.1^	571-L3.1 571-A5	Graduates know and understand the functional components of an information system.
2.0.1.1^	571-L3.1 571-A5	Graduates are able to use this framework to analyze and critique an information system, such as library.
2.3.1,1.4#	571-L4.2 571-A6	Graduates are able to construct a conceptual data schema for a given information system BT 2.3.4.2
2.3.4.1#	571-L2.2 571-L4.2	Graduates are able to analyze the knowledge structure of an existing information system and apply the results to • judging the adequacy of this schema with respect to the queries to be answered • using the knowledge of the schema to exploit fully the possibilities of obtaining answers from the information system RT 2.5.2.1
2.3.5	571	Graduates are able to apply schemes of organizing / representing / modeling of data/ information/ knowledge – metadata, cataloging, classification, ontology, and vocabulary standards – in cataloging, indexing, and metadata creation, and searching and to deploy automated and computer-assisted metadata creation tools.
2.5.2.1~	571	Graduates are able to conduct good searches that are responsive to user needs NT 2.3.2.1, RT 2.3.2, 2.3.4.1

The Inner Workings of the Embase Database: An Analysis

1. Introduction

The main purpose of this essay is to analyze Embase, a biomedical literature database, in order to better understand it's inner workings and expand my knowledge of medical databases. In the summer of this year I took LIS 586 Health Science Librarianship. We studied Medline as part of our course, as it was one of the database made available by Buffalo General Medical Center's medical library (the library we were studying in). We also covered MESH, the Medical Subject Headings, but only from a very basic and practical standpoint. That is we learned how it worked but not why it worked the way it does. After taking this class on organizing recorded information, where we dove deeper into the inner workings of how a database is designed and functions, I wanted to apply this knowledge to a medical database. However I also wanted to explore something that I hadn't before. It was suggested to me that I try Embase, a database that claims to have over six million more records and 2,500 more journals that is covered by Medline (Elsevier, 2013a). All too often we as researchers become complacent with the databases we are already familiar with and this was the perfect opportunity to break out of my comfort zone. It would also improve my overall skills as a librarian, as I would have a deeper understanding of a prominent database, an understanding that could easily translate over to other databases, medical or otherwise.

2. Those who use Embase and their needs.

A wide range of scientific and medical professionals uses Embase. However there are certain key professions who, at least according to the Elsevier website, are most likely to find this database useful (Elsevier, 2013b). These professions, while liked through similar fields all have unique needs that Embase claims to be able to fulfill.

2.1 Embase Users as Claimed by Embase

a. Biomedical and Clinical Researchers: these are professionals that are performing medical or pharmacological research. They need information on their specific topic of research; be it drug, disease, health, or past, current, or experimental treatment options, from as wide a variety of trusted sources as possible. This information will most likely be used for original research or a compliment to lab-based research (Elsevier, 2013b).
b. Information Managers: There are individuals who control the flow of information within and between offices, primarily those at private and public agencies which deal primarily with biomedical sciences. They need a database that will allow them to keep track of the development of in-house drugs and competitor drugs, access existing pharmacological research on the drug being developed, information to help manage a clinical trail database, meet the published literature requirements of regulatory agencies and most importantly, disseminate all that information throughout the organization (Elsevier, 2013b).
c. Medical Librarians: these are professionals whose goal is to help support medical decision-making and biomedical research. They need their database to contain a wide variety of high quality, up to date materials that are easy to navigate through for both themselves and their patrons. By easy to navigate, it means that the information must

d. be both quick and easy to find and comprehend (Elsevier, 2013b).
e. Regulatory Specialists: these are professionals whose goal is to ensure that the disease management procedures and pharmaceuticals that have or are being developed are safe and effective. They need to be able to track various studies on a single drug or procedure, compare the effectiveness of multiple drugs and procedures on a single illness, and collect information that they can present as evidence that the creators or distributors of a specific type of therapy is in compliance with current legislation and regulations (Elsevier, 2013b).

2.2 Why Such a Large User Base?

The fact that Embase is advertising that it caters to such a wide variety of clients on it's own website suggests that they are trying to be all things (or nearly all things) to all possible people in their field. While this might seems like spreading themselves a bit thin, it makes sense from a budget and collection development perspective. This term I took LIS 587 Collection Development with Rand Bellavia and one of the things we discussed was the high cost of journals and databases in libraries. Because of the ever-rising cost of journals and the increasingly common cuts to libraries budgets, many libraries are being forced to choose which databases to maintain and which to discontinue. Perhaps by having as large a range of journals and topics as possible, Elsevier is trying look like it has that greatest amount of materials per dollar and therefore has greater value than smaller or more specialized databases.

2.3 Why are Students not Mentioned?

The one thing I found particularly unusual about the large user base advertised by Embase is the fact that students seemed to be left out of their target audience. Now one could argue that students are included under the other professional titles or that Embase actually isn't trying to cater to college students and instead is trying to be a database just for professionals. In either case I think this is a mistake for one important reason, namely brand loyalty. When people begin any kind of research, they often start in databases with which they are already familiar. I myself am occasionally guilty of starting my research in databases that I am familiar with rather than seeking out the database most in line with my search subject. By not explicitly focusing on the current students in the biomedical field along with their professional counterparts, Embase runs the risk of their system being unfamiliar and less likely to be used or demanded by professionals in the future.

3 ER conceptual data schema

We learned about how to create Entity-Relationship conceptual data schemas back in lecture 2.2 (Soergel, 2013a). Although Embase is obviously a bit more complicated that than the recipes we did in class, the principles remain the same.

3.1 Entity Types and Relationship Types

Entity Type	Relationship types
Document	Document <isoriginal> Original or Non-Original
	Document <waswritteninlanguage> Language
	Document <haspublishingdate> Publishing Date
	Document <isrecordedin> Database
	Document <waspublishedin> Journal

	Document <hasmajorterm> Major Term
	Document <hasminorterm> Minor term
	Document <haschecktag> Human Study Type
	Document <haschecktag> Animal Study Type
	Document <haschecktag> Subject Age
	Document <haschecktag> Subject sex
	Document <haschecktag> Miscellaneous (systematic review, controlled study, & diagnostic test accuracy study)
	Document <haschecktag> Clinical trial
Clinical trail	Clinical <hasclinicalnumber> Clinical trial number
Original Document	Original Document <istype> Publication Type
Non-original Documents	Non-original Document <haschecktag> Meta analysis
	Non-original Document <haschecktag> Systemic analysis
Journal	Journal <hasprioritytranking> Priority Ranking
Index term	Index term <isclassified> Type of index term
Drug term	Drug term <ispartof> Drug group name
	Drug term <classifiesasother> Other drug type
	Drug term <hassubheading> Drug Subheading
	Drug term <isadministered> Route of Drug Administration
	Drug term <hasregistrynumber> CAS Registry Number
	Drug term <classifiesdrugas> Real drug
Real drug	Real Drug <ismadeby> Drug manufacturing company
Device term	Device term <hastradename> Device trade names
Device trade names	Device trade names <ismadeby> Device Manufacturers
Disease term	Disease term <hassubheading> Disease Subheading
General term	General term <isa> Non-chemical or drug term
Amino acid or nucleic acid	Amino acid or nucleic acid <hasmolecularnumber> Molecular sequencing number

4. Index Language

Embase uses Emtree or the Elsevier Life Science Thesaurus to index its articles. It contains over 60,000 biomedical preferred terms and 170,000 synonyms. These terms are divided into 15 facets. These facets are topic-specific taxonomies. Emtree is a document with multiple purposes. It works both as a key indexing aid, essentially a guideline that indexers use to determine the preferred terms to use when indexing and article or if a new term needs to be created and a search tool, where it acts as a visual representation of the database's language hierarchy, allows searchers to find preferred terms to build their queries from and a search function that allows users to browse rather than just search through the database's documents (further explored in section 6.5 *Browsing Emtree*) (Elsevier, 2012 January).

4.1 Issues with the largest facet in Emtree

The largest facet is dedicated to the topic of "Chemical and Drug", which accounts for about 50% of the preferred terms and 70% of the synonyms (Elsevier, 2012 January). The main

reason for this particular facet being so large and varied is that it includes terms that while technically fit into this category might not typically be thought of as drugs or just as chemicals such as endogenous compounds (compounds created naturally within an organism like endorphins) and environmental toxins (like PBCs leached from plastics). While this was probably done for the sake of simplicity on the part of the indexers, the potential definitions for this facet are so broad and varied it runs the risk of being useless for both the searcher and the indexer. We discussed this in lecture 8.2a, when we learned the importance of using clear language in the creation of a vocabulary and to handle the use of certain classes of words, like polysemy words, very carefully (Soergel, 2013c) terms like Chemical or Drug may not seem like they have multiple meanings in common language, as I learned from the biology classes I took my freshman year of undergrad, one could easily argue that just about everything in bioscience from drugs, to viruses, to human anatomy are essential just sets of chemicals and chemical reactions. Fortunately Embase resolves this potential confusion by giving clear instruction on exactly what terms are defined as chemicals and drugs in the Embase Indexing Guide of 2012 (Elsevier, 2012 January)

4.2 Index term categories

In addition to being filed under the 15 facets in Emtree, Embase index terms are assigned to one or more of eight specific categories within their documents (Elsevier, 2012 January):

- General terms- this is the catchall category for terms in all facets except for those filed under the topic of “Chemical and Drug”.
- Check tags- these are special kinds of general terms that fall into specific
- Drug terms- all chemicals and drugs, including endogenous compounds, laboratory chemicals, environmental toxins and therapeutic drugs. This category is further broken down into real drugs (drugs with clinical or potential clinical use), drug group names, and other drug terms (all other drugs that do not fall into the first two categories).
- Drug trade names and manufactures: all real drugs are indexed with the registered (legal) trade names if the trade name is mentioned in the article.
- Device trade names and manufacturers: all real drugs are indexed with the registered (legal) names if it’s manufacturer if the manufacturer is mentioned in the article
- Clinical trail numbers: the numbers under which a clinical trail is registered in one of three databases: ClinicalTrails.Gov, Current Controlled Trials, or the European Clinical Trails Database.
- Molecular sequence numbers: accession numbers under which nucleic acids acid or amino acid sequences are found
- CAS registry numbers: Chemical Abstract Service Registry Number generated for all drug terms and displayed with the drug name.

4.3 Subheadings

Subheadings are another layer of special classification that can be placed onto indexed terms. Subheadings are used as concept modifiers for drugs and diseases. They are matched to a term, typically a general term, within Emtree (Elsevier, 2012 January). So for example the drug term Acetaminophen could be modified with the drug subheading route of administration: buccal drug administration. In this way the subheading helps to clarify what kind of drug is being used in a document and how it is being used or tested. How these subheadings work to makes searches more effective will be discussed section 6.7 *Benefits and Drawbacks of Search Functions*.

4.4 Non-Emtree Index terms

While Embase does contain a large number of unique journals and articles, it also has approximately 3,000 journals that are also covered in MEDLINE and have already been index by them (Elsevier, 2012 January). Embase deals with this overlap by independently re-indexing each document using it's preferred terms in Emtree. Embase claims that all MeSH terms are included in Emtree as well as most of MeSH subheadings and MEDLINE's supplementary concepts. In the even a MeSH or MEDLINE term is only slightly different from it's Emtree counterpart the indexers will replace it with an appropriate Embase term or if the term is a numerical code (Clinical Trail number or Molecular Sequencing Number) the existing number in MEDLINE is used to generate a corresponding Embase code. In all these case there is some similarity or link between the MEDLINE term and the Emtree term, however there are cases when the MEDLINE term simple falls completely outside the scope of the Emtree language (Elsevier, 2012 January). These are referred to candidate terms and the exact process of their creation will be discussed in section 5.3, *Creating and Indexing a Candidate terms*.

4.5 Drug and Medical Device Trade Name and Manufacturers

Interestingly, there is one particular class of terms that Embase treats like a de facto candidate terms. These are the drug and medical device trade names and manufacturing companies. While Emtree does include generic names and general descriptions as a part of it hierarchy, terms that are classified as drug and medical device trade names or manufacturers names are largely left out of the Emtree hierarchy (Elsevier, 2012 January). This means that you can search for Tylenol in Embase, however the database will try to redirect you to more generic names for it like paracetamol.

4.5.1 Possible Reasons Trade Names are non-Emtree terms

It would seem that it would be easier to simple place common Drug and Device manufactures and trade names in Emtree to allow for searching by browsing or allow for straight searching by the name rather than having to use the generic name. However as I learned in the nursing classes I took in undergrad, the copyright laws and business practices of drug manufactures can make searching for drugs based only on their trade names or the name of their manufactures somewhat tricky (and I image this information translates easily to medical devices and the manufactures themselves). Some of the main difficulties in indexing trade names that I consider were:

A. Manufacturers can change– the trade name of who manufactures a given drug or device can change for a number of reasons including mergers, buyouts, and copyright lawsuits.
B. Multiple Manufactures can make the same drug or device over time- this can happen if a company obtains the rights to make a drug or device from another company or if one company makes the trade name of a drug and a separate company or division of that company makes a generic version of the same drug with a different name.
C. Drug & Device trade names change- again this can happen for a lot of reasons including mergers, buyouts, and particularly changes of copyrights to a specific drug or device.
D. Drugs & Devices can have multiple trade names- this is again can happen for many reasons such as lawsuits or copyrights changing hands, but one of the more common

ways this could happen is that after copyright on a trade name drug run out, a manufacture may create a generic version or alternative version with minor cosmetic changes in order to renew their copyright. The alternative or generic drug may have the same active chemical makeup, but small cosmetic changes and a different trade name.

In chapter 14.1 of Organizing Information, it was pointed out that a hierarchy does not exist merely for its own sake. It serves a number of specific functions of helping with indexing and query formation, assists in determining how general a term is, and allows for specific indexing and more general filing arrangements (Soergel, 1985b). If the Emtree hierarchy were to include such volatile and easily altered concepts as trade names, there would need to be near constant revisions and reworking's of its branches. While this is possible in Embase's current digital form, the constant and potential changes would make effective indexing and searching though Emtree's hierarchy difficulty if not impossible. Eventually indexers and searches alike would begin to distrust and eventually abandon use of the hierarchy, making the whole exercise of including trade names in the first place an exercise in futility.

5. Indexing Process and Parameters

Articles in Embase are indexed both manually and automatically, however manual indexing is by far the most preferred method and used for nearly all articles (Elsevier, 2012 January).

5.1 Manual Indexing

Professional indexers with backgrounds in biomedical science process the manually indexed articles. They do this by reading the full text of the article, identify the relevant concepts within the document and then index those relevant concepts according to the Emtree Thesaurus. The indexers also catalogue articles by important concepts that are left out of the Emtree Thesaurus. These left out concepts include drug and medical device trade names and manufacturer names (Elsevier, 2012 January).

5.2 Automatic Indexing

Only three types of document; conference abstracts, Articles in Press, and In-Process records, are index automatically using an algorithm. Only one of these document types, the conference abstracts, remaining permanently indexed in this manner with the other do being reevaluated and re-indexed by human indexers eventually. The algorithm categorizes these special types of articles based only on the text of the title and abstract. However the algorithm seems to be considerably more limited in its ability to completely index articles when compared to the manual indexing procedure. While the algorithm indexing process is able to catalogue articles using Emtree terms and is able to differentiate between major and minor terms, it is unable to index, subheadings, non-Emtree terms such as medical device, or candidate terms.

This limited indexing ability is probably why Articles in Press and In-Process records are only provisionally indexed in this fashion until human indexers are able to look them over and re-index them properly (Elsevier, 2012 January). While the most obvious reason for having an automated indexing process in the first place is to save time and have fewer indexers, one must question the actual benefits of such a practice if 2/3 of the documents indexed automatically are going to require revisions by human indexers anyway.

5.3 Creating and Indexing a Candidate term

Candidate terms are officially defined in Embase as terms that are proposed by indexers in order to cover "concepts discussed in articles that are not adequately covered by an existing Emtree term" (Elsevier, 2012 January). These indexers are also expected to create a more general or broad term to classify their newly proposed term under. So if there is a completely new type of drug is discovered, the indexer must include, in addition to the specific drug's name, the newly discover drug family and connect it back to the even broader term unclassified drugs. Now if this new candidate term does not appear in significant numbers in other documents, it will continue to be classified this way. However in the event that a candidate term shows signs that it is being indexed quite frequently, it will be reviewed for possible inclusion in Emtree (Elsevier, 2012 January).

5.4 Candidate Term Becoming a Preferred term

If a candidate term becomes a preferred term, meaning it is included within Emtree; a search is made for synonyms for the term. These synonyms may be preexisting within Emtree, but many of them will likely have already been indexed as candidate terms themselves. In the event the new preferred term is a drug term, a CAS Registry Number is assigned (when possible) and other candidate terms that may have been used to describe this drug in earlier works (ex: chemical name, trade name, generic name) are replaced by the newly made preferred term (Elsevier, 2012 January).

5.5. Benefits of Candidate terms

All of this help to make Emtree a living index language that is constantly evolving and updating itself along with all its documents. This keeps the database's index plastic enough to allow for new discoveries and terms to be added or changed as needed. Such plasticity is especially important for a biomedical database like Embase because it is a rapidly expanding field and new terms are being assigned to old concepts (or newly discovered concepts) frequently in an attempt in standardize the language. The one possible negative of the Emtree Thesaurus for users is because Embase indexes all of it's documents base on it's own unique language, it is possible that a person who has grown accustomed to searching on MEDLINE or another database might not be able to use their familiar search term or strategies

5.6 Indexing terms as Major or Minor

Once a candidate term is created or a term within a document with an existing Emtree counterpart is mapped by either the human indexers or the algorithm, the indexed term is given the additional quality of being a major or minor term in relation to its article. A major term is an index term that represents the main focus or core message of an article (ex: for a paper about the effects of diabetes medication, diabetes is a major term). A minor term is the opposite, an indexed term that represents an important term within the document but does not pertain to the main focus of the document (ex: In the same paper on diabetes, any mentions of other types of non-pharmaceutical therapies that are recommended for future study are considered minor terms). There are typically only three to four major terms per document, but there can be up to fifty possible minor terms attached to each document (Elsevier, 2012 January).

5.7 Benefits and Drawbacks of Major and Minor Indexing

This method of indexing has its benefits and drawbacks. The most obvious benefit is that it creates a built in weighting system for each documents index terms. While it is not identical to the numerical weighting system described in chapter 16 it works in much the same way (Soergel, 1985c). By indexing document terms with the quality of being a major or minor term, searchers are able to form more specific queries. They can clearly state if they want documents where their search terms are the main focus of the document, which ideally would make their searches more discriminating and the results more relevant. If course whether this actually happens in reality depends on a lot of other factors within the Embase system, such as how well the search is able to form the query and how well the documents are indexed (Soergel, 1985c).

This drawback common to all weighting systems, namely that it depends heavily on proper indexing, is particularly troublesome in this system of indexing terms major or minor. Due to the fact that this weighting system is binary, rather than a sliding scale, the decision of what is and isn't the focus is more of a qualitative than quantitative. Qualitative decisions are difficult to make and sometimes even harder to justify as inevitably it depends at some point on the well informed but still arbitrary opinion of the indexer. In addition, because each document is limited to a maximum of four major terms, those documents who have more that four major points, however rare they may be, will be forced to either drop one of their major terms into one of the minor slots or the indexer may be forced to try to cram two concepts into one term. Both of these might lead to problems in indexing to document as a whole and in being able to retrieve it effectively.

6. Searching Embase

Embase's primary search is divided into eight main functions, five of which are search function: Quick, Advanced, Drug, Disease and Article, and three of which are browsing functions: Emtree, Journal and Author. Each search function is slightly different with it's own unique purpose.

6.1 Quick Search

Quick search is the starting home page for the database and is a simple Boolean search box. The example given below the search box suggests that this search function is meant for searching single terms, even through more complex Boolean phrases can be searched using it. There is also a single check box call extensive search that allows for three simultaneous functions. These functions are mapping, where the search term is automatically mapped or matched to the Emtree preferred term and the search is done on all index fields; explosion where narrower (but oddly enough not broader) terms are searched, and keyword search, where all the words and phrases are searched in every field. The Quick Search also has an autocomplete that offers suggestions of terms found in the database's index. This is supposable to help the searcher find the best terminology for their search (Elsevier, 2012d). While that is not untrue, another possible reason for the autocomplete is to strongly encourage users to used Embase's preferred terms, or at the very least common synonyms so the at the user doesn't get too frustrated by constantly getting empty searches due to poor spelling or use of common slang terms. The obvious purpose of this search function is for simple, one word or phrase searches.

6.2 Advanced Search

Advanced Search is much like the Quick Search but offers the user the ability to fine-tune his or her search in a way that the Quick Search does not. Advanced search has Boolean search box that is about the size of a small paragraph and the examples given underneath it show queries with Boolean connects rather than a single term. This seems to suggest that Embase it encouraging their users to use the Advanced Search page for more complex Boolean searches. This larger search box, like it's smaller Quick Search counterpart, also allows for searching in multiple fields at once (this is essentially the keyword search function) but unlike the Quick Search keyword function it is possible for the searcher to query for terms only in one field or a chosen set of fields at a time. For example the searcher can enter the name of a country and query under Country of author and Conference location, without also turning up articles with the country's name in the title or the index.

In addition to allowing for a more refined keyword search, Advanced search also has checkboxes, which allow the searcher to choose to query their terms with only mapping or explosion search rather than being forced to do both at the same time. It also allows for searching by synonyms, by major terms (which was explained in section 5.6, *Indexing Terms as Major or Minor*) and allows for what Embase refers to as free text searching. Free text search, according to Embase, allows the user to search their query in all index fields while also allowing for searching in all child term related to the original search term (Elsevier, 2012a). A searcher can further refine their query by other limits, such as publishing date range, areas of focus, age groups of test subjects and so on, that work in much the same way the checkboxes.

6.3 Disease and Drug Search

The Disease and Drug search functions are like a combination of the Quick and the Advanced Search option, both having the narrow search box of the Quick Search while having largely the same type of limiter checkboxes that Advanced Search had. The main differences however is that unlike the Advanced Search their isn't an option to restrict the search to a single field and it has additional specialized limiters unique to their specific topics like Drug subheadings, Routes of Administration and Disease Subheadings.

6.4 Article Search

The Article search function allows for a search of specific articles, rather than a specific topic. Rather than having a single general search box, it has multiple boxes already pre-designated for index terms most likely to be used to find a specific article such as author name, journal title, abbreviated journal title, ISSN, CODEN, volume issue and first page. It also like the other search functions allows for searching for documents by a range of years of publication. The only other somewhat unique aspect to this search function is that there is a check box that allows the searcher to choose to look for exact or not exact journal titles or abbreviated journal titles. This is useful as it's not uncommon for searcher to misremember the exact name of a journal they found a particular article in.

6.5 Browsing Emtree

In addition to direct searches, Embase also allows a person to browse rather than directly search using the Emtree vocabulary hierarchy tree. This is essentially a way of using a digital version of the Emtree thesaurus to find exact terms in the database's controlled vocabulary hierarchy. This search function has a distinct advantage over the direct search options mentioned

above as it allows for serendipity. By serendipity, I mean that if one is not exactly sure what terms or even what subject to look for, browsing through Emtree will give the searcher the opportunity to stumble across the appropriate term. This is an opportunity that is often lost in common search engines like Google. However you cannot use it to search for trade names of devices, drugs or manufacturers. The possible reasons for this I already discussed in section 4.5.1 *Possible Reasons Trade Names are non-Emtree*.

6.6 Browse by Journal or Author

Browsing by Journal or Author is a somewhat unstructured experience. The Journal browsing function is simply a list of journals organized alphabetically by title. Under each title you can browse by individual volumes, issues and tables of contents (Elsevier, 2012c). Before taking Organizing Information I might have consider this to be a meaningful form of organization, but as I learned in. Alphabetical vs. meaningful sequences in Lecture 5.2a, this isn't the case (Soergel, 2013b). By organizing journals only by their spelling, Embase makes it considerably harder to find a particular journal, especially if you don't know the exact journal you are looking for, than if they organized them by, for example, main topics covered in them.

Browsing by Author works in much the same way as Journal except that you have to type in a name or initial into a search box before you are given a list of names, again organized alphabetically. This in some ways is even worse than the Journal browsing because at least by reading through the list of alphabetized journals, there is a chance something might catch you eye or jog your memory if you don't remember the exact name of the journal. In the Author Browsing however you absolutely need to remember at least a part of the authors name in order to find it (Elsevier, 2012b), which completely removes the advantage of serendipity that I mentioned earlier with Emtree.

6.7 Benefits and Drawbacks of Search Functions

All of these different search functions and the elements within them work together in order to try to increase the relevance of the searcher's queries without having to sacrifice recall. It's common wisdom that was precision and recall are inversely related. But as we learned in Chapter 8.1 Answer Quality, that is not always the case in reality. In fact in reality if a query and a system are properly matched it is possible to have high relevance without sacrificing high recall (Soergel, 1985a*)* It is possible that by having six different types of search functions and having a wide variety of easy to check limiters, Embase is encouraging searchers to input as much information as possible through checkboxes, subheadings and so on (to increase relevance) while at the same time indexing as many terms and synonyms as possible, even if they don't fit in the Emtree hierarchy and having as many documents available as possible, even if there are already indexed in another database (to increase recall). It's almost like a birdshot method of reaching the relevance and recall search ideal; just provide every possible item and every possible option for finding it and inevitably the user will find something that fits their needs.

The obvious downside of having so many search functions is that having too much variety in your search functions can lead to confusion and choice paralysis on the part of your users. Having to learn how to properly use six different search functions and then having to decide which one will be most effective at any given time is quite a challenge even for professional in the biomedical field. A challenge that many searchers may forgo in the name of time and convenience, causing them to just automatically select the Quick Search as it is both the simplest to understand and the first search page. Embase tries to overcome some of this

confusion by offering webinars in how to use the database, but again this requires time and commitment that many database users may not have the time or patience for.

7. Final Assessment

After look in detail at the Embase database, does it live up to its claim of being all things to all likely users? Well it certainly makes a great effort. The Emtree Thesaurus is very clean and easy to use, the autocomplete in the search functions is a great help for correcting spelling errors or ill chosen words, and Embase is nothing if not thorough in its indexing. I particularly like the weighting of indexed terms to allow a searcher to clarify if their query terms are meant to be the main focus of the documents they are search for or not.

The one main fault however I find in this database is that it has almost a bit too much variety in its search functions. While it is true that a person can just use the Quick search function to find their articles, the design of Embase's other search functions suggest that Quick search will not always provide the most relevant or largest pool of results that Advanced search or Drug search might make available. This combined with the fact that some index terms are indexed and showing through the autocomplete, but are not made available in Emtree

Since I spend time to learn and get to know this database, I can say with mild confidence that I would be able to navigate through it and use it effectively, but I am not sure if the same can be said for the average user. I think that Embase, like so many other professionals, sometimes forgets that just because a person is an expert in pharmacology, pediatrics or any other kind of bioscience that does not necessarily mean they are experts at searching for their specialized topics. While it is true that Embase openly admits that it is designed for certain individuals who should have taken the initiative and participated in the available webinar series, like medical librarians and information managers, I wonder if researchers and regulatory specialists would do the same.

Overall I would say that this was an enriching experience. While I cannot not say with absolute Embase is objectively better than Medline, my analysis of it did make me more able and more likely to make use of this database in the future By looking deeply into Embase as a database a learned a great deal not just about the database itself but the philosophies and business practices that lay under them. I recognize that I might not be able to do with every new database that I come across in my life as a librarian due to time constraints. However it did encourage meet to seek out trains in databases that I am unfamiliar with and not only use the ones with which I have become comfortable and familiar. To be in a sense adventurous in my research methods.

Work Cited

Soergel, D. (2013a). *Lecture 2.2 Knowledge representations*. Personal Collection of Professor Dagobert Soergel, University at Buffalo, Buffalo, New York.

Soergel, D. (2013b). *Lecture 5.2 Document Design (information design) for people*. Personal Collection of Professor Dagobert Soergel, University at Buffalo, Buffalo, New York.

Soergel, D. (2013c). *Lecture 8.2a Vocabulary Control*. Personal Collection of Professor Dagobert Soergel, University at Buffalo, Buffalo, New York.

Soergel, D. (1985a). *Organizing information: Principles of data base and retrieval systems*. (pp. 120-122). Orlando, Florida: Academic Press Inc.

Soergel, D. (1985b). *Organizing information: Principles of data base and retrieval systems*. (p. 252). Orlando, Florida: Academic Press Inc.

Soergel, D. (1985c). *Organizing information: Principles of data base and retrieval systems*. (p. 337). Orlando, Florida: Academic Press Inc.

Elsevier, B. V. (2013a). *About embase - boost your biomedical research*. Retrieved from http://www.elsevier.com/online-tools/embase/about

Elsevier, B. V. (2013b). *Embase: From pharmacovigilance to evidence based medicine*. Retrieved from http://www.elsevier.com/online-tools/embase/who-uses-embase

Elsevier, B. V. (2012, January). *Embase indexing guide 2012: A comprehensive guide to embase indexing policy*. Retrieved from http://cdn.elsevier.com/assets/pdf_file/0009/126873/Embase-indexing-guide-2012.pdf

Elsevier, B. V. (2012a). *Embase medical answers: Also search as free text*. Retrieved from http://www.embase.com.gate.lib.buffalo.edu/info/helpfiles/search-forms/advanced-search/also-search-as-free-text

Elsevier, B. V. (2012b). *Embase medical answers: Authors*. Retrieved from http://www.embase.com.gate.lib.buffalo.edu/info/helpfiles/search-forms/authors

Elsevier, B. V. (2012c). *Embase medical answers: Journals*. Retrieved from http://www.embase.com.gate.lib.buffalo.edu/info/helpfiles/search-forms/journals

Elsevier, B. V. (2012d). *Embase medical answers: Quick*. Retrieved from http://www.embase.com.gate.lib.buffalo.edu/info/helpfiles/search-forms/quick-search

An excellent description and analysis of this system showing mastery of course concepts.

Growing a Garden Database: an Analysis of the Collaboratively Developed Plant and Vendor Databases on davesgarden.com

Comment [d1]: Purpose sentence is missing. Should be here even though the title is informative

1 Analysis of user needs

►

Dave's Garden is an online community of gardeners whose users access and share information about plants and gardening. Two features are examined here:

1 PlantFiles, a collaboratively developed plant database, and
2 The Garden WatchDog, a collaboratively developed vendor database.

The site proclaims that it is by gardeners, for gardeners. So what do gardeners want? Who are these gardeners?

This site is aimed primarily at gardening hobbyists. Certainly, nursery professionals and other experts such as botanists are represented among the user base, but its core mission is to serve the non-professional. They are predominantly middle-aged (35-55) and middle class (half have a household income over $75,000), and female. The vast majority of the site's users live in the United States. They are computer literate enough to turn to a website for information, but they are most likely not expert searchers. A core of power users, acknowledged as PlantFiles pioneers and Uber-Gardeners, are responsible for most of the data contained in the PlantFiles database.

This study of user needs is primarily gleaned from the present users (a close examination of the various forums uncovers many of the issues that preoccupy gardeners), and my own experience as a gardener, mail-order nursery customer, and Dave's Garden user. The following is a list of plant and vendor information required/desired by gardeners. The unifying principle is information that helps the gardener with the tasks of plant evaluation and/or sourcing.

Note: Highlighted entries are not included in the databases:

1.1 Needs

Comment [d2]: These could be arranged in a more meaningful order/

- Plant identification—what is the scientific/common name for this plant? (aids in purchasing the correct plant, and locating other cultivars in the same class—also, the multiplicity of common names can introduce confusion).
- Plant suitability—what plants will thrive in my garden's location and conditions?
- Plant's appearance attributes—what color flowers, foliage, how tall, etc.
- Plant's timing attributes—when does it bloom, when does it break dormancy, will it go dormant, does it have winter interest—this aids in choosing plants that will coordinate with each other
- Plant's growing requirements—sun exposure, soil conditions, space requirements, water requirements (xeriscaping is a hot topic in gardening now)
- Plant's other attributes: fragrance
- Plant's type-- annual, biennial, perennial, bulb, tree, shrub, grass
- Plant's use in landscaping--ground cover, alpine and rock gardens, ponds/aquatics
- Plant's use after harvesting: edible, suitable for cut flower, suitable for drying, medicinal etc
- Plant's impact on the environment—does it attract bees and butterflies? Is it a native or an exotic? Is it invasive?

- Plant's ease/method of propagation—from seed? Self-sow? Require division? How often?
- Plant maintenance—pruning, fertilization, deadheading, special needs (for example—managing the color of hydrangea)
- Is the plant dangerous? If ingested? If touched? To pets?
- Is the plant deer resistant?
- How long until harvest?
- Is the plant susceptible to particular diseases or pests?—actually, it is in there for certain genera, but not on the top level of the Advanced Search
- What plants are suitable for growing in containers?
- Where can I buy this plant?
- Is this a reputable nursery?
- Which nurseries specialize in this plant?
- What will I receive (plant start, bare root, bulb, seeds)?
- Where can I find plants grown organically?
- Plant's patent status

1.2 Wants

- Plant combinations: what plants look well together? What plants grow well together?—there is currently the ability to link to a picture of the plant in a garden
- Plant ratings by other users, (with their location)—recommended or not
- Plant reviews by other users, (with their location)—particulars about how the plant performed in their garden
- Plant specifications and plans for predesigned gardens suitable for garden conditions
- Vendors that offer plans and plants for predesigned gardens
- Ratings and reviews of those predesigned gardens

Comment [d3]: How are these different from needs? One meaningful sequence would be better.

2 Entity-relationship (ER) conceptual data schema.

►

2.1 PlantFiles database Raw Entities and Relationships:

Entity Types, Raw	Relationship Types, Raw
Plant (Note: Plants are listed by Common Name)	Plant <isClassifiedUnder> Family
Category (annuals, bulbs, biennials, perennials, herbs, vegetables, ground covers, shrubs, etc. – categories commonly used by nurseries)	Plant < isClassifiedUnder> Genus
Genus	Plant < isClassifiedAs> Species
Species	Plant < isClassifiedAs> Cultivar
Family	Plant <hasHybridizer> Person
Cultivar	Plant <hasHybridizer> Company
Height	Plant <isInCategory> Category
Spacing	Plant <hasAttribute> Height
Hardiness	Plant < hasRequirement> Spacing
Sun Exposure	Plant <hasAttribute> Hardiness
Danger	Plant <hasRequirement> Sun Exposure
Bloom Color	Plant <hasAttribute> Danger
Bloom Time	Plant <hasAttribute> Bloom Color
Foliage	Plant <hasAttribute> Bloom Time
Other details	Plant <hasAttribute> Foliage
Soil pH requirements	Plant <hasAttribute> Other Details
Patent Information	Plant <hasRequirement> Soil pH requirements
Propagation Methods	Plant <hasAttribute> Patent Information
Seed Collecting	Plant <hasMethod> Propagation Method
Person	Plant <hasMethod> Seed Collecting
Family	Genus <isClassifiedUnder> Family
Genus	Species <isClassifiedUnder> Genus
Species	Cultivar <isClassifiedUnder> Species
Cultivar	Plant <hasRating> Rating
Rating	Plant <hasReview> Document
Review (Document)	LegalEntity <givesRating> Rating
DaysAmount	LegalEntity <givesReview> Review
Image	LegalEntity <postsReview/Rating> Date
	Plant <hasAttribute> Maturity
	Plant <reachesMaturityIn> DaysAmount
	Plant <hasPhoto> Photo
	LegalEntity <postsImage> Image

Formatted: Space Before: 4 pt

Comment [d4]: One entity type Taxon

Comment [d6]: Taxon *<isClassifiedUnder>* Taxon

Comment [d5]: Repeated

2.2 PlantFiles database Final Entities and Relationships:

Entity Types, Final	**Relationship Types, Final**
Plant (common name) Category (annuals, bulbs, biennials, edible fruits and nuts, vegetable, etc.) Classification (family, genus, species, cultivar) Attribute Requirement Method LegalEntity Person Organization Rating Document Date DaysAmount Image	Plant <*isa*> Category Plant <*isClassifiedAs*> Classification Plant <*hasAttribute*> Attribute Plant <*hasRequirement*> Requirement Plant <*hasMethod*> Method Plant <*hasHybridizer*> LegalEntity Plant <*hasRating*> Rating Plant <*hasReview*> Document Plant <*reachesMaturityIn*> DaysAmount LegalEntity <*givesRating*> Rating LegalEntity <*givesReview*> Document LegalEntity <*postsReview/Rating*> Date Plant <*hasPhoto*> Photo LegalEntity <*postsImage*> Image

2.3 The Garden Watchdog Vendor database Raw Entities and Relationships:

Entity Types, Raw	**Relationship Types, Raw**
LegalEntity (Vendor, user) Mailing Address (Location) Phone Fax Email Website Facebook page Twitter Offering (seed, bulb, plant) Company Comment (Document) Review (Document) Rating (Positive, Neutral, Negative) Date PlantScout **From Plant Database:** :Plant :Category :Other Details **From Garden Products Database:** :Garden Product **From Garden Bookworm Database:** :Garden Books	LegalEntity <hasMailingAddress> Mailing Address LegalEntity <hasPhone> Phone LegalEntity <hasFax> Fax LegalEntity <hasEmail> Email LegalEntity <hasWebsite> Website LegalEntity <hasFBpage> FB page LegalEntity <hasTwitter> Twitter LegalEntity <hasSpecialty> :Plant LegalEntity <hasSpecialty> :Category LegalEntity <hasSpecialty> :Plant <hasAttribute> Other Details LegalEntity <hasSpecialty> :GardenProduct LegalEntity <hasSpecialty> :GardenBookProduct LegalEntity <isRankedTop30> Watchdog30 LegalEntity <hasOffering> Offering LegalEntity <hasCompanyComment> Document Legal Entity <hasRating> Rating LegalEntity <hasReview> Document LegalEntity <hasNumberOfProductsIn> :PlantScout LegalEntity <writesComment> Comment LegalEntity <givesRating> Rating LegalEntity <writesReview> Review LegalEntity <postsRating/Review> Date LegalEntity <updatesRating/Review> Date

2.4 The Garden Watchdog Vendor database Final Entities and Relationships:

Entity Types, Final	**Relationship Types, Final**
LegalEntity Person Organization Location Phone Fax Email Website SocialMediaAccount Facebook Page Twitter, etc. Document Rating Date :PlantScout Product	LegalEntity <*hasLocation*> Location LegalEntity <*hasPhone*> Phone LegalEntity <*hasFax*> Fax LegalEntity <*hasEmail*> Email LegalEntity <*hasWebsite*> Website LegalEntity <*hasSocialMediaAccount*> SocialMediaAccount LegalEntity <*hasOffering*> Offering LegalEntity <*authoredCompanyComment*> Document LegalEntity <*authoredReview*> Document LegalEntity <*hasRating*> Rating LegalEntity <*gaveRating*> Rating LegalEntity <*postedRating/Review*> Date LegalEntity <*hasNumberOfProductsListed*> :PlantScout LegalEntity <*hasSpecialty*> Product LegalEntity <*isRankedTop30Watchdog 30*> LegalEntity

2.5 Comparison between User Needs Analysis and Conceptual Data Schema.

A comparison between the user analysis and the conceptual data schema uncovers only a few types of data that are currently missing but might warrant inclusion. Plant care and maintenance is obviously important to gardeners, and knowledge of the degree of maintenance a plant requires certainly affects purchasing decisions. The current lack of this information is presently offset by the gardener's ratings and reviews of a particular plant; gardeners very often share information about maintenance in their reviews. That said, it is unlikely that was a consideration in this omission. Plant maintenance fits into the current conceptual schema under Plant <hasMethod> Method. The current PlantFiles database includes deer resistance (under Other Details), in the top level of the Advanced Search, but omits information about susceptibility to other pests and diseases. This information could be included in a new entity, Pest/Disease, that could include deer, other animals, and insects. Deeper searching reveals that pest and disease resistance is included under Other Details in entries pertaining to particular plant groups—I found that information included with rose cultivars. It appears that Other Details is a catchall for stray concepts but also for categories that were not included in the original design. For example, four aspects of water requirements are included in Other Details; conceptually, these fit into Plant <hasRequirement> Requirement and could be grouped together as Water

Requirement. The current PlantFiles database is oriented to the plant as a single specimen; each plant is considered in isolation. Gardening is all about combining plants, so adding recommended plants to combine with a specified plant would be a useful feature. Given the collaborative nature of the database, gardeners could easily update existing plant records to include their recommendations; they currently have the ability to add a link to a picture of the plant in a garden. Many nurseries offer pre-planned collections of plants; these could be specified in the Products entity in The Garden Watchdog, and users could have the ability to rate and review these pre-planned gardens.

3 The index language(s) used and its (their) suitability for the needs of the intended audience.

Dave's Garden uses free-text searching. As noted in the user needs analysis above, the average user is middle-class, a computer user but not a computer expert, and between the ages of thirty-five and fifty-five. They are most likely already familiar with free-text searching, but much less likely to be familiar with Boolean operators.

Comment [d7]: Plants should all be indexed and their scientific name and found

The limitations of free-text searching are addressed in the following ways:

- In the Advanced Search option, the user interface imposes a degree of terminological control, essentially presenting a graphically displayed entity-oriented index that allows the users to select descriptors (search keys) to formulate a query. The absence of cross-references has a negative effect on recall. (more on this in sections 4 Indexing Process and 5 Searching).
- The website offers a search tutorial to teach the user how to search for the information they need.
- Instructions on the Specific Search screen remind the user that spelling and punctuation matter, and that partial words are acceptable search input
- In the Specific Search option, users are restricted to entering data into a particular (appropriate) field, for a field-restricted free-text search. (more on this in section 5 Searching).
- One enterprising user in the PlantFiles How-to forum recommends typing "Asclepias tuberosa Dave's Garden" (for example) into Google to find a particular plant in the PlantFiles database. This takes advantage of Google's ability to offer relevant spelling and content (different cultivars, for example) alternatives.

The General Search is a frustrating and inefficient way to search for plants by characteristic. Precision—the measure of relevance and discrimination against the data collection--is woefully low. There are just too many relevant entries missed, and too many irrelevant entries returned. There is a reason Dave's Garden is able to charge a subscription premium for the Advanced Search.

4 Indexing process

The conceptual data schema determines what data will be entered into the database. To add a new plant to the database, the user is asked to supply the information associated with classification: common name, family, genus, species and cultivar. He may also include additional cultivar names, hybridizer, year of registration, grex name (for orchids only) and clonal name (for orchids only). Once the new entry is uploaded, the user is presented with a checklist of descriptors—the entities that relate to the plant's attributes, requirements, and methods listed in the conceptual schema. The user who creates or edits an entry checks off all entities, or aspects of entities, that apply. The Advanced Search takes advantage of the classificatory structure by treating entities/aspects as descriptors, resulting in a search engine with vastly improved precision than the General Search.

Dave's Garden relies on volunteer editors (much like Wikipedia) to oversee the accuracy of the data. Users are provided a link to notify the site of any errors that they spot.

Comment [d8]: So there is a controlled vocabulary

5 Searching

►

Plant identification, acquisition and care are the drivers for most of the searches done on PlantFiles and The Garden Watchdog. There are three ways to search the PlantFiles database; General Search, Specific Search and Advanced Search. The Specific Search and Advanced Search rely on the user interface to guide input; the General Search offers no such guidance. As one would guess, the performance of the guided searches far outperforms the general search in terms of returning relevant results.

The Specific Search allows the user to search for a plant by its scientific or common name. Users may enter information into the following fields: common name, family, genus, species, cultivar, hybridizer, orchid grex or orchid clonal name. Limiters regarding cultivars (show with cultivar only; show without cultivar only), and pictures (show entries with pictures only; show entries without pictures only) are options, as is the ability to sort by popularity, Latin name or cultivar. There is a reminder to use correct spelling and punctuation (free-text); and the user is advised that partial words or names are fine.

The Advanced Search treats the entities detailed in the conceptual schema as descriptors in an entity-oriented index. Scrollable fields containing the aspects of these entities are presented to the user. The user may choose from any number of fields to narrow his search; he may also choose to add multiple aspects from within a field. (For example, Find a perennial that is 18"-24" tall, USDA hardiness 6B, has red blooms and needs full sun or sun to part shade.)

Comment [d9]: I am not sure the index is entity-oriented. Many of the descriptors used are derived from anticipated requests.

Both of these searches do a good job of returning relevant results; the possible query statements are clear and built into the graphical representation of the search fields. The Advanced Search takes typing out of the equation, eliminating a common source of user error. The Specific Search interface does require typing, though the reminder to use correct spelling and punctuation, or partial names if unsure of the precise renderings, reminds the user of the limitations of the free-text search. A search tutorial on the site does a good job of explaining how to work around issues of uncertainty about precise spelling. Looking for the Mr. (Mister? Mr?) Lincoln Rose? If the expected result is not returned, try "Lincoln Rose."

The General Search is the most frustrating to use. Search for "white rose" and thousands of results are returned; most are irrelevant. Although judicious use of quotes ("white rose" as

opposed to white rose) can bring results down to a reasonable number, it does so at a huge cost to recall (relevance is still not that good, either). Seven of the results for that query are not roses at all, and six hundred and sixty nine of the roses returned by the Advanced Search failed to be found. Without access to the descriptor pertaining to "category" available in the Advanced Search, the General Search cannot distinguish between a rose and a potato. (Potato 'White Rose' *Solanum tuberosum* is the second of the twelve results returned to the query "white rose.")

To offset the frustration of the general search, Dave's Garden offers the ability to browse the most popular cultivars (weighted by reviews and ratings) from a selection of thirty-six plant groups (chosen by the editors, these are mostly the largest genera in the database, with the largest number of cultivars). A user interested in roses can browse the top ten (user-ranked) PlantFiles roses, the top ten most photographed roses, and they can browse an alphabetical listing of rose cultivars--not a single potato in the lot.

The Garden Watchdog offers a mix of restricted-field free-text searching and browsing. Companies in the highly ranked WatchDog 30 may be browsed through their listing on the Watchdog's top-level page. The most recently added vendor ratings are also featured on that page, for easy browsing. There is also browsing by first letter of company name, category (product category), country, North American state/province (Canada and U.S.), and a list of company ownerships that details what companies own what nurseries.

Users can search for vendors by zip code and by company name. The search by zip code returns a list of vendors ranked in order of nearness to the requested zip code (see section 5.2 below), a useful feature for users who might want to visit the nursery, rather than mail order. The search by company name suffers from the same problem with free-text searching already seen in the General Search of PlantFiles: spelling matters. A failed search brings up an Advanced Search feature that offers the user the following fields to refine the search:

- Search Text: (spelling still matters!),
- Specializing in: a scrollable drop-down box of product categories (as seen in the PlantFiles Advanced Search, these entities behave as descriptors or search keys),
- Miscellaneous: (another drop-down box, this one relating to business status: wholesale, retail, organic, association memberships, etc)
- State:,
- Country
- Sort by: (company name, rating, stte/province.country, and date added).

It is odd that a failed search is required to bring up the Advanced Search feature. The absence of cross-referencing in the product categories has a negative impact on that field's recall: a user must know that a failed search for bulbs:daffodils should be adjusted to bulbs:Spring-blooming and dormant potted bulbs (see section 5.2 below).

5.1 Sample searches—PlantFiles

PlantFiles Advanced Search:
Bloom Color: White (w)
Limit to: Roses
676 plants found.

PlantFiles General Search:
12 plants were found that matched your search terms. ("white rose")
All 12 have “white rose” in their names, including the second result, the ‘White Rose’ potato. Compared to the 676 plants found in the Advanced Search, this search has very low recall. Seven of the twelve results are not roses, so relevance is low as well.

12,187 plants were found that matched your search terms. (white rose)
Searches for white AND rose and white OR rose yielded the same results. Many of these results are not roses or white, or either. This search has extremely low relevance.

Comment [d10]: precision. Relevance is a property of the relationships between the user's need or a query with an item in the collection.

5.1 Sample searches—PlantFiles, continued

PlantFiles Specific Search:
Query: Common name “rose” AND Cultivar “crepuscle”
Did you mean: Cultivar:**Crepuscule**?
Did you mean: Cultivar:**Creamsicle**?
No results found!

Corrected Specific Search based on suggestion:
Query: Common name “rose” AND Cultivar “crepuscule”
Plant Thumbnail
Noisette, Tea Noisette Rose 'Crepuscule'
Rosa
Hybridized by Dubreuil, 1904
Additional info: (aka Crépuscule)

Query: Common name “rose” AND Cultivar “mister lincal”
Did you mean: Cultivar:**Wisteria**?
Did you mean: Cultivar:**Sister Magic**?
No results found!
(suggestions much less helpful in this case—a very costly typo)

Query: Common name “rose” AND Cultivar “mister lincol”
1 exact match
Plant Thumbnail
Hybrid Tea Rose 'Mister Lincoln'
Rosa
Hybridized by Swim-Weeks, 1964
Additional info: (PP2370, aka Mr. Lincoln)

5.2 Sample searches--The Garden Watchdog
Search by zip:

Here are the companies that are closest to 10960 (Nyack, New York):
Page 1 | 2 | 3 | 4 | 5 Next »

Miles	Company	Location
5 miles	International Bulb Company	Montvale, New Jersey
8 miles	Matterhorn Nursery, Inc.	Spring Valley, New York
9 miles	Waterford Gardens	Saddle River, New Jersey
9 miles	WaterSunTogether.com	Haverstraw, New York
9 miles	Sprainbrook Nursery's Gardening Things	Scarsdale, New York
10 miles	Practically Wholesale Plants (div. of Arnat Corp.)	Yonkers, New York
11 miles	Northern Border Tree Farms LLC	Bergenfield, New Jersey
13 miles	GreenGardening.com	Greenwich, Connecticut
13 miles	Christianne	Port Chester, New York

5.2 Sample searches--The Garden Watchdog, continued

13 miles The Bulb BarnFair Lawn, New Jersey
13 miles Frame-It-All Scenery Solutions Port Chester, New York
14 miles Oopsa Daisy,LLC. Hawthorne, New Jersey
16 miles Orchid Select Haledon, New Jersey
16 miles S & K Wildflower Rescue and Nursery LLC Oakland, New Jersey
16 miles Ultra Sleeve™ Garden Kits (JoMar Greetings) Oakland, New Jersey
17 miles Garden Shield Iris Borer Deterrent Cos Cobb, Connecticut
17 miles New York Succulents Bronx, New York
18 miles Katonah Nursery Katonah, New York
18 miles Peekskill Nurseries Shrub Oak, New York
19 miles Shanti Bithi Nursery, Inc. North Stamford, Connecticut

Page 1 | 2 | 3 | 4 | 5 Next »

Search by company name:
Brent and Beckys
No companies found!

Search by company name:
Brent & Beckys
No companies found!

Search by company name:
Brent
1 company found.
Company Location
Brent & Becky's Bulbs Gloucester, Virginia (United States)

The Brent & Becky's search reveals the same issues as the free-text searching in the General Search in PlantFiles: spelling matters!

Advanced Search:
Specializing in: Daffodils
State: Virginia
No companies found!

Specializing in: Bulbs:Spring-blooming and dormant potted bulbs
State: Virginia
2 companies found.
Company Location
Brent & Becky's Bulbs Gloucester, Virginia (United States)
W.R. Vanderschoot, Inc. Chesapeake, Virginia (United States)

In the absence of cross-referencing between daffodils and spring-blooming bulbs, the burden is on the user to recognize the relationship and refine the search. The fact that the product categories appear in a drop-down box does mitigate the difficulty: at least the user knows what options are available.

6 Conclusion

►

Given that plant identification, acquisition and care are the drivers for most of the searches done on PlantFiles and The Garden Watchdog, (and PlantScout, the search feature that bridges the two databases) how well does Dave's Garden help gardeners to achieve these goals? To realistically address the question, one must weigh the needs of the users against the performance of the system. Dave's Garden's users are primarily hobbyists: the financial stakes are low. There are two membership levels: free and subscriber. Free members can use the General Search and Specific Search in Plant Files; the Advanced Search is reserved for subscribers. The Garden Watchdog is completely available to everyone.

Gardeners are always looking for plants, and the plant database contains information and photos for 214,960 of them. Users can research plants they are considering for their gardens, and gather information about what plants do well where they live, and what plants should be avoided. They can discover what plants need which specific conditions to thrive, and what plants might fit into their landscaping due to their foliage, flower color, bloom time or height. They can learn what methods should be used to propagate and collect seed of a particular plant. They can learn that though Dame's rocket has several aliases: damask violet, dame's-violet, dames-wort, dame's gilliflower, night-scented gilliflower, queen's gilliflower, rogue's gilliflower, summer lilac, sweet rocket, mother-of-the-evening and winter gillyflower, there is only one Hesperis matronalis.

Subscribed members of Dave's Garden can do these things far more easily than non-subscribed ones: the Advanced Search is a much better tool than the General Search. The sample searches show the General Search returns too many results (mostly irrelevant ones) to be an efficient search tool. The burden on the searcher to formulate a better query is unlikely to met by the typical user of the site. The ability to browse popular cultivars among selected plant groups might be enough of a mitigating factor for the casual gardener, but the serious gardener would be better served by the vastly superior performance of the Advanced Search. The sample searches show that the Advanced Search performance could be improved with the addition of cross-references, so that a user looking for a nursery that sells daffodils will find the nursery that sells spring-blooming bulbs.

The conceptual schema addresses gardener's needs very well. Anyone familiar with plant nursery catalogs will recognize the logic behind the attribute/method/requirement schema. The Advanced Search does a good job of leveraging the conceptual schema into an index, resulting in a much-improved search tool.

An additional feature that emphasizes the plant's role in a garden (as discussed in section 2.5) would be a welcome addition. The ability to link to a photo of the plant in a garden (if such a photo is on the site) hints that this is already recognized as a desirable feature.

So how well do these databases serve the Dave's Garden user? In their own rating parlance: for subscribers, they get a positive rating, for free members, a negative one—perhaps raised to neutral if they primarily use the browse popular cultivars feature.

Paper grade A

Overall, this is a thoughtful, preceptive, critical analsysis of OED Online applying course concepts.

I have many detailed comments I want to go over in person.

Learning objectives demonstrated:

2.0.1ˆ 571-L3.1
571-A5Graduates know and understand the functional components of an information system.
2.0.1.1ˆ 571-L3.1
571-A5Graduates are able to use this framework to analyze and critique an information system, such as library.

2.3.1,1.4# 571-L4.2
571-A6Graduates are able to construct a conceptual data schema for a given information system
BT 2.3.4.2

2.3.4.1# 571-L2.2
571-L4.2 Graduates are able to analyze the knowledge structure of an existing information system and apply the results to

- judging the adequacy of this schema with respect to the queries to be answered
- using the knowledge of the schema to exploit fully the possibilities of obtaining answers from the information system RT 2.5.2.1

2.5.2.1~ 571 Graduates are able to conduct good searches that are responsive to user needs NT 2.3.2.1, RT 2.3.2, 2.3.4.1

The Oxford English Dictionary's Online Iteration: Information Retrieval Subsystem Analysis

1. Objectives

1.1 Statement of Purpose

The objectives of this paper are to describe, analyze, and critique the online iteration of the Oxford English Dictionary (OED) whilst also providing a brief historical overview of the original version's origins and evolutions to its present state.

1.2 Specified Account of Objectives

This paper will begin with a brief historical overview of the OED. User information needs' for the system will be analyzed and the fulfillment of, or lack of, those needs by the information retrieval subsystem will be determined. A conceptual data schema will be constructed using the entity-relationship model. The indexing language and process will be discussed, with an acknowledgement of the OED as possibly the first crowd sourced reference work. An analysis of the searching process will also be conducted, and the results of sample searches will be reviewed. An overall critique of the OED's online iteration will conclude the paper.

2. Historical Overview of the OED

2.1 Chronology of the OED's Editions and Versions

	Chronology of the OED's Editions and Versions
1879	Murray work began on the OED
1884	1st volume (or fascicle) published
1928	Last fascicle published; 1st edition of OED complete
1957	Burchfield hired to create OED supplement
1972	1st volume of OED supplement published
1984	Digitization project of OED began
1986	Last volume of OED supplement published
1989	2nd edition print version of OED published, combining 1st edition and supplement
1992	CD-ROM version of OED published
2000	OED online website launched
2011	Updated version of OED online launched with modifications

2.2 Origins of the OED's First Edition

To attempt a "complete re-examination of the language from Anglo-Saxon times onward" proved an ambitious endeavor (Oxford University Press, 2013). However, in 1879 the Philological Society of London enlisted the Oxford University Press and James A. H. Murray to transform this ambition into a multi-volume reference source (Oxford University Press, 2013). Unlike lexicographical collections prior, Murray used a unique process to gather and verify words and their definitions. Quotations of usage from actual readings jotted down on scraps of paper were submitted from readers all over the world (Lerer, 2007). These random scraps were reviewed and organized in what

came to be called the Scriptorium, the central processing center for the dictionary, and served as the components that formed the word entries of the OED (Lerer, 2007). This method allowed for a new means of interpreting language, for it was through these usages in quotations that the origin and evolution of a word could be gleaned, from inception through various cultural contexts (Winchester, 1998). "Only be finding and showing examples could the full range of a word's past possibilities be explored" (Winchester, 1998, p. 26). However, due to the labor-intensive nature of the work, it took forty-four years to complete the first edition, with the last volume published in 1928 (Oxford University Press, 2013).

2.3 Supplemental Volumes to Fill the Historical Gap

While unique in method and beneficial in its use, the cataloging method employed to create the OED had some negative consequences. The examples and language of the original edition were very much products of the Victorian societal perceptions from whence the project began (Lerer, 2007). Many of the meanings encapsulated in the dictionary were now outdated (Winchester, 1998). In 1957, Robert Burchfield was hired to begin creating a supplement to the OED to bridge this historical language gap (Rosenheim, 2008). The supplement to the OED was published between 1972 and 1986, resulting in two separate alphabetical classifications of words (Rosenheim, 2008).

2.4 Digitization of the OED Creates 2nd Edition and CD-ROM Version

In order to combine these two separate multi-volume print works into a unified whole, a digitization project was conceived and began in 1984 (Rosenheim, 2008). The digitization of the OED led to the publication of a multi-volume second edition print dictionary in 1989, the CD-ROM version in 1992, and the inevitable launch of the OED online in 2000 (Rosenheim, 2008).

2.5 OED Online Website Launched in 2000

The initial OED online launch elicited strong reactions from the academic community. The most glaring criticism regarded the lack of Boolean search capabilities (Black, 2000). Similarly, the interface did not support combined field searching within the same query (Black, 2000). While the ability to search by author of the quotations included in the word entries proved a valuable tool, the lack of clear explanation regarding the abbreviations used for particular authors nullified that value in some queries (Black, 2000). Other repeated criticisms included the perceived excessive amount of information displayed on the screen, and the high cost that would prevent the average individual from subscribing (Booth, 2000). And if, as shown based on high subscription prices, libraries and other educational institutions were the target users, the system website failed to anticipate their needs by not including the capability to conduct advanced searches.

2.6 OED Online Updated Over a Decade Later

In 2011, the OED online underwent its first major update that altered the graphic user interface tremendously. The new interface was touted as more user-friendly and provided features that were lacking in the 2000 edition of the website (Black, 2011). The new advanced search features support Boolean searching with all operators accessible through drop-down menus (Black, 2011). This type of advanced search also allows for combined searching of different fields within the same query

(Black, 2011). Furthermore, this current iteration of the OED online allows for other methods of searching through hyperlinks incorporated into word entries (Black, 2011). "…the OED online is taking one of the greatest works of the English language, a collected record of our past, into the modern world" (Cliff, 2000, p. 2)

3. User Needs

3.1 Libraries as Target Institutional Users

Clearly based upon the substantial subscription cost of the OED online, the target users for this system are individuals that are participants in academic, and other educationally oriented, institutions. Libraries are the primary institutions within the institutional user group, as libraries constitute the primary purchasers of database subscriptions. The types of users that would access the OED online database vary amongst many demographical factors. For academic libraries, the users could include students of varying levels, professors, alumni, administrators, and librarians. Similarly for elementary school and secondary school libraries, the users could include students of varying levels, teachers, administrators, and librarians. For public libraries, the users could include all citizens of the community that access the library either physically or remotely.

3.2 Conducted Assessments of User Needs

Once the digitization of the components from the original OED and the supplemental volumes was in progress, the development of prototypes of the online system began. The lexicographical and technical staffs were the creators of the prototypes, each staff contributing specialized knowledge to the project (Elliott, 2000). "To rely solely on the user's statement about his or her needs means to abdicate professional responsibility and would be akin to a doctor's relying on a patient's assessment of necessary treatment" (Soergel, 1985, p. 98) Thus, the corresponding reasoning for the internal production of prototypes regarded that the dictionary content as the primary entity was known most intimately by lexicographers. Whilst the system technicians were then able to take the content as the focal point and provide the functionality for its access in designing the interface, and its correlative information storage and retrieval (ISAR) subsystem.

A genuine needs assessment is an obligation shared between both the information professionals and the users of the system (Soergel, 1985). Thus, the prototypes were toured throughout libraries to elicit feedback from the users before launching the website (Elliott, 2000). Several alterations were made based upon the criticism from the users of the prototypes, including a simplification of the graphical user interface (Elliott, 2000). It is assumed that the assessment of user needs was conducted either solely or predominantly with librarians, rather than the patrons of the libraries.

3.3 User Expectations and Unbeknownst Information Needs

At the time of digitization and prior to the inevitable online launch, the OED already merited a level of respect amongst academics as a dictionary and, furthermore, a historical academic achievement. Thus users of the online iteration of the OED would anticipate the same components of word entries retrievable in the multi-volume print format including pronunciations, etymologies, definitions, and

quotations. Additionally users would anticipate that the OED online would contain all the information found within prior print editions, and that new terminology would be added to the database that did not exist in any of the prior print editions. The accompanying expectation would be that the same level of quality the OED is reputed for would continue with all new entries added to the online iteration.

From this new online format, users would desire that the information be accessible in ways that were impossible for the print versions. By having more control over the retrieval of the information from this ISAR system, users would be able to interpret the words and accompanying entries in novel ways. "Where necessary, the system can repackage substantive data into a format that the user can understand or add documents that provide the background needed to understand a relevant document" (Soergel, 1985, p. 133). Thus, with well-executed hyperlinks, information needs previously unbeknownst to the users would be met by the OED online. These formerly unknown information needs could range from searching for synonyms to finding all words in the database originating from the fictitious literature of a particular author.

4. OED Online's Conceptual Data Schema

4.1 Entity Types

headword	quotation	part of speech	number
definition	etymology	usage	citation
pronunciation	author	subject	citation style
language	date	edition	quick definition
lemma	alt spelling	region	language status
function	chrono rank	title	

4.2 Relationship Types

headword <isDefinedAs> definition	headword <isPronouncedAs> pronunciation
headword <hasRelatedTerm> headword	headword <OriginatedFrom> headword
headword <OriginatedFrom> language	headword <QuotedIn> quotation
etymology <containsOriginOf> headword	etymology <containsOriginFrom> language
quotation <whenQuoted> date	quotation <QuotedBy> author
quotation <chronoRanked> chrono rank	quotation <QuotedWithin> title
headword <hasEtymology> etymology	headword <UsedAs> part of speech
headword <hasNarrowerTerm> headword	headword <hasBroaderTerm> headword
definition <SubjectCategorized> subject	definition <UsageExtentCatgeorized> usage
headword <lastUpdated> edition	headword <lastUpdated> date
headword <AlphabeticallyOrdered> number	quotation <exemplifiesDefinition> definition
headword <isCitedAs> citation	citation <hasCitationStyle> citation style
headword <hasQuickDefinition> quick definition	headword <categorizedByLemma> lemma
headword <isAltSpelt> alt spelling	headword <usedInRegion> region
headword <hasLanguageStatus> language status	headword <functionsGrammar> function

4.3 Example of the Entity-Relationship Types Using the Headword "Philosopheress"

- philosopheress <isDefinedAs> A female philosopher; (also) the wife of a philosopher.
- philosopheress <isPronouncedAs> fɪˈlɒsəf(ə)rɪs
- philosopheress <hasRelatedTerm> philosophes
- philosopheress <OriginatedFrom> philosopher
- philosopheress <OriginatedFrom> English
- philosopheress <QuotedIn> She's a Philosophresse Augure, and can turn Ill to good as well as you.
- philosopher n. + -ess suffix <containsOriginOf> philosopheress
- philosopher n. + -ess suffix <containsOriginFrom> [none indicated]
- She's a Philosophresse Augure, and can turn Ill to good as well as you. <whenQuoted> 1631
- She's a Philosophresse Augure, and can turn Ill to good as well as you. <QuotedBy> G.Chapman
- She's a Philosophresse Augure, and can turn Ill to good as well as you. <chronoRanked> first
- She's a Philosophresse Augure, and can turn Ill to good as well as you. <QuotedWithin> *Warres of Pompey & Caesar*
- philosopheress <hasEtymology> philosopher n. + -ess suffix
- philosopheress <UsedAs> n.
- philosopheress <hasNarrowerTerm> [none indicated]
- philosopheress <hasBroaderTerm> philosopher
- A female philosopher; (also) the wife of a philosopher. <SubjectCategorized> philosophy
- A female philosopher; (also) the wife of a philosopher. <UsageExtentCategorized> rare
- philosopheress <lastUpdated> 3rd
- philosopheress <lastUpdated> 2006
- philosopheress <AlphabeticallyOrdered> 142,478
- She's a Philosophresse Augure, and can turn Ill to good as well as you. <exemplifiesDefinition> A female philosopher; (also) the wife of a philosopher.
- philosopheress <isCitedAs> "philosopheress, n.". OED Online. December 2013. Oxford University Press. http://www.oed.com/view/Entry/142478?redirectedFrom=philosophress (accessed December 1, 2013).
- "philosopheress, n.". OED Online. December 2013. Oxford University Press. http://www.oed.com/view/Entry/142478?redirectedFrom=philosophress (accessed December 1, 2013). <hasCitationStyle> Chicago
- philosopheress <hasQuickDefinition> [none indicated]
- philosopheress <categorizedByLemma> [none indicated]
- philosopheress <isAltSpelt> philosophress
- philosopheress <usedInRegion> [none indicated]
- philosopheress <hasLanguageStatus> [none indicated]
- philosopheress <functionsGrammar> feminine

5. OED Online's Multiple Index Languages

5.1 Natural Language as the OED Index Language

A dictionary, peculiarly and traditionally, is an alphabetical index of words occurring in natural language. Therefore the OED itself is a kind of index that was created by the crowd-sourcing method begun with Murray in the late nineteenth century. As the foundation of the OED online is the entries that comprised the 1st edition of the OED and it's multi-volume print supplement, it is by extension also a type of natural language index in essence.

However, it is important to distinguish natural language from "free-text" in this instance. While the headwords that denote the entries of the OED online are constituted with natural language, there is an index of those terms that underlies the basic searching of the database. The user may enter any number of "free-text" terms within the search but due to the high degree of order used to sort entities in highly specified classes, the results may not prove desirable using this search methodology. Outside of the headwords, the corresponding definitions and quotations are recorded in a full-text manner and thus can be searched as free-text with presumably better results.

5.2 Exhibiting High Degree of Order

The degree of order is assessed by analyzing three components which are "the number of criteria used in sorting (sortkeys), the specificity of the descriptors (broadly defined) used for sorting, and the number of access points per entity" (Soergel, 1985, p. 199). The OED online uses a large number of sortkeys for the organization of all the entry related information as demonstrated by the entity list in the conceptual data schema. Furthermore the descriptors are highly specified, as they each are representative of the individual word entries. There are no pre-combined or post-combined descriptors, unless they occur as such in the natural language. For each word entry of the OED online, there are many different access points across the spectrum of sortkeys employed. And while the retrieval subsystem did not initially support searching in accordance with all the types of access points, the most recent version of the OED online after the 2011 update does support an increased access using all the sortkeys in a variety of ways.

5.3 Entity-Oriented Indexing vs. Request-Oriented Indexing

The OED does not fit neatly into either the concept of entity-oriented indexing or request-oriented indexing. An argument could persuasively, and logically, be made for either approach to the process when all factors of its creation are taken into account.

Because the words included in the OED online are resultant from the numerous submissions of people throughout the world, and throughout time, it could be argued that the indexing process was conducted from a request-oriented approach. The Victorian Era readers who submitted the first quotations incorporated into the entries of the OED, were essentially providing the information that they would wish to seek through queries. "…every word…tells the story of the men and women who defined it, marked their books and handed in their slips" (Lerer, 2007, p. 245). Every submitter was partaking in the process of request-oriented indexing by consciously determining which words, definitions, and supporting quotations would be sought out by future dictionary users.

Yet, the argument could also be made for the entity-oriented indexing approach. For while the general public were submitting the quotations on slips of paper, it was Murray and his lexicographical team that were conducting the indexing and editing of the word-related information. When alphabetizing and creating the volumes of the original OED, the indexing was done for each entry in relation to the entity or headword.

Regarding the OED online, at first the entity-oriented approach to indexing was dominant. Most reviewers were highly critical of the lack of advanced search features, as previously noted. There was little to no thought conducted regarding how users would search for the information and what queries would be used. The primary goal at that point in the OED's evolution was to convert the digitized version into an accessible online version, where each entry could be reached by very few access points. The alterations to the interface and addition of advanced search features in 2011 exhibit a request-oriented approach to indexing. "Usefulness for searching is the primary criterion in descriptor selection. The descriptors are arranged in a meaningful hierarchy that communicates the user's conceptual framework to the indexer" (Soergel, 1985, p. 231). The addition of the Historical Thesaurus to the OED online in 2011 is a pertinent and beneficial example of this sentiment in the request-oriented indexing approach.

5.4 Historical Thesaurus as Faceted Hierarchical Classification

The OED online's Historical Thesaurus is a hierarchical index containing all the dictionary word entries that supports full browsing capabilities. Prior to its inception, word related information could be searched utilizing various means but there was neither a hierarchical nor conceptual classification of the words. The dominant organizational structure was an alphabetized list, which digitization was able to circumvent through basic interface searching. Thus, this Historical Thesaurus for the first time allows the users to conceptually arrive at terms using their own knowledge and perception as a frame of reference.

All dictionary terms are arranged hierarchically under the three main facets of the External World, the Mind, and Society. "Faceted classification as a method of file organization is concerned with just that problem of sorting sequence for human searching" (Soergel, 1985, p. 200). This solves several retrieval problems a searcher may have when using the OED online ISAR system. Previously, if the user did not know the correct spelling of a word of interest there was a chance that retrieval would yield no results. But if the user has a conceptual understanding or even a broad conceptual framework to place the word of interest in, the Historical Thesaurus may help locate the entry. A similar retrieval problem exists when a user is trying to find a word that defines a concept but does not know what specific word is being sought for. With the Historical Thesaurus, the concept can be known and the word that defines the concept can be discovered.

Links to the Historical Thesaurus are fully integrated throughout all the dictionary entries. If there are several definitions for a headword, each will contain a link to the Historical Thesaurus. When the link is activated, a small window opens exhibiting the hierarchically inherited chain of terms for the chosen definition of the word. If there are multiple locations for the chosen definition, multiple hierarchically inherited chains of terms are displayed. Additionally, there are links within these small windows that will then take the user to the actual Historical Thesaurus if the choice is made to leave the webpage of the entry where they are currently located.

6. Digitizing and Indexing the OED

6.1 Digitization Process

In 1984, the project of digitizing the OED began which created the foundation from which the website version was built upon. At the time, the purpose for digitization was to merge the two separate alphabetical listings of words contained in the 1st edition of the OED and the supplemental volumes. It was a massive and costly undertaking that Richard Charkin, the deputy academic publisher at the time, had enough foresight to suggest (Rosenheim, 2008). The government of the United Kingdom, the International Code Council, the International Business Machines Corporation, and experts from the University of Waterloo assisted the Oxford University Press in the project's realization (Rosenheim, 2008).

A central objective to the digitization process was to convert the information contained in the multiple print forms of the OED into machine-readable text (Elliott, 2000). A small army of typists was enlisted and over a period of eighteen months entered all of the dictionary content into computers (Elliott, 2000). The only differentiation the typists were asked to include were the variations in typeface used in the print volumes. "…the next imaginative move was made: the typefaces were converted automatically to codes identifying the various components of the Dictionary – including headwords, numbered sense definitions, etymologies, quotations, and cross-references" (Elliott, 2000, From hot metal to computer section, para. 1).

6.2 Standard Generalized Markup Language Employed

The computer scientists working on the project used generalized markup language to encode the various dictionary components of the text (Elliott, 2000). This method was chosen due to the coding's ease of interpretation by both machines and humans. Another benefit to using generalized markup language was the flexibility of transferring the data from one machine to another while still retaining its interpretability (Elliott, 2000). Once standard generalized markup language (SGML) emerged, which was a standardized version of markup language, an additional version of the OED was produced conforming to SGML coding. OED was never converted into HTML as "all that useful information on text structure would be lost", since HTML specializes in storing information pertaining to the aesthetics of the text rather than conceptual content related information (Elliott, 2000, Were there alternatives to SGML? section, para. 1).

The Oxford University Press then partnered with High Wire Press in 1999 to host and maintain OED online once it was launched (Elliott, 2000). High Wire Press was known to begin any online publication project by converting data to SGML, thus the Oxford University Press only had to provide the source files already existing in SGML for the launch of the OED online in 2000 (Elliott, 2000).

6.3 Polyhierarchical Indexing Yields High Specificity and High Exhaustivity

The way in which the OED online is organized allows for various search methodologies. As previously mentioned, the indexing has a high degree of order due to each word constituting its own descriptor. The indexing is also polyhierarchical in nature for an individual word can be located by searching from numerous components that make up the word entry (Soergel, 1985). Each component that makes up a

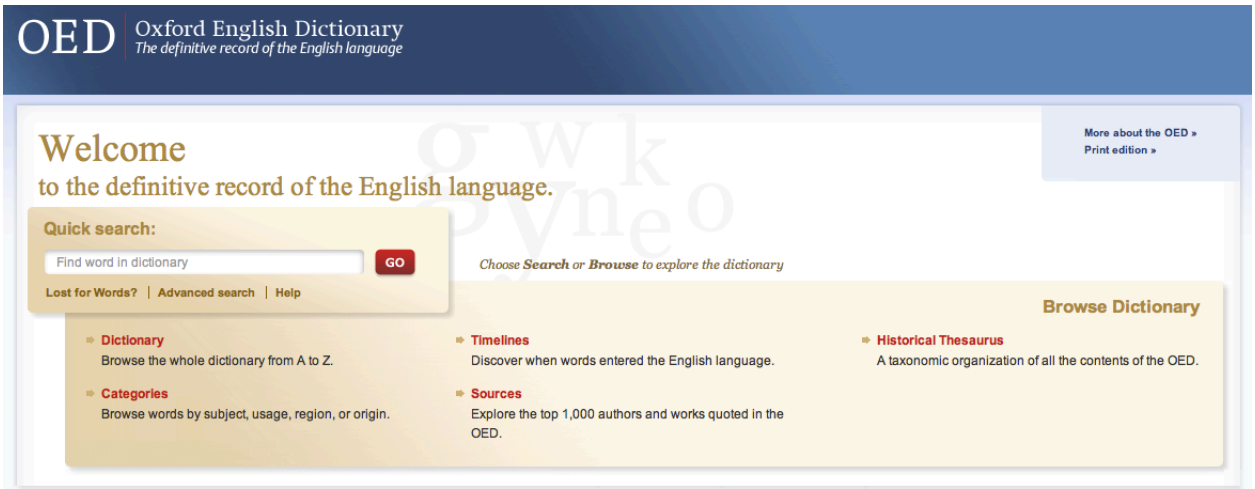
word entry can be likened to subsisting within a different facet. For example a word can be searched by geographical region of use, by an author of a particular quotation, or by the etymological language of origin. Each of these represents a different conceptual facet to which the word, and all words, could belong depending on the perspective of inquiry. In a monohierarchical index, the word would only be located by one of these attributes and thus any queries related to other perspectives would yield null retrieval sets.

This type of indexing yields high specificity and high exhaustively, which must be considered for retrieval purposes. Intelligent query formation will allow for adaptability regarding the exhaustivity of indexing used in a system (Soergel, 1985). In a system indexed with a high level of specificity, the user can use very specific terms or descriptors and locate the information desired. With a purely alphabetical system, searching may prove more challenging as with the print version of the OED. But the OED online allows for specific searches to be conducted across various facets. Furthermore, with the inclusion of the Historical Thesaurus, users can now conduct inclusive searches without needing to alter their queries based upon level of specificity but rather focus on the conceptual context of the desired word entry.

7. Searching the OED Online

7.1 OED Online's Homepage

The first search box users are presented upon arriving at the OED online's homepage is a "Quick search" box. Within the box are instructions indicating what the box is used for in light gray lettering stating, "Find word in dictionary". This is creative and intelligent way to alert users to this search box's specific purpose without cluttering the interface with extraneous text.



Additionally within the "Quick search" area are links to other types of searching. "Lost for Words?" is an inventive feature that when activated presents the user with four seemingly random words. This is ideal for individuals passionate about linguistics, or even those individuals just curious about words. "Advanced search" is also linked within this section, with various search features that will be explored

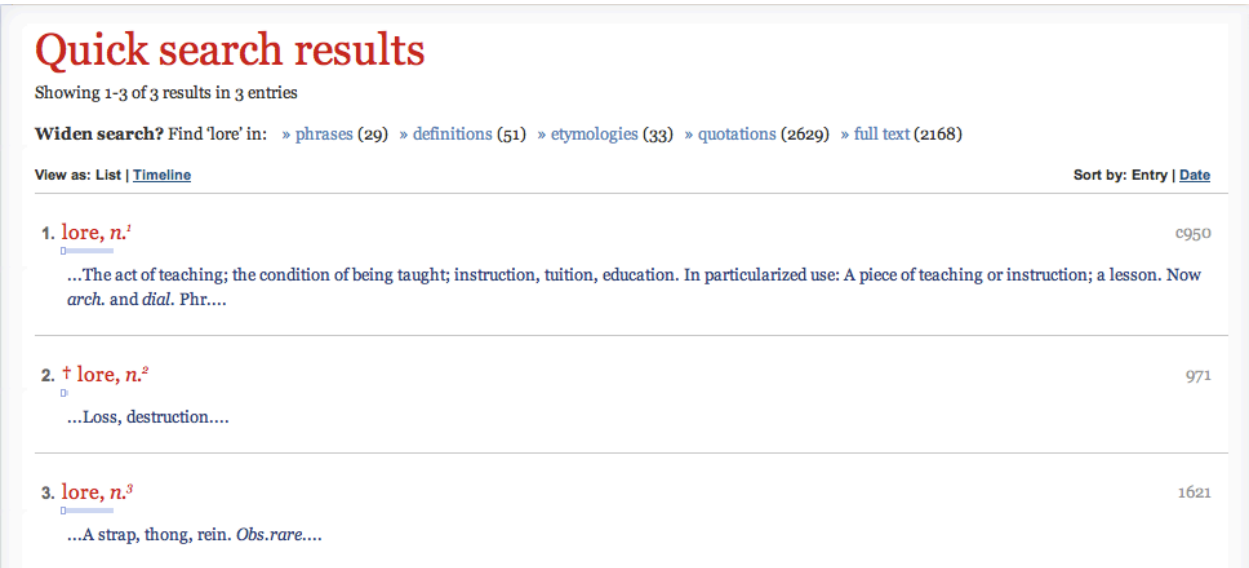
in turn. Clicking on the "Help" link leads the user to a "Quick start" guide with explanations for the various search and browse options offered on the homepage.

Users who arrive at the OED online's homepage are also offered several different browsing options including browsing through the faceted classification of the Historical Thesaurus, or chronologically through the Timelines. Additionally links are present for users that arrived here hoping to gain background knowledge regarding the OED, or who require information about the print edition. The decision to include both of these links together on the homepage is representative of thinking about the users, their reasons for arriving to this page, and how to redirect those that do not wish to search or browse the dictionary-related content.

Further features are provided such as a "Word of the Day" which leads to that word's entry, and includes another link to sign up for "Word of the Day" emails. This is another creative manner to attract logophiles and linguists alike, or just those individuals with a desire to increase their vocabularies for differing reasons. A short list of "Recently Published" words are provided as well, each linked to their respective entry.

7.2 Quick Search Using the Query "Lore"

While the term "Quick search" may imply basic to some, the OED's interface provides a multitude of ways to navigate the website regardless of the stage of the search process the user is in. The interconnectedness of the hyperlinks boasts a complex and intricate web between many locations throughout the site at large. For example once the query "lore" is inputted into the quick search box, several different options are presented for continuance of the search along with the results list.



As inferred from the "Widen search?" options presented, the quick search box defaults to searching amongst the headwords only. From this point if the user was not satisfied with the three results presented but rather was attempting to conduct a full-text search of all entries and find "lore" in quotations, the option to proceed along that search path is provided. This is another example of the

intuitive anticipation of the user's information need, or want, built into the interface. For even if the user mistakenly searches through the quick search when actually needing to conduct a search in fields only offered in the advanced search, the system is highly forgiving and allows alternate routes to achieve the same end goal.

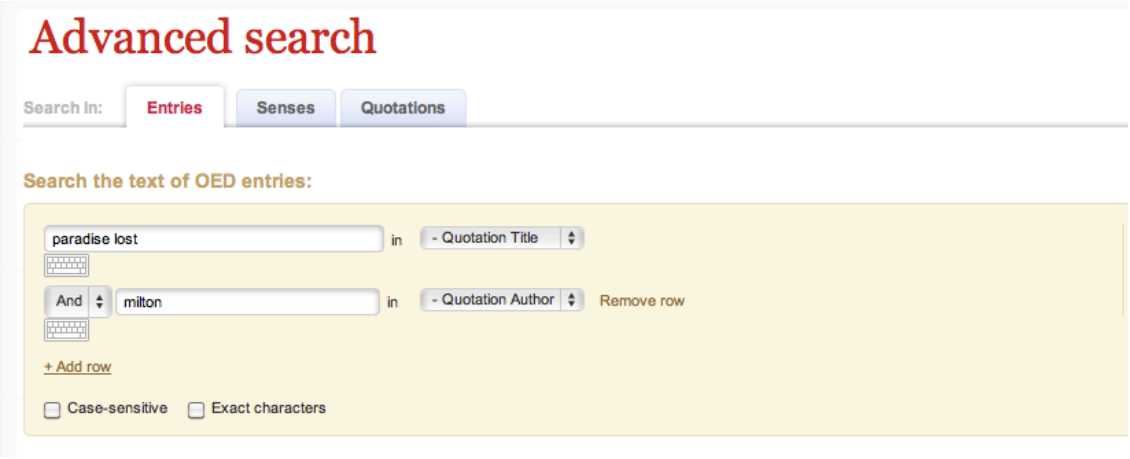
There are also options offered for the user to "Refine your search" by subject, language of origin, usage, and other criteria. This allows for a retrieval set exhibiting high recall, perhaps unexpectedly, to be decreased to a sizable result list for sifting. The user, even with prior knowledge, may not be aware of the copious amounts of entries a word presents within the OED. These limits present a solution to this potential user conundrum.

The user is also offered choices for the way in which the results are viewed. The default settings display the results in list form, sorted by entry. However as the OED is reputed for its historical knowledge of words and their origins, the system supposes that some users may be interested in displaying results according to date or viewing results as a timeline.

Within each result listed the headword, identical to the query, is displayed with an accompanying excerpt from that entry's definition. This allows the user to see from the results list if any are relevant to the search.

7.3 Advanced Search Using Boolean and Combined Field Searching

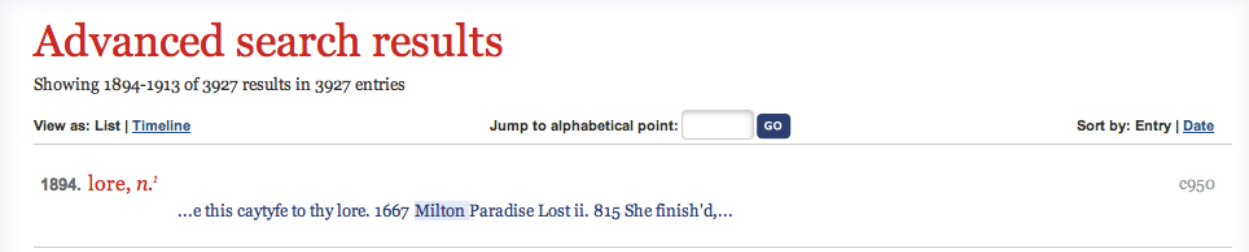
There are a multitude of search methods that can be employed using the "Advanced search" interface. Searching can be conducted within dictionary entries, senses, or quotations. Within these search interfaces are search boxes with corresponding drop-down menus for fields, and drop-down menus with Boolean operators in between. There is a separate drop-down menu to employ proximity searching, with a selection box to indicate that the sequence of the terms as appeared in the results matters. The same limits that can be used to refine results from the results list can be applied in the advanced search interface, such as refining by part of speech or geographic region.



The focus of this search is employing Boolean searching through the interface, while combining two terms to be searched in different fields. The term "paradise lost" was inputted with the "Quotation

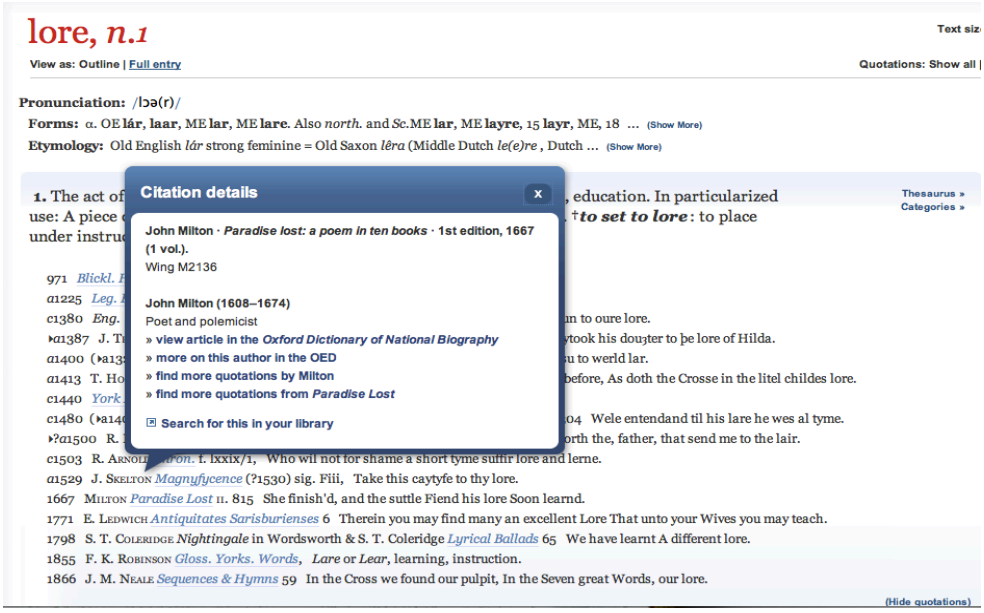
Title" field selected. The default Boolean operator "And" was left, and the term "Milton" was inputted into the second search box with the "Quotation Author" field selected.

The search features accompanying the results list were identical to those from the quick search, without the abilities offered to widen the search. The retrieval set contained 3,927 instances of a quotation from John Milton's *Paradise Lost* throughout the entirety of the OED. A tool above the results list allows the user to "Jump to alphabetical point" within the results list. By inputting "lore" in this search box, the system displayed the 1,894th result with the alphabetical continuance of results following.



7.4 Search Capabilities from within a Dictionary Entry

Once a user selects an entry to view from the results list, there are numerous additional ways provided to continue searching the OED and other Oxford University Press online databases. Links abound to entries of the word in other dictionaries, word phrases using the word within the same entry, and the Historical Thesaurus including all terms within all hierarchical chains used to reach the selected headword. The "word wheel" displayed allows the user to see the chosen word's place within the alphabetical listing of all OED headwords and offers the option to jump to another point in the listing.



Furthermore, activating a link corresponding to a text title of a quotation opens a small window containing more links for further searching that text or the author of that text within the OED, or in other Oxford University Press online publications.

8. Laudatory Critique

8.1 Perception of OED's Online Information Retrieval Subsystem

The information retrieval subsystem of the OED online is well structured and intelligently designed. A user who is unfamiliar with the OED may feel overwhelmed upon first using the system. However the target user group consisting of scholars, librarians, and other intellectuals will find the interface intuitive and intricate. The design allows for navigating through all the dictionary entries in seemingly limitless ways, while having the user feel in control of the direction of that navigation.

8.2 OED Online Maintains the High Quality of Repute

While reviews of the OED online's original launch overwhelmingly criticized the lack of search capabilities and functionality, the major interface update of 2011 has addressed these deficiencies and exceeded expectations. The result is an innovative and poignant iteration of this scholarly masterpiece, and a continuance of an academic tradition that dates back to the OED's original inception in 1879.

9. References

Booth, C. (2000). Online product review: Oxford English Dictionary online, chief editor: John Simpson; Oxford University Press. *Law Librarian, 31*(4), 251-252.

Black, K. (2000). Oxford English Dictionary online. *Booklist, 96*(18), 1770-1771.

Black, K. (2011). OED online's makeover. *Booklist, 107*(18), 68.

Cliff, P. D. (2000). The Oxford English Dictionary online. *Ariadne, 23*. Retrieved from http://www.ariadne.ac.uk/issue23/oed-review

Elliott, L. (2000). How the Oxford English Dictionary went online. *Ariadne, 24*. Retrieved from http://ww.ariadne.ac.uk/issue24/oed-tech/

Lerer, S. (2007). Pioneers through an untrodden forest: The Oxford English Dictionary and its readers. In *Inventing English: A Portable History of the Language* (pp. 235-245). New York, NY: Columbia University Press.

Oxford University Press. (1903). Lore, n. In *OED online*. Retrieved 12/6/13, from http://www.oed.com/view/Entry/110333?rskey=6UXpAD&result=1&isAdvanced=false#eid

Oxford University Press. (2006). Philosopheress, n. In *OED online*. Retrieved 12/1/13, from http://www.oed.com/view/Entry/142478?redirectedFrom=philosopheress

Oxford University Press. (2013). History of the OED. In *OED online*. Retrieved 9/30/13, from http://public.oed.com/history-of-the-oed/

Rosenheim, A. (2008). Oxford still loves the 'OED'. *Publishers Weekly, 255*(50), 8-9.

Soergel, D. (1985). *Organizing information: Principles of data base and retrieval systems*. Orlando, FL: Academic Press, Inc.

Winchester, S. (1998). *The professor and the madman: A tale of murder, insanity, and the making of the Oxford English Dictionary*. New York, NY: Harper Collins Publishers, Inc.

A perceptive essay showing good understanding of course material **A**

Application of UBLIS571 Course Concepts to Academic Librarianship

1 Introduction

This paper is a reflective essay that describes the application of course concepts from UBLIS571 to a future career as an information professional in an academic library setting.

To begin with I must first discuss what information professionals are. As defined in one of the UBLIS571 readings, titled *Competencies for Informational Professional of the 21st Century*, informational professionals are individuals who use information strategically in their work to help their organizations and or clients meet key goals. I must also discuss what an academic library is so that we can understand the relationship between information professionals and the academic library. The Online Dictionary for Library and Information Science defines an academic library as a library that is an integral part of a college, university, or other institution of postsecondary education, administered to meet the information and research needs of its students, faculty, and staff.

As discussed in lecture 1.1c, there are five major roles of information professionals

1. Answer questions and find things
2. Organize things so they can be found
3. Help people produce information
4. Teach
5. Develop and set up systems for all of the above

These five roles can be applied to academic librarianship. The information professionals working in an academic library have the responsibility to make sure information is accessible to students, faculty, and staff. Additionally, the information professionals need to educate the users on effective ways to locate and use this information to help meet their needs. The application of these five roles in an academic library will be discussed in detail in relation to the UBLIS571 course concepts learned.

2 Roles of Information Professionals in an Academic Library

The five major roles of information professionals mentioned above can be categorized into two main aspects of librarianship: user services and technical services. User services entail the information professional interacting directly with the library users to find information. An information professional working in the aspect of user services should expect lots of interaction with library users. On the other hand, information professionals working in technical services tend to have little contact with library users. They serve as behind the scenes librarians. The primary role of technical services is to acquire, prepare, and classify library materials so that it can be made available to users. After taking UBLIS571 I feel I have a good understanding of what it is like to work in these two aspects of librarianship. When I began the semester I did not know what type of librarianship I wanted to pursue. Through UBLIS571 and UBLIS518: Reference Sources and Services, I have discovered I am really interested in working in an academic library setting. I can see myself applying the course concepts learned in UBLIS571 while demonstrating the five major roles of information professionals.

2.1 Answer Questions and Find Things

It is likely that a person who goes to a library is there to find information about something. The problem though, is that the person might not know how to find the information they are seeking. That is when the information professional comes into the picture. In an academic library, the reference librarian is available to help with user's informational needs. The user could be a student looking for sources for their research paper or possibly a professor looking for materials to help teach their course. It is important to understand that users' informational needs come in all varieties. As an academic librarian, I would need to be prepared to help answer a variety of questions and help users find information from a variety of sources. All information cannot simply be found in one place.

To be able to answer a question posed by a student or faculty member and help find information in the library, as an academic librarian I would need to do a few things. I need to explore the user's information need by understanding three things

1. The user's problem
2. What the user knows already
3. How the user thinks

Once I have an understanding of these three things then I can help direct the user to places to find the information they are seeking. First, my role as the academic librarian will be to work to identify and understand the needs of the user. In order to accomplish that, I will need to conduct a reference interview between myself and the user seeking information. As noted in lecture 2.1, an information professional has to think about a user's question and develop different clues that could lead to finding relevant information sources. During a reference interview, an information professional interacts with a user to analyze the user's current problem or question to help determine what the user's informational need is. Once the information professional and user identify the informational need then they can begin searching for items to help meet the need.

As an academic librarian it is likely I will be dealing with the Library of Congress Classification System (LCC) and Library of Congress Subject Headings (LCSH) since that is what is commonly used by academic libraries. This means I need to be familiar with the design of the classification system and terms used as subject headings so that I will be able to assist library users in finding materials in the library. UBLIS571 has helped familiarize me with both of these systems used for subject access.

The Library of Congress Classification System (LCC) is a systematically arranged scheme of subject classes. Each class is identified by a class number, called the Library of Congress Call Number, which marks its place in the classified arrangement. The notation used for LCC is simple. Letters are used to identify the main class and integers are used for the divisions. For example, the class number BJ 2139 represents an item about *Etiquette for airplane travel*. The main class BJ is *Ethics. Social usages, Etiquette.* and then the integers 2139 after the main class BJ represent a specific division of the main class which in this case is etiquette for airplane travel. LCC is used for the systematic subject arrangement of books on the shelves. One problem with LCC as discussed in lecture 8.1 is the fact that books are customarily shelved in

Comment [d1]: Class number. A call number is given to a book; it consists of the class no. appropriate for the book's subject and an appended Cutter number, usually derived from the author name

only one place so only one LCC class is assigned to a book which provides only one access point. A book may deal with more than one class but the cataloger has to choose only one to assign to the book. It is important for me to know each of the classes under LCC so that if a user asks a question about where certain books are located I would be able to direct them there. For example, a student who is studying music education might ask which section of the library has books about music or where musical compositions are stored. I would direct them to the LCC main class M *Music and Books on Music*. To help students and faculty be able to find items in the library collection that are shelved based on the LCC, the library needs to have clear labels identifying where items are stored.

The Library of Congress Subject Headings (LCSH) is an alphabetical list of subject terms. These terms are used to index books to make them searchable in an Online Public Access Catalog. Usually, several subject headings are assigned to a book to provide multiple access points. These subject headings are especially important for me to understand as an academic librarian so that I can help users formulate queries based on the descriptors identified by the Library of Congress Subject Headings.

2.2 Organize Things So They Can Be Found

As a method for organizing things so that they can be found, information professionals catalog books using the Machine-Readable Cataloging (MARC) format and the Anglo-American Cataloging Rules. This is something I have become quite familiar with while taking UBLIS571. No matter what type of library setting an information professional works in, there will always need to be records created for items. As I learned in lecture 4.2 a MARC record is an extended frame with information about a document. The MARC record incorporates many statements that link the document to some other entity. An entity in a MARC record could include a person, a date, or a subject for example. The statements in a MARC record could be represented as a series of statements or could stand on their own. For example, a statement from a MARC record might read document <has title> text. This is known as a triple.

Comment [d2]: Now mostly RDA.

My favorite concept covered in UBLIS571 was hierarchical inheritance. It is a skill that I believe I have mastered and I look forward to implementing it in a future career. This concept can be applied to many different contexts. One key thing I discovered was that hierarchical inheritance can be used to represent information in a more meaningful way. Hierarchical inheritance helps eliminate redundancy and present information more clearly.

2.3 Help People Produce Information

As an information professional, an academic librarian's job does not end once information is retrieved for a user. In addition to helping users find information in the library, one of my duties as an academic librarian will be to assist those users further by helping them edit and format documents and interpret the information that they have found. Academic librarians also need to help users understand the importance of citing their sources used in research and the correct ways to do so. I can help familiarize users with the different citation

styles such as the American Psychological Association (APA), Modern Language Association (MLA), or Chicago Manual of Style. Those are all commonly used formats for reference citations.

Creating and or maintaining a library website will also be one of my responsibilities as an academic librarian. I will need to make sure that the academic library website meets the needs of all of its users. This is an area where knowledge of document structure and efficient web design come into play. I will be able to use my knowledge of web design acquired fromUBLIS571 and UBLIS506: Introduction to Information Technologies to increase information organization on the website. I will make sure the website is structured in a recognizable format for easier searching and understandability. As discussed in UBLIS571 good web design takes into account who the users of the system will be and what their needs are. Effective web design also arranges information in a meaningful order and uses chunking to keep related pieces of information together .A well designed website can be a powerful communication and information tool between the information professionals in the academic library and the users of the academic library including students, faculty and staff. It is common for academic library websites to provide research strategies to users. Having a library website will allow for 24/7 access to information about the library and might help answer some of the user's questions about their informational needs without having to talk to an actual information professional.

2.4 Teach

Information professionals have the responsibility to teach people the following things

1. How to find information which requires the information professional to teach the user about information organization
2. How to access and evaluate information
3. How to use and integrate information
4. How to present information

As an academic librarian I will not be in the role of a classroom teacher. However, I will still have numerous opportunities to educate library users about how to find information to help meet their needs. The UBLIS571 reading, titled *Competencies for Informational Professional of the 21st Century*, addresses the fact that information professionals are educated professionals who understand the value of developing and sharing their knowledge. It is highly likely that an information professional will be able to find information easier than a novice library user seeking information. It is important for the academic librarian to work with the user seeking information to help them develop their search skills. Information professionals have the responsibility to teach users to become more information literate. This means the user will improve upon their ways in searching for, interpreting, and handling information.

Comment [d3]: In many academic library positions you will spend a good portion of your time teaching in classrooms, either in dedicated user education classes or giving a lecture or two to teach information skills in an introductory class in any subject. In an interview they may well ask you how you would go about prepare a lesson plan. It is a good idea to take LIS 523 User Educations.

The concepts I have learned in UBLIS571 have helped me be a better searcher for information. I know that searching for information can be a frustrating process for many people. As an academic librarian, I want to be able to teach students and faculty effective ways to search for information without feeling frustrated. In the Soergel, *A general model for searching linked*

Comment [d4]: This belongs also under 2.1.

data or design of an integrated information structure interface, reading it is stated that all searches have things in common as listed below.

1. A search consists of one or more search steps.
2. Each search step starts from something known and a type of link that are both indicated by the user
3. The search leads to something wanted but unknown.
4. The system that the search was conducted in finds or creates what is wanted

Helping users to understand these commonalities of searching will make them better searchers and gain better insight as to what they are trying to find to meet their informational needs.

One of the important concepts that I learned in UBLIS571 was entity types and entity relationships. Prior to this course, I had never heard of the term entity. Now I think about entities and entity relationships often in many different contexts. Entity relationships play a role in finding information. Since I am now familiar with entities and entity relationships I believe I am able to search for information better. As an academic librarian, I can use my knowledge of entity relationships to help students and faculty find information. One of my roles of information professional as teacher could be to help students and faculty understand the concept of entity relationships and how information is linked through these relationships. Understanding some entity relationships will help library users be able to formulate stronger queries and thus conduct more relevant searches. For example if a student is looking for a document by a specific author then the two entities they are dealing with is person and document. The entity relationship for those two entities would be <has creator>. Essentially what the user would be telling the information system to retrieve would be anything that matches the statement document <has creator>author.

Something that is difficult for users to comprehend is the idea that not all searches for information can be conducted in the same way. It is important for the academic librarian to work with users and help them understand different ways that searches can be conducted using various systems. The academic librarian should teach users to think about the possible conceptual data schema for a given system. The users might not be familiar with what a conceptual data schema is but they should be able to identify the type of information they are looking for and think about possible places they could look to find that information. For example, if a person is looking for a full text scholarly article from a periodical, they are not going to consult the library catalog to find an article. They should understand that a library catalog is used to store records of items that the library has in its physical collection such as books or dvds. A database such as Academic Search Complete, which is a comprehensive scholarly full text database, would be a better place for the user to conduct a search to find a scholarly article. An academic librarian can serve as an information guide and mentor to help direct users to proper places to conduct searches for various types of information. As mentioned earlier, it is important to understand that all information cannot simply be found in one way and in one place. This is something that became very clear to me while taking UBLIS571.

Comment [d5]: Some libraries have set up a system where the user can search the library catalog and outside sources simultaneously in one search. That makes it much easier for the user who just wants information regardless of where she can find it.

The academic librarian also needs to teach users how to formulate good queries in the different systems. To be able to formulate good queries, the user must first understand what a

query is. As discussed in lecture 2.1 a query formulation is a definition. It defines what it means for a document to be relevant for the user. To help formulate queries into an information system, users need to have an understanding of Boolean operators. The use of AND and OR in a query formulation can help retrieve relevant documents. When using the Boolean operator AND in a query statement, in order for a document to be considered relevant both aspects of the query statement must be present. If a user is looking for documents about airports in New York for example, they would use the Boolean Operator AND to search for relevant documents. The two descriptors *airports* and *New York* would be used in the search. The information system would retrieve only documents that deal with both airports and New York. A document about noise control at airports would be ignored whereas a document about subway access to New York airports would be retrieved because it deals with both descriptors. The other common Boolean operator is OR. When using the Boolean operator OR in a query statement, in order for a document to be considered relevant it must deal with at least one of the concepts in the query statement. In the query statement airports OR airways for example, a document would be relevant if it mentioned airways but not airports, or mentioned airports but not airways, or it mentioned both descriptors. As an academic librarian, it is important to inform the users about using Boolean Operators to help find relevant sources in their search. Not everyone is familiar with the term Boolean operator so it is the information professional's role to educate users about this search option. Often times in a library catalog, a user can select to do an advanced search where more than one search term can be entered and the user can select the Boolean operator AND or OR for their search. Before taking UBLIS571 and UBLIS518 the term Boolean operator was unknown to me. I am glad I became familiarized with this concept and can use it to conduct more meaningful searches and will be able to teach users about it someday in my future career as information professional.

2.5 Develop and Set Up Systems for All of the Above

One of my duties as an academic librarian might entail setting up bibliographic and other databases including library catalogs. A library catalog would be an example of an information retrieval system. Users would access the library catalog to search for items available in the library including books and media. I would need to establish a system that models an entity relationship conceptual data schema. As discussed in lecture 7.1b a catalog is a database that contains identifying and descriptive data about objects. According to one of the UBLIS571 readings by Cutter, some objectives of a library catalog include

1. To enable a person to find a book for which either the
   A. author
   B. title
   C. subject
   is known

2. To see what a library has
   D. by a given author
   E. on a given subject
   F. in a given kind of literature

I need to keep these objectives in mind if I am responsible for adding or creating records for items in a library catalog. This goes along with the entity relationships mentioned earlier because the catalog will allow users to search for items based on entity relationships such as searching for an item by a certain title. It is important to note that a library catalog is capable of more than just searching for a book. Instead of using the term book, Cutter should have mentioned that a library catalog's objective is to enable a person to find an item. This could include a book, dvd, map for example and does not limit the search condition to only a book.

As an academic librarian I might also be responsible for setting up document templates for easy creation of documents. This could involve working with the academic faculty to help them design a uniform template for course syllabi or possibly lesson plan templates. As discussed in lecture 6.1b document templates make document creation so much easier and thus save a lot of work. Having a template designed for the academic faculty would save them a lot of time in creating their course syllabi which is a requirement for all academic courses. Good document structure makes reading and understanding documents easier and allows for pinpoint retrieval of relevant document sections. It is highly likely I will be required to create document templates of some sort throughout my career as an information professional. UBLIS571 has helped educate me about the importance of good document structure and design to help organize and present information in a clear and understandable way.

3 Conclusion

Taking UBLIS571 has made me think about information in a new way. I never realized how complex the information storage and retrieval process is. I have learned to appreciate all the labor that is put into organizing information and making it accessible to people with informational needs. Before taking this course I thought I was pretty good at being able to find information but I did not realize all the different possibilities available in finding information. I now consider myself more information literate after taking UBLIS571. Without a doubt the skills I have learned and acquired from this course will help me be successful in my career as an information professional. Currently I am leaning towards becoming an academic librarian and working in the aspect of user services but that may change. Regardless, I will still find myself applying many of the skills I acquired from UBLIS571 into my career as an information professional in whatever setting that may be. I will be demonstrating the five major roles of an information professional as discussed in this paper.

References

Cutter. *Objectives of the library catalog*. UBLIS571 Course Materials

Needham, C.D. (1964). *Organizing knowledge in libraries*. UBLIS571 Course Materials

Soergel, D. (2015). UBLIS571 Lecture Notes.

Soergel, D. A general model for searching linked data OR Design of an integrated information structure interface. UBLIS571 Course Materials

Soergel, D. (1985). *Organizing information: Principles of database and retrieval systems*, San Francisco, CA: Morgan Kaufmann Publishers.

Special Libraries Association. (2014). *Competencies for information professionals of the 21st century*. UBLIS571 Course Materials

Tagging One Image At A Time:
How Folksonomy Affected the Discovery of Images

A perceptive well-researched analysis. Grade A

Abstract

This paper seeks to investigate how the use of folksonomies by users and information professionals has affected the discovery of images on social platforms and in Web 2.0 environments. As more images from archives and museums become available via the web, information professionals are tasked with adapting traditional taxonomies to effectively index images by expanding their efforts to include user-generated metadata.

Introduction

This paper seeks to investigate how the use of folksonomies by users and information professionals have lent to increased discovery of images on social networking platforms and in Web 2.0 environments. As more images from archives and museums become available via the web, information professionals are tasked with adapting traditional taxonomies to effectively index images by expanding their efforts to include user-generated metadata. Though critics argue the consistency and intent of the tagging community, it is undeniable that user-generated data offers a more expeditious and cost effective manner of indexing. The motivation of this paper is to investigate how this user data can add value to the content by linking data to discovery.

Folksonomy in Action:

Folksonomy is a term that was coined in late 2004 by Thomas Vander Wal to express the act of tagging content by information consumers to develop user-generated metadata. The creation of the term came soon after the inception of the social platforms Del.icio.us and Flickr. Taken

An example of a tag-cloud serves as a visual way to represent the frequency of appearance within a specific system, the larger the font the higher the frequency, and the less frequent the smaller the word appears.

from Vander Wal's professional website he defined "folksonomy" as, "the result of personal free tagging of information and objects for one's own retrieval" (Vander Wal, 2007). As the study and understanding of folksonomy has developed and tagging systems have matured, information professionals acknowledge the potential value for information discovery, "Spiteri (2007) studied tags in three large sites for their conformance to national standards for thesaurus construction, and found some close correspondence, leading her to suggest guidelines for incorporating tagging into OPACs. Some libraries see tagging as a way of augmenting the retrieval tools in OPACs (examples include LibraryThing for Libraries and the University of Pennsylvania's PennTags), and the ubiquitous tag cloud is being used to display everything from queries in the Ann Arbor District Library's catalog search cloud to classification number assignments in OCLC's Dewey browser" (Schwarts, 837).

Folksonomy specifically relating to the indexing of images has been steadily executed over the past two decades through the undertaking of large-scale digitization projects and inception of social media platforms. Technology has supported the digitization and manipulation of images, but indexing practices have not grown accordingly with the technology. "Digitization has created a need for more extensive image description to facilitate image discovery in the digital environment. A considerable amount of indexing work accompanies image digitization in library and museum settings…professional catalogers following standards and using controlled vocabulary tools. This approach represents traditional document-oriented indexing where items are classified a priori by professional catalogers with little or no input from the end-users. The web, however challenges this world of clear boundaries and distinct authority roles" (Matusiak, 283).

It is clear to information professionals that if an image is not given adequate description during the indexing process that it will remain buried in a database, never to be viewed by an end-user. Due to the complexity of images and their lack of textual identifiers, before an image is indexed it is inaccessible through way of search. There are two basic ways of approaching image indexing; content-based and context-based. As advances have been made in computer software programs the identification of content-based

features in an image can automatically be extracted by programs or added by a person; while, context-based indexing is conducted by a person identifying semantic relationships through description. There are other theories surrounding image classification that include other categories of data; such as subjective, organizational, factual, personal, etc., but for the purpose of this paper we will focus on the content and context of images as they relate to indexing in museums and social platforms.

Museums

Philadelphia Museum of Art

With the opportunity to engage with a large audience, online museum collections provide an excellent prospect for information professionals to develop tools and theories that offer insight into the study of folksonomy. In the process of writing this paper I contacted the Assistant Director of Collection Information of the Philadelphia Museum of Art, Jessica Milby to discuss the museum's Social Tagging Project. In the brief exchange she made clear that their project allowed users to tag the online collection, but the museum did not ingest any of the information into their internal collection information or library systems. The Social Tagging project was turned on in 2007 and the functionality has not been changed or adapted since the inception of the project. It was inferred that the development of the Social Tagging Project was implemented as an added search function for users, but was not intended for expansion or to inform usage of their system. This museum was not a part of the collaborative steve.museum project, and was using a platform created internally.

Steve.museum And The Metropolitan Museum of Art

The Steve.museum project is an open collaboration of institutions that pool raw data, resources, and research results to develop tools and techniques that facilitate the engagement of online museum collections and tagging communities (Trant, 2006). An example of a preliminary test supplied to a group of volunteers conducted with art from the Metropolitan Museum of Art: Ford Motor Company Collection. The following image is called "The Octopus" and was taken by Alvin Langdon Coburn in 1912.

"The art historian describes it physically and stylistically: Couched in the soft velvety nap of the platinum paper, composed in the languid lines of Art Nouveau, and softly focused, this photograph of New York's Madison Square employs many elements of Pictorialism at its best. However, the dizzying effect of Coburn's aerial view and his fascination with the skyscraper are distinctly and precociously modern. The blend of Pictorialist technique and fresh vision was characteristic of the transitional moment when Alfred Stieglitz, Coburn, Karl Struss, and Paul Strand began to celebrate contemporary urban experience" (Trant, 2).

The record that the Metropolitan Museum of Art retains:

"Alvin Langdon Coburn
(British, born America, 1882–1966)
The Octopus, 1912
Platinum print; 41.8 x 31.8 cm (16 7/16 x 12 1/2 in.)
Ford Motor Company Collection, Gift of Ford Motor Company and John C. Waddell, 1987 (1987.1100.13)

This image was presented to a group and they were asked to provide words or short phrases that describe the image. The pre-test returned 57 unique terms, such as empty park, New York City, winter, urban landscapes, park-goers, etc., that simplified and expanded the discovery of the image; what the unique terms failed to do was ground the image in a historic context. The implementation of social tagging in institutions such as museums is to connect the community and the collection. The image's record still remains relevant to create context and a greater meaning to the image, but social tagging allows the user to more easily access the image.

Social Platforms

Since the inception of Flickr and Del.icio.us, social platforms have had explosive growth. The availability and use of social platforms has far outgrown its expectation, and is now being consulted for a wider range of use than initially expected. Social media platforms allow users to be creators and catalogers of content.

Flickr:

A popular online photo sharing service supported by Yahoo!, allows users to share their images and create descriptive tags that apply to their images and others' images. The resources in photo sharing services, "photos", "are supposed to facilitate the exchange and communication between the platform's users, and above all provide added value for the individual users. The resource 'photo' is particularly enticing in this respect: photographs provide glimpses of other lives, show other points of view, 'say more than a thousand words' and are personal, even if no person should be visible on them" (Peters, 70). Daniel H. Pink eloquently writes about the structure of Flickr and social platforms alike, "Grass-roots categorization, by its very nature, is idiosyncratic rather than systematic. That sacrifices taxonomic perfection but lowers the barrier to entry" (Pink, 2014). The rapid growth of Flickr over the past decade only supports the earlier commentary about the impracticality of professional indexers applying metadata to all the digital images being placed on the web. Based on the claims of Svilian and Jorgensen (2009) there were 250 million photos on flickr.com in 2007, "it would take 100 indexers over 60 years (assuming three minutes per image and forty hours per week) to catalog the 250 million images on Flickr in 2007 "(Willey, 3). In this platform images that are uploaded are only accompanied by their tags and user description, traditional indexing does not complement tags on this platform.

Twitter:

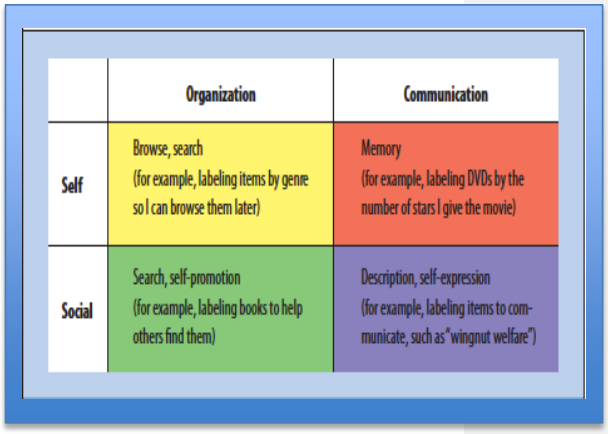
Similar to the concept of tags used in social bookmarking and tags associated with images on Flickr, Chris Messina introduced the hashtag to Twitter in 2007. The hashtag, originally introduced in internet relay chat (IRC) in the late 1990s, was reworked and introduced onto the social media platform in 2007 to connect social conversations with back-end functionality programming. "Twitter uses Scala for its back-end programming. Hashtags are a piece of logic in the program that enables hashtags to be clickable and searchable. Once hashtags are used in a tweet, they are stored in a database and are searchable in Twitter's search API for about a week… Hashtags can also be used for data analytics. A very popular Twitter feature has been its lists of trending topics, comprising the most widely used hashtags, keywords, and conversation topics" (Alfonzo, 20).

Twitter serves as an example of a self-contained database with many folksonomies where users are both the creators and organizers of content. Twitter is not the only social platform that supports hashtags, other platforms that adopted hashtags are Facebook, Pinterest, Instagram, YouTube, Tumbler, with user bases in the millions. Twitter and other social platforms that support free-form categorization face similar challenges. "The fact that anyone can create a hashtags leads to several different hashtags describing the same thing. For example, the hashtags #martinlutherkingjr, #mlk, #drmartinluterking—the list goes on—have all been used when talking about Martin Luther King Jr…. While there are no formalized rules for hashtag use, Twitter recommends using no more than two hashtags per tweet" (Alfronzo, 21).

Folksonomy in Research

Tagger Motivation:

As presented in the research of Morgan Ames and Mor Naaman there are two different motivations for social tagging. The researchers produced a two-by-two model that identified the action of tagging by who the tag was intended for, social or self, and what the tag was intended to accomplish, organization or communication. The table featured to the right was construct in the research of Shilad Sen and John Riedl to simplify the original research of Ames and Naaman to identify the motivation behind tags. As stated earlier, because folksonomy is not a traditional taxonomy and does not abide by a strict set of rules it is important that these factors be taken into consideration when the data is consulted. At the University of Minnesota a GroupLens research group was established to survey users on why they tagged, in 2006 participants cited that they were "having fun" and "giving back to the community". "Experts suggest that tagging systems work because of a sort of reverse tragedy of the commons. Individual taggers



	Organization	Communication
Self	Browse, search (for example, labeling items by genre so I can browse them later)	Memory (for example, labeling DVDs by the number of stars I give the movie)
Social	Search, self-promotion (for example, labeling books to help others find them)	Description, self-expression (for example, labeling items to communicate, such as "wingnut welfare")

work for selfish reasons, labeling items for their own purposes…Meanwhile, the tags they add accidentally create social good for others, serving social purposes…" (Sen, 98).

Tagger Vocabulary and Discovery

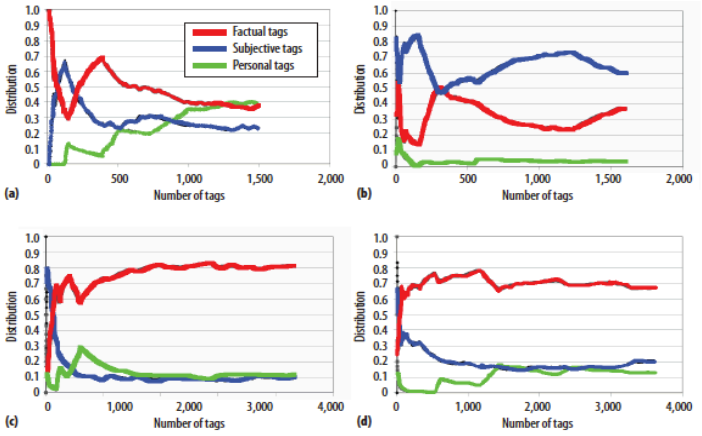
In a traditional system of indexing an end-user would be successful in finding images reflecting their interest only if their vocabulary matched that of the indexer. Though the concept is similar in a social platform, the example provided about numerous tags used to describe the presence of Martin Luther King Jr. in content shows that the opportunity for discovery is increased. A shared vocabulary with a professional cataloger indicates the end-users understanding of the official classification of the subject matter, but a shared vocabulary with a fellow tagger can signify shared interest or background. Professionals such as "Bearman and Trant (2005) recognize that 'we may be alienating a user community by not speaking their language.' Many practitioners feel that traditional document-oriented indexing techniques are insufficient for image indexing in the web environment…" (Matusiak, 287). Users participating in social platforms dependent on their motivation of tagging want to either share or remember content that they have found. The activity of social tagging creates a give and take relationship with all of the participants on the platform. The tagger provides data through the process of tagging and receives additional data through the review of other users' tags. "The interlinked system of tags supports browsing activities and serendipitous discovery of images in the digital environment. The most important strength of social tagging, however, is its close connection with the users and their language. Mathes (2004) points out that it directly reflects user 'choices in diction, terminology, and precision. The vocabulary is current and flexible as it quickly absorbs newly-created terms and neologism invented by web users" (Matusiak, 289).

Comment [d1]: Only in the sense that if a user happens to enter one of these idiosyncratic has tags she would find some relevant tweets but not all of them. Many users will not think of even one of these tags and not find anything. The solution is a system that collects all the synonymous hashtags and some additional synonymous terms that users are likely to use in searching into synonym sets and search for all the tags in the synonym set to which the user's search term belongs. Of course it would take some effort to create and maintain such a system.

Peer Pressure and Tags

Consideration has been given to the concept of social pressure in tagging environments. Studies have shown that the structure of the tagging interface can have a significant effect on the tags users create. A study conducted by the GroupLens team out of the University

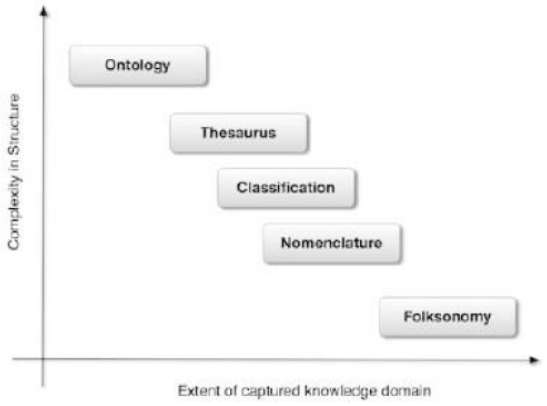
of Minnesota set up four tagging interfaces that used different algorithms to display data. The first interface was a platform that was considered “unshared”, the users would see only tags that they created, but no one else’s tags. The second interface featured a platform that was considered “shared”, users would see tags in their group in a random order. The third interface was considered “shared-pop”, meaning that users would see tags created in their group based off the tags popularity. The final interface featured “shared-rec” where users would see users tags that were created in their group and the system would also display additional tags that were of similar rankings. The study sought to find patterns in the way that users were creating tags. The GroupLens team categorized 11,443 tags by 635 users over the course of 100 days in factual, subjective, or personal classifications. The following graphs are a representation of their findings based on the use of the four platforms.



(a) unshared (b) shared (c) shared-pop (d) shared-rec

With the findings in this study the GroupLens team deducted that when taggers cannot see others tags besides their own they are not motivated to create unselfish tags. When the taggers were able to see the tags of others the number of personal tags drastically decreased in all three of the “shared” tagging interfaces. The team also found that, “Beginning taggers appeared to be more heavily influenced by community tagging norms, while experiences taggers’ behavior reflected their investment in a personal tag vocabulary. These findings suggest…users in a community perceive a particular tagging norm, they adjust their behavior to be more consistent with that norm and the new tags they create strengthen that norm” (Sen, 100).

Semantic Relations



"Classification of the Methods of Knowledge Representation with regard to Expressiveness and Coverage of the Knowledge Domain. Source: Adapted from Peters & Weller (2008, 101, Fig. 1)."

In social tagging environments users are able to apply free word association to images or content for their own use or they can add tags to benefit others. The act of creating free word association that is not tied to an ontology creates a broader "Weak methods of knowledge representation, such as folksonomies, can be applied to a large knowledge domain but are relatively inexpressive since they make no use of semantic relations. Any extension of knowledge organization systems from the right side of the spectrum to the left increases the semantics and thus the expressiveness of the vocabulary; however, its applicability to larger or separate knowledge domains is severely restricted…"(Peters, 122).

Research has gone further regarding the semantic web to bridge ontology and folksonomy. Tom Gruber originally introduced "Ontology of Folksonomy" and developed TagOntology. The concept behind TagOntology was to create ways to identify and formalize the tagging process. Gruber's process sought to create context in folksonomies by constructing clusters that would establish semantic relations between tags (Ungrangsi, 2011).

Conclusion

The process of digitization has brought images into Web 2.0 environments, but tagging has brought digital images into a social setting. Social tagging has enabled users to engage with images and data by adding their opinions and viewpoints. The theories first presented in this paper regarding social tagging represent low-level classification modalities for image indexing. The use of free-word association serves as a non-hierarchical classification that allows for singular connectivity. The development of semantic relations in folksonomies, and connection to ontologies represents high-level classification that rely on the context established by concepts such as clustering and co-occurrence.

By engaging users in a social context, the process of classification became a social activity; traditional classification never allowed indexers to engage with the end-user. The advantages of social tagging on image discovery are endless, user-generated metadata will constantly evolve with the changing trends and will remain connected to older tags for classification. With more images and data being created each second, information professionals and social taggers alike still have a lot to learn from each other.

Works Consulted

"Adaptive image classification based on Folksonomy." 2010.Web. */z-wcorg/.*

Alfonzo, Paige. "Using Twitter Hashtags for Information Literacy Instruction: The Ubiquity of Hashtags has Opened the Doors for the Teaching of Advanced Searching Concepts to a Much Wider Audience than in Past Years." 09; 2014/12 2014: 19+. *Information Science and Library Issues Collection; Gale.* Web. <http://go.galegroup.com/ps/i.do?id=GALE%7CA381665413&v=2.1&u=nm_p_newmex&it=r&p=PPIS&sw=w&asid=ea7aa41c96309fab31c2942c5b5e4bcc>.

Andrews, Pierre, and Juan Pane. "Sense Induction in Folksonomies: A Review." *Artificial Intelligence Review* 40.2 (2013): 147-74. Web.

Arms, Caroline R. "Getting the Picture: Observations from the Library of Congress on Providing Online Access to Pictorial Images." *Library Trends* 48.2 (1999): 379. Web.

Bar-Ilan, Judit, et al. "The Effects of Background Information and Social Interaction on Image Tagging." *Journal of the American Society for Information Science and Technology* 61.5 (2010): 940. Web.

Cantador, Iván, Ioannis Konstas, and Joemon M. Jose. "Categorising Social Tags to Improve Folksonomy-Based Recommendations." *Web Semantics: Science, Services and Agents on the World Wide Web* 9.1 (2011; 2010): 1-15. Web.

Carmel, David, et al. "Folksonomy-Based Term Extraction for Word Cloud Generation." *ACM Transactions on Intelligent Systems and Technology (TIST)* 3.4 (2012): 1-20. Web.

Cattuto C, Loreto V,Pietronero L,. "Semiotic Dynamics and Collaborative Tagging." *Proceedings of the National Academy of Sciences of the United States of America* 104.5 (2007): 1461-4. */z-wcorg/.* Web.

Cattuto, Ciro, Vittorio Loreto, and Luciano Pietronero. "Semiotic Dynamics and Collaborative Tagging." *Proceedings of the National Academy of Sciences of the United States of America* 104.5 (2007): 1461-4. Web.

Cox, Andrew M. "Flickr: A Case Study of Web2.0." *Aslib Proceedings* 60.5 (2008): 493-516. Web.

Fu, Wai-Tat, et al. "Semantic Imitation in Social Tagging." *ACM Transactions on Computer-Human Interaction (TOCHI)* 17.3 (2010): 1-37. Web.

Godoy, Daniela, Gustavo Rodriguez, and Franco Scavuzzo. "Leveraging Semantic Similarity for Folksonomy-Based Recommendation." *IEEE Internet Computing* 18.1 (2014): 48-55. Web.

Golbeck, Jennifer, Jes Koepfler, and Beth Emmerling. "An Experimental Study of Social Tagging Behavior and Image Content." *Journal of the American Society for Information Science and Technology* 62.9 (2011): 1750-60. Web.

Huang, Hong, and Corinne Jörgensen. "Characterizing User Tagging and Co-Occurring Metadata in General and Specialized Metadata Collections." *Journal of the American Society for Information Science & Technology* 64.9 (2013): 1878-89. Web.

Huang, Hong, and Corinne Jörgensen. "Characterizing User Tagging and Co-occurring Metadata in General and Specialized Metadata Collections." *Journal of the American Society for Information Science and Technology* 64.9 (2013): 1878-89. Web.

Kern, R., Granitzer, M.,Pammer, V.,. "Extending Folksonomies for Image Tagging." (2008): 126-9. */z-wcorg/*. Web.

Konkova, Elena, et al. "Social Tagging: Exploring the Image, the Tags, and the Game." *Knowledge Organization* 41.1 (2014): 57-65. Web.

Matusiak, Krystyna K. "Towards User-Centered Indexing in Digital Image Collections." *OCLC Systems & Services: International digital library perspectives* 22.4 (2006): 283-98. Web.

Oded N.O.V., Chen Y. E.,. "Why do People Tag? Motivations for Photo Tagging." *Commun ACM Communications of the ACM* 53.7 (2010): 128-31. */z-wcorg/*. Web.

Petek, Marija. "Comparing User-Generated and Librarian-Generated Metadata on Digital Images." *OCLC Systems & Services: International digital library perspectives* 28.2 (2012): 101-11. Web.

Peters, Isabella.,Becker, Paul.,. "Folksonomies indexing and retrieval in Web 2.0." 2009.Web. */z-wcorg/*.

Pink, Daniel H. "Folksonomy." *The New York Times Magazine* 11 Dec. 2005: 69(L) *Literature Resource Center*. Web. 20 Dec. 2014.

Sen S., Riedl J.,. "Folksonomy Formation." *Computer Computer* 44.5 (2011): 97-101. */z-wcorg/*. Web.

Soergel, Dagobert,. "Indexing and Retrieval Performance: The Logical Evidence." *ASI Journal of the American Society for Information Science* 45.8 (1994): 589-99. */z-wcorg/*. Web.

Stvilia, Besiki, Corinne Jörgensen, and Shuheng Wu. "Establishing the Value of Socially-Created Metadata to Image Indexing." *Library and Information Science Research* 34.2 (2012): 99-109. Web.

Stvilia, Besiki, and Corinne Jörgensen. "User-Generated Collection-Level Metadata in an Online Photo-Sharing System." *Library and Information Science Research* 31.1 (2009): 54-65. Web.

Trant, J.,with the participants in the steve.museum project,. "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept." *New Review of Hypermedia and Multimedia* 12.1 (2006): 83-105. */z-wcorg/*. Web.

Ungrangsi, R., and C. Anutariya. "A Comparison Study of Flickr's Folksonomies and Ontologies". *Computer Science and Software Engineering (JCSSE), 2011 Eighth International Joint Conference on*. Web.

Willey, Eric, "A cautious partnership: The growing acceptance of folksonomy as a complement to indexing digital images and catalogs" (2011). *Faculty and Staff Publications – Milner Library.* Paper 57.

Wu L., Jin R.,Jain A.K.,. "Tag Completion for Image Retrieval." *IEEE Trans Pattern Anal Mach Intell IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.3 (2013): 716-27. */z-wcorg/*. Web.

Yang X.S., Cheng R., Mo L., Kao B., Cheung D.W., 29th International Conference on Data Engineering,ICDE 2013,. "On Incentive-Based Tagging." *Proc Int Conf Data Eng Proceedings - International Conference on Data Engineering* (2013): 685-96. */z-wcorg/*. Web.