Utah State University

DigitalCommons@USU

Instructional Technology and Learning Sciences Research

Instructional Technology & Learning Sciences

10-1-2009

Changing Higher Education Learning with Web 2.0 and Open Education Citation, Annotation, and Thematic Coding Appendices

Heather Leary
Utah State University

M. Harrison Fitt
Utah State University

David Wiley
Brigham Young University - Utah

Follow this and additional works at: https://digitalcommons.usu.edu/itls_research

Part of the Educational Assessment, Evaluation, and Research Commons, and the Instructional Media Design Commons

UtahState University
MERRILL-CAZIER LIBRARY

Appendix A: Del.icio.us (del.icio.us)

Researcher: Bobbe Allen <bobbemcghie@gmail.com>

Literature Review

Reference	Synopsis/Findings	Themes
Al-Khalifa, H. S. and Davis, H. C. (2006). Folksonomies versus automatic keyword extraction: An empirical study. In IADIS Web Applications and Research 2006 (WAR2006), 15-19 May.	The assumption in this work is that folksonomies carry more semantic value than keywords extracted by machines. They measured the percentage overlap between the folksonomy set and a machine generated keyword set in an automatic sense; and by asking a human indexer to evaluate the same. The result of the experiment can be considered as evidence for the rich semantics of folksonomies, demonstrating that folksonomies such as del.icio.us can be used in the process of generating semantic metadata to annotate web resources. After completing the three phases of the experiment it is clear from the results that the folksonomy tags agree more closely with the human generated keywords than the automatically generated ones. In addition, the purpose of this experiment was satisfied by proving that folksonomies can be semantically richer than the keywords extracted using a major search engine service like Yahoo. The experiment also showed the percentage of overlap between folksonomies and automatically extracted keywords for a given website. (Source: http://eprints.ecs.soton.ac.uk/14292/1/IADIS_FvK_2006.pdf)	Folksonomy Keywords Searching Machine vs. human Semantic metadata Semantic value

Reference	Synopsis/Findings	Themes
Al-Khalifa, H. (2007). Automatic document-level semantic metadata annotation using folksonomies and domain ontologies. Unpublished doctoral dissertation, University of Southhampton, United Kingdom.	The growth of user-generated metadata presents many opportunities from which researchers can benefit. This study utilizes delicious data to produce semantic metadata. The results demonstrated that tags are good enough for creating semantic metadata; an analysis of tagging behavior showed that tags include formal and informal metadata; tags may not be perfect or complete but they have potential; semantic web technologies have contributed to retrieval of learning resources; tags are better than automatically generated keywords; and tags show the power of producing meaningful metadata by aggregating people's intelligence and without requiring their consensus in choosing tags. (Source: http://citeseerx.ist.psu.edu/viewdoc/summary;jsessionid=99BA503930 AC62D8CC9256D01CD56C17?doi=10.1.1.106.9841)	User metadata Semantic metadata Tag credibility Tagging behavior Machine vs. human Semantic web
Al-Khalifa, H.S. And Davis, H.C. (2006). Delicious learning resources. In: International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering, December 4 - 14, Online.	This research uses keywords generated by folksonomies and ontology based semantic annotations to annotate learning resources. The system does not use an algorithm to extract keywords and it generates semantic metadata not standard metadata. They used del.icio.us resources to show that folksonomies, guided by domain ontologies can be used to generate semantic metadata. (Source: http://eprints.ecs.soton.ac.uk/13197/)	Keywords Folksonomy Semantic metadata Ontology
Bateman, S., Brooks, C.H., and McCalla, G. (2006). Collaborative tagging approaches for ontological metadata in adaptive e-learning systems. In proceedings of the Annual International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL), June 20,	Semantic web ontologies are being used to create metadata for e-learning resources. This paper proposes a method of handling shortfalls created by social bookmarking. They experiment with del.icio.us data. The model facilitates users such as students to create machine consumable metadata about learning objects. It is called CommonFolks Tools and uses annotations that can be made with terms that exist in WordNet (a large lexical database of English where nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms that express a concept). (Source: http://www.win.tue.nl/SW-EL/2006/swel06-ah06-program.html)	Semantic web Ontology User metadata Machine metadata Social bookmarking

Reference	Synopsis/Findings	Themes
Binkowski, P.J. The effect of social proof on tag selection in social bookmarking applications. A Master's Paper for the M.S. in I.S. degree. December, 2006.	The social psychological factors of users tag choices are examined in this paper using del.icio.us data. Social proof is the principle applied to social bookmarking because it can predict that tags will be more similar to each other if the users are provided a list of suggestions. If web content is confusing to users, the effects of social proof becomes pronounced. Knowing this has the potential for improving the usability of collaborative tagging systems. Users can improve their search for similar content because tags will have the same name. (Source: http://etd.ils.unc.edu/dspace/bitstream/1901/358/1/philipbinkowski.pdf)	User tags Social bookmarking Collaborative tagging
Campbell, D.G. (2006). A phenomenological framework for the relationship between the semantic web and user-centered tagging systems. Advances in classification research, Vol. 17: Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop (Austin, TX, November 4, 2006), ed. Jonathan Furner andJoseph T. Tennis.	Husserl's theory of phenomenology is used to discover a relationship between user-centered tagging systems and the Semantic Web, which in contrast is a highly structured system. They found that the Semantic Web resembles the traditional library system while social tagging works on implied principles of emergence. 'Perhaps the most painful limitation is the glib equation of consciousness, as Husserl envisions it in his work, with the Web as a collective mind, in which outward and inward movement, phenomena and intentionalities are manifested in Web resources and Web behavior.' The Semantic Web is highly structured and elaborate but based on decisions that are machine understandable. 'The surprising, even absurd patterns that emerge through social tagging will form a necessary counterpoint to this infrastructure….'	User tags Semantic web Tagging patterns Library

Reference	Synopsis/Findings	Themes
Cattuto, C. Loreto, V. & Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. In proceedings of the National Academy of Sciences of the United States of America, 105, 1461-1464.	The authors investigate the statistical properties of tag cooccurrence utilizing a stochastic modeling of behavior by the user. There are two aspects of collaborative tagging "1) A frequency-bias mechanism related to the idea that users are exposed to each other's tagging activity; 2) a notion of memory, or aging or resources, in the form of a heavy-tailed access to the past state of the system." They found that even though our cognitive processes are complex and there is no global coordination of tagging activity there is a universal behavior between users. (Source: http://www.pnas.org/cgi/doi/10.1073/pnas.0610487104)	Tagging behavior Tag occurrence Collaborative tagging
Cattuto, C., Baldassarri, A., Servedio, V. D.P., Loreto, V. (2007). Vocabulary growth in collaborative tagging systems.	How does the number of different tags grow as a function of a suitably defined notion of time? Using del.icio.us data the researchers attempt to answer this question. They look at the number of distinct tags as well as the evolution of local vocabularies. At the global level they found a power-law growth with exponents smaller than one, of the number of distinct tags. For tagging with more popularity, they found a 'sub-linear growth with exponents sharply peaked around a characteristic value (slightly different from the global one), while for less popular resources' they observed 'exponent values slowly shifting towards 1.' Their work brought up many more questions and possible problems that could impact future designs of bookmarking systems. (Source: http://arxiv.org/PS_cache/arxiv/pdf/0704/0704.3316v1.pdf)	Tag growth Social bookmarking Tagging behavior
Cattuto, C., Loreto, V., and Pietronero, L. (2006). Collaborative tagging and semiotic dynamics.	Using del.icio.us data, the researchers investigate the statistical properties of tag co-occurrence. They use a stochastic model of user behavior to look at data. They discovered that users have a universal behavior pattern despite the selfish and uncoordinated nature of social tagging. They believe this is a starting point for future studies that seek to understand, predict and control the 'Semiotic Dynamics of online social systems.' (Source: http://arxiv.org/PS_cache/cs/pdf/0605/0605015v1.pdf)	Tag occurrence Tagging behavior

Reference	Synopsis/Findings	Themes
Chi, E.H. And Mytkowicz, T. (2007). Understanding navigability of social tagging systems. In proceedings of CHI.	This study looks at the efficiency of 'organically evolved vocabulary' in navigating to information sources now that social tagging systems are gaining much popularity. They used a custom web crawler and screen scraper to collect data, computing the frequency and probability distribution of documents being bookmarked. They used various kinds of entropy to discover that del.icio.us is getting hard to navigate, that users are having a harder time tagging as the bookmark collections grow. (Source: http://www.viktoria.se/altchi/submissions/submission_edchi_0.pdf)	User tags Social bookmarking Navigation
Golder, S. A., and Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. Journal of Information Science, 32 (2) pp.198-208.)	The authors analyze the structure and dynamic aspects of collaborative tagging systems. They found that users tag in variety of ways, some use many tags some use just a few. However, stable patterns did emerge with popular tags coexisting with minority tags without disrupting the stable consensus. Also, tagging is done for personal benefit not public benefit. Delicious functions as a recommendation system with tagging such as to read.	Structure Tagging patterns Tagging behavior
Golder, S.A. & Huberman, B.A. (2005). The structure of collaborative tagging systems.	The authors analyze the dynamic aspects and structure of collaborative tagging systems using two sets of Delicious data. One set was comprised of popular URLS's in a given time period and the other set was a random sampling of the people or users in the same time period. They found that some users use many tags and others just a few tags. There is a large variety among sets of tags, the frequency in which they are used and in what they describe. However, across all tagging a stable pattern does emerge. It also appears that users tag for personal reasons, not public. The authors feel that looking at users choices on a large scale could be useful for describing and organizing how web documents interact with one another. (Source: http://arxiv.org/abs/cs/0508082)	Structure Tag occurrence Tagging patterns

Reference	Synopsis/Findings	Themes
Guy, M., and Tonkin, E. (2006). Folksonomies: Tidying up tags? D-Lib Magazine 12(1).	They question the sloppiness of folksonomy tagging and look for solutions to the problem by creating a system for searching, sorting and classifying. Then they begin to question the approaches that would 'tidy up' folksonomies and find that the changing of folksonomies into tidy neatness has the potential risk of 'losing the very openness that that has made folksonomies so popular.' The authors suggest that the chaotic character of tagging is not the problem but the fact that they are suppose to fulfill two very different things, tagging for the personal collection and for the collective collection. The idea of training the user to use a restricted choice of tags may defeat the very purpose of a folksonomy. It would be applying a destructive solution. They also suggest that there is possibly more to folksonomy tagging and researchers need to delve deeper into it rather than find ways to 'tidy it up.' (Source: http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/january06/guy/01guy.html)	Folksonomy Tagging patterns Openness Tagging behavior
Halpin, H., Robu, V., and Shepherd, H. (2006). The dynamics and semantics of collaborative tagging.	Concerned about cohesion and consensus within distributive tagging systems, the authors investigate the distribution of tag frequencies and whether or not they are stable. They attempt to understand the basic dynamics that go into collaborative tagging. Using empirical data they describe the tagging distributions as stabilizing into power law distributions. They suggest that this means there is consensus, by tagging behaviors, around the categorization of information and that the evidence suggests that tagging has a type of latent classification scheme and taxonomic structure. They also suggest that folksonomies and ontologies are 'not mortal enemies' because tag-based categorization can evolve into stable classification schemes which can then be formalized as ontologies. (Source: http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-209/saaw06-full01-halpin.pdf)	Tagging consensus Tagging patterns Tagging behavior Classification scheme Folksonomy Ontology

Reference	Synopsis/Findings	Themes
Han, P., Wang, Z., Li, Z., Kramer, B., and Yang, F. (2006). Substitution or Complement: An Empirical Analysis on the Impact of Collaborative Tagging on Web Search. Web Intelligence archive. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence.	An empirical study was conducted using data collected from del.icio.us. They were looking at how users descriptions of web resources could be used to enhance the existing web search paradigm. Users annotations to web pages were categorized into search and exploration keywords then analyzed and experimented with. Based on the data, they propose a framework for web searches that can increase the accuracy of a search substantially. (Source: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4061467)	User annotations Keywords Searching
Hassan-Montero, Y., and Herrero-Solana, V. (2006) Improving tag-clouds as visual information retrieval interfaces. In proceedings of the International Conference on Multidisciplinary Information Sciences and Technologies, Merida, Spain, October 25-28.	This paper presents a novel approach to Tag-Cloud's tags selection, and proposes the use of clustering algorithms for visual layout, with the aim of improve browsing experience. The results suggest that presented approach reduces the semantic density of tag set, and improves the visual consistency of Tag-Cloud layout. (Source: http://nosolousabilidad.com/hassan/improving_tagclouds.pdf)	Tag selection Clustering Browsing Semantic density Visual consistency
Heymann, P. and Garcia-Molina, H. (2006). Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. InfoLab Technical Report, 2006-10.	Investigation led the authors to discover an algorithm for 'converting a large corpus of tags annotating objects in a tagging system into a navigable hierarchical taxonomy of tags.' Their system worked with Delicious because it is 'high density, high overlap' and more even specificity distribution. They also determined that graph centrality is as valid a way of determining importance in collaborative tagging systems as it is in social networks. (Source: http://labs.rightnow.com/colloquium/papers/tag_hier_mining.pdf)	Tagging characteristics Collaborative tagging

Reference	Synopsis/Findings	Themes
Hotho, A., Jaschke, R., Schmitz, C., and Stumme, G. (2006). Emergent semantics in BibSonomy.	This paper uses data from del.icio.us to perform experiments for taking first steps towards emergent semantics. By adapting data mining techniques and information retrieval approaches for detection, they specify a formal model for folksonomies. At first they exploited and enhanced existing algorithms, but felt that more sophisticated ones are need with the growing amount of users and relationships between tags. They discuss the need for more structure than just flat tagging, but a structure that the user does not have to maintain. It would be the application of an ontology learning technique. They extend their knowledge in context of BibSonomy. Source: http://www.kde.cs.uni-kassel.de/stumme/papers/2006/hotho2006emergent.pdf)	Semantics Data mining Folksonomy Structure Tagging
Hotho, A., Jaschke, R., Schmitz, C., and Stumme, G. (2006). Trend detection in folksonomies. In proceedings of the 1st Workshop on Applications of Semantic Technologies, Dresden, Germany, October 6.	This paper looks at how difficult it is to keep up with ones own interests as the number of resources on the web exceeds ones ability to track them all. Social bookmarking is overcoming this bottleneck in knowledge acquisition. They use data from del.icio.us to experiment with their idea of discovering topic-specific trends within folksonomy-based systems. Their approach does not rely on the internal structure of documents but allows for all data types. It is felt that in the future they can predict how a folksonomy will change its structure during growth, that they can predict communities that are stable or volatile and can generate recommendations. (Source: http://www.tagora-project.eu/wp-content/2007/05/hotho2006trend.pdf)	Tracking resources User interests Social bookmarking Folksonomy Predictible communities Recommenders

Reference	Synopsis/Findings	Themes
Hotho, A., Jaschke, R., Schmitz, C., and Stumme, G. (2006). Information retrieval in folksonomies: search and ranking.	This paper looks at del.icio.us in its early days. Success was attributed to ease of use but the information retrieval support was considered limited. The researchers used data from del.icio.us and created a search algorithm, FolkRank, to exploit its structure. They compare it to a Google-like search within folksonomy based systems. They argued enhanced search facilities are of primal importance for emergent semantics within folksonomies. Users and community members can more easily get to know each other and the resources created if they are extracted from the whole and made more explicit. They suggest that spam and link farms will cause problems in the future as social bookmarking systems grow. (Source: http://www.kde.cs.uni-kassel.de/hotho/pub/2006/search2006hotho_eswc.pdf)	Usability Searching Folksonomy Semantics Social bookmarking
Kipp, M.E.I, and Campbell, D.G. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. In proceedings of the Annual American Society for Information Science & Technology Conference, Austin, Texas, November 3-8.	The work examines how users analyze tagging patterns to see if their behavior supports and/or enhances the traditional ways of document classification and indexing. They discovered that tagging practices do emulate the traditional forms with small numbers of tags emerging through 'unspoken consensus.' When inconsistencies occurred, they discovered that they were actually predictable and could be anticipated. They also discovered a practice between users relating to time and task which could lead to a new way of modeling subject access. (Source: http://eprints.rclis.org/archive/00008315/fullmetadata.html)	Tagging patterns User behavior Classification Tagging consensus Tagging behavior

Reference	Synopsis/Findings	Themes
Klerkx, Joris and Erik Duval. (2007). Visualizing social bookmarks. In proceedings of the 1st Workshop on Social Information Retrieval for Technology-Enhanced Learning & Exchange, Sissi, Lassithi - Crete Greece, 18 September.	The researchers designed a visualization application for social bookmarks using del.icio.us data. The design is meant to enhance the implicit structures, visually, between tags, users and bookmarks. The underlying belief here is that users would benefit from seeing other bookmarks that might be of interest to them. Also, this design would benefit researchers who study social information to see how people use retrieval tools (Source: KU Leuven; http://www.cs.kuleuven.ac.be/~hmdb/infovis/deliciousdel.iciou.us%20visualization_files/Visualizing_bookmarks_Klerkx.pdf	Social bookmarking Visual consistency
Kome, S.H. (2005). Hierarchical subject relationships in folksonomies. Unpublished Masters thesis, University of North Carolina, Chapel Hill.	The author looks at folksonomies for retrieval and organization effectiveness by analyzing folksonomy metadata for hierarchal semantic relationships. The results showed a hierarchical relationship in folksonomies. He also discusses the potential value between the development of folksonomies and thesauri developers. (Source: http://hdl.handle.net/1901/238)	Folksonomy Structure Semantic metadata
Lee, Kathy J. (2006). What goes around comes around: An analysis of del.icio.us as social space. In Proceedings of the Conference on Computer Supported Cooperative Work, November 4-8, Banff, Alberta, Canada.	An analysis is made of del.icio.us users to investigate the effects of perceived social presence and the impact it might have on their posting behavior. They used a linear regression model to discover that when a user is aware of others they subscribe to other bookmarks, sign their homepages, and are more likely to include annotations that would be helpful to others. The researcher feels that this knowledge could help the designers of information sharing mechanisms and knowledge repositories. The del.icio.us model blends creates a benefit to the collective efforts as well as the personal goals of users. (Source: http://portal.acm.org/citation.cfm?id=1180875.1180905&coll=GUIDE&dl=ACM&CFID=28764231&CFTOKEN=39793131)	User behavior Social bookmarking User interests User annotations

Reference	Synopsis/Findings	Themes
Lin, X., Beaudoin, J.E., Bui, Y. and Desai, K. (2006) Exploring characteristics of social classification. Advances in classification research, Vol. 17: Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop (Austin, TX, November 4, 2006), ed. Jonathan Furner and Joseph T. Tennis.	This work provides three empirical studies on the characteristics of social classification. The first study found little overlap between social tags and controlled vocabularies and title-based automatic indexing. The second study looked at whether tags could be categorized to improve searching and browsing. And the third study wanted to see what led to the most significant impact on tag convergence. The study discovered that when the number of users increases the tags become similar to controlled vocabulary indexing more than automatic indexing. Second, it is feasible to categorize tags into meaningful and stable groups, which has implications for searing and browsing. And third, the convergence of tags could be predicted at the 30/70 ratio rather than the expected 20/80 ratio. These studies lead to the usefulness of studying social classification characteristics. (Source: http://dlist.sir.arizona.edu/1790/01/lin.pdf)	Classification characteristics Social tagging Controlled vocabularies Tag consensus Tagging User growth
Marinchev, I. (2006). Practical semantic web – tagging and tag clouds. Cybernetics and Information Technologies, 6(3).	This study looks at the incentives and trends of tagging and tag clouds utilizing del.icio.us users. He defines the users he looks at as smart – someone who has found more engaging material than he has found on a certain subject. He uses RSS to 'intellectually draft' off that person and found that some users where intellectually drafting from his tagging. The conclusion of his paper discusses various ideas about tagging. He talks about an inverted tag in which the tag will depict all of the focuses attached to it. He also discusses search engines that will look for tags instead of web pages. He suggests that a tag could spawn a magnifying lens of related tags. (Source: http://www.iit.bas.bg/Cit_en/CIT_06_en/v6-3/33-39.pdf)	Tagging trends User incentive User characteristics

Reference	Synopsis/Findings	Themes
Mason, B.L. & Thomas, S. (2007). Tags, network's, narrative: Exploring the use of social software for the study of narrative digital contexts. In proceedings of the 18th conference on Hypertext and Hypermedia, Manchester, UK.	Utilizing del.icio.us participants who are tagging 40 different websites, the researchers looked at the role and utility of using folksonomy in trans-disciplinary communication. In their analysis, they saw how participants faced the dilemma of who does one tag for, self or a broader, unknown audience. They found that tagging by users is heavily affected by the cultural context. In the several questions they asked, they hope to inform the theory and practice of use of these technologies across disciplines. (Source: http://delivery.acm.org/10.1145/1290000/1286252p39-mason.pdf) (Keywords: Folksonomy, Hypertext, Narrative, Social Software, Tagging, Transliteracy, Web2.0)	Tagging Folksonomy Tagging behavior Cultural context
Mika, Peter. (2007) Ontologies are us: A unified model of social networks and semantics. Web Semantics: Science, Services and Agents on the World Wide Web, 5(1), 5-15.	This paper argues that Semantic Web ontologies produced by knowledge engineers do not match the ontological drift that occurs over time within a community. They propose an emergent semantics by defining the 'ontology as an emergent feature of a system of autonomous agents acting in dynamic, open environments.' Using del.icio.us to investigate the semantics of a large scale social network, they show how to enrich the representation by extracting clusters of related concepts and taxonomical relationships. Then they apply their ideas toward extracting community ontologies. Using co-occurrence analysis they evaluate it against the results of traditional web mining. Mika's research suggests that combining the actors and their concepts in ontologies creates an important and sustainable structure. (Source: http://www.google.com/search?ie=UTF-8&oe=UTF-8&sourceid=navclient&gfns=1&q=%3A+http%3A%2F%2Fwww.cs.vu.nl%2F%7Epmika%2Fresearch%2Fpapers%2F)	Semantic web Ontology Taxonomy Sustainability

Reference	Synopsis/Findings	Themes
Paolillo, J.C. and Penumarthy, S. (2007). The social structure of tagging internet video on del.icio.us. In proceedings of the Annual Hawaii International Conference on System Sciences.	This work analyses the social network of Internet video tagging in del.icio.us. They discover that tagging by a tight group of users is coherent with not much re-use of tags across different content or many users that tag more than a few similar items. This suggests that tagging is highly local and not likely to be useful in global navigable classification systems. Their findings suggest that problems exist in folksonomic tagging, because they are weaker than expected probably due to uncontrolled vocabulary applied by non-specialists. Alternative versions are used extensively for the same data and there is a problem in that different copies of the same video are linked by different users. Thesauri tag suggestions might prove useful. The small group users seem to influence a large share of the semantic range. How do outsiders come into contact with this group, and would they understand the tagging conventions used? Is this an impediment or a resource for information organization? (Source: http://csdl2.computer.org/comp/proceedings/hicss/2007/2755/00/27550085b.pdf)	Tagging behavior Folksonomy Controlled vocab Classification
Plangprasopchok, A. and Lerman, K. (2007). Exploiting Social Annotation for Automatic Resource Discovery.	The resource discovery task relies on a user to manually discover and integrate information resources. Using data from del.icio.us, they describe a probabilistic model of the user annotation process to find resources in a particular domain. They find that this is a great method for automating the resource discovery task and describe future work in this area. (Source: http://arxiv.org/PS_cache/arxiv/pdf/0704/0704.1675v1.pdf)	User annotation Resource location Searching

Reference	Synopsis/Findings	Themes
Schmitz, C., Hotho, A., Jaschke, R. and Stumme, G. (2006). Mining association rules in folksonomies.	Delicious was used as a dataset to demonstrate how association rule mining could be used to analyze and structure folksonomies. They share the belief that the results can be used for a variety of purposes such as 'recommending tags, users, or resources, populating the supertag relation of the folksonomy, and community detection. They also include ideas for future work using 'association rule mining, FolkRank ranking, and graph clustering.' (Source: http://www.kde.cs.uni-kassel.de/hotho/pub/2006/schmitz2006asso_ifcs.pdf)	Folksonomy Association Community detection Recommendors
Specia, L. and Motta, E. (2007) Integrating folksonomies with the semantic web. In proceedings of the Annual European Semantic Web Conference, Innsbruck, Austria, June 3-7.	The researchers feel that there is more to collaborative tagging systems than just searching and navigating resources. A collective classification schema is possible when more than one person uses the same tags. They use both del.icio.us and Flickr data and report that their experiment is feasible and very promising. Using clustering techniques and co-occurrence analysis and derived with meaningful groups of tags which correspond to concepts in ontologies. Querying ontologies could reveal relationships within tags in each cluster. (Source: http://www.eswc2007.org/pdf/eswc07-specia.pdf)	Collaborative tagging Classification schema Clustering Ontology

Reference	Synopsis/Findings	Themes
Spiteri, L. F. (2007). Structure and form of folksonomy tags: the road to the public library catalogue. Webology, 4(2).	This paper looks at the structure of tags used in folksonomies. The folksonomy sites used were Del.icio.us, Furl, and Technorati. They were evaluated against the National Information Standards Organization (NISO). Tags were collected over a thirty day period and found that they corresponded closely to NISO guidelines in the 'concepts expressed by the tags, the predominance of single tags, the predominance of nouns, and the use of recognized spelling.' Problem areas were the inconsistent use of singular and plural forms of count nouns, homographs and unqualified abbreviations or acronyms. The conclusion was that folksonomies should be used by librarians to increase user-friendliness, interactivity of public library catalogues and could also be used to encourage more activities such as informal online communities of readers and advisory services for user-driven readers. (Source: http://www.webology.ir/2007/v4n2/a41.html) (Keywords: Collaborative tagging; Controlled vocabularies; Folksonomies; Guidelines)	Tag structure Folksonomy Tagging Tagging characteristics
Szekely, B. & Torres, E. (2005). Ranking bookmarks and bistros: Intelligent community and folksonomy development. Unpublished.	Using delicious for datasets, researchers looked at tagging as a possible mechanism for rating a restaurant, for instance. Sharing and classification of information are parts of the fabric of semantic techniques just as they are in online communities. The investigators used UserRank (modeled after Google's PageRank system) and TagRank showing that these algorithms do represent the community consensus. They conclude that caution should be taken to make sure the ranking methodologies remain trustworthy and accurate, alluding to the possibility of advertisers and spammers finding a way to manipulate the data. They used a series of ranking algorithms that they feel can be applied to any online community. (Source: http://torrez.us/archives/2005/07/13/tagrank.pdf)	Tagging Semantic techniques Ranking

Reference	Synopsis/Findings	Themes
Tonkin, E. (2006). Searching the long tail: Hidden structure in social tagging. In proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop, Austin, TX, November 4.	The authors look at a way to take apart compound tags in order to improve search indexes, 'extraction of semantic information' and benefits to usability. They analyze tagging habits, which leads them to conclude that social tagging systems have both a formal component of metadata as well as an informal one which utilizes annotations and descriptions. They found that a majority of tags are informal. They think there needs to be a way to improve searching across data and developed an approach using a sample set In their conclusion they demonstrated a certain class of multi word compound terms, segmented compound terms and concatenating words could be retrieved from component terms; but they question the value.	Tagging characteristics Semantic metadata Usability Tagging behavior
Trevino, E. M. (2006). Social bookmarks: personal organization and collective discovery on the web. Unpublished Masters thesis, University of Illinois at Chicago.	This is a qualitative study that focuses on the users of del.icio.us. Interviews as well as a content analysis were employed in order to discover how users understood the information and the site structure. The researcher discovered that tagging for individuals is used as a memory aid and signifies web pages a user values. When looking at the aggregate of taggers, you can see group formations around an interest as well as identify social trends. Trevino suggests that the data indicates the posts of users 'often derived significance from prior knowledge of the individual.' She also comments on how the design of del.icio.us balances the tensions between the personal and the public. (Source: http://blog.erickamenchen.net/Trevino-SocialBookmarking2006.pdf)	Users Tagging Classification Structure

Reference	Synopsis/Findings	Themes
Wash, R. and Rader, E. (2007). Public bookmarks and private benefits: An analysis of incentives in social computing. Unpublished paper.	This is a case study that looks at the incentives in social computing within the del.icio.us user community. At the basis is the belief that tags are everything. They are how users organize their bookmarks and how users who are seeking new information find it. The question becomes how can designers of software applications induce users to produce information that is useful for others? They found that metadata which reflects who bookmarked the page helps the user seeking information than 'free-form keyword metadata (tags).' They attribute this find to the way del.icio.us motivates users to provide tagging which has personal benefits. (Source: http://bierdoctor.com/papers/delicious-incentives-asist-submitted.pdf)	Incentives Tagging Tagging characteristics
Wash, R. and Rader, E. (2006). Tagging with del.icio.us: Social or Selfish? In proceedings of CSCW'06, November 4-8, 2006, Banff, Alberta, Canada	Is tagging done selfishly or socially and what are the implications? Past research shows that less than 20% of people will generate the same tag for the same object. Selfish tagging is done when the user gives no thought to the public audience. Selfish tagging creates a diversity of tags; however, such diversity actually enhances findability but at the expense of 'community convergence.' They use data from del.icio.us to answer the question about whether or not there is a 'vocabulary problem' in the way users select tags for web pages. Because del.icio.us is able to store a large number of tags for any object, the user can bypass the problem of tagging inconsistency so that there is a greater probability of successfully finding the web page they seek. The researchers feel that you can overcome the shortfall inherent in selfish tagging by using a human or algorithmic indexer, or using 'editorial control over tag synonyms and usage' in order to eliminate the tags that are inconsistent while keeping the tags that are rare. (Source: http://bierdoctor.com/papers/Rader_CSCW_Abstract_Final.pdf)	Social tagging Tagging characteristics Editorial control

Reference	Synopsis/Findings	Themes
Wu, X., Zhang, L., Yu, Y. (2006). Exploring Social Annotations for the Semantic Web. In proceedings of the 15th International conference on World Wide Web, Edinburgh, Scotland.	There is much interest today in the Semantic Web, or web resources that can be read by machines. For this to happen, web resources would have to be annotated with metadata that is understandable by a machine. This paper identifies three areas where an ontology for a large number of web resources is difficult: producing a common ontology with so many different people and applications with varying viewpoints; the vocabularies and web resources change very quickly; in order to use ontologies one must have skill in ontology engineering (which most normal web users may not have). Social annotations are looked at which are freely chosen by the user "without any a-priori dictionary, taxonomy, or ontology to conform to." They define social as a large number of web users with no ontology between them. It reflects the dynamics of the vocabulary of the moment and how it grows. They analyze delicious data quantitatively in order to show that you can infer things statistically from emergent semantics. This is a bottom-up approach. They analyze the annotation data using a probabilistic generative model. Their experiment shows how a global semantic model can be statistically inferred using folksonomies used to annotate web resources. They feel that this model helps to disambiguate tags and group synonymous tags together in concepts. And they demonstrate that even if web resources are not tagged by the query tags and do not contain any query keywords, the emergent semantics can be used to search and discover semantically-related web resources. (Source: http://www.manticmoo.com/bib/uploads/pdf/4071.pdf)	Semantic web Machine metadata Ontology Vocabularies Dynamic vocabularies Folksonomy Tagging characterstics

Reference	Synopsis/Findings	Themes
Yanbe, Y., Jatowt, A., Nakamura, S., and Tanaka, K. (2007). Can social bookmarking enhance search in the web? In proceedings of the Annual Joint Conference on Digital Libraries, Vancouver, BC, Canada.	Experimenting on datasets from del.icio.us, the researchers investigate enhancing Web searches. They discuss the advantages of social bookmaking data: high dynamics, attached metadata, available temporal and sentiment information, etc.' They combine the link structure ranking method with social bookmarking and find that incorporating popularity statistics improves page quality, relevance is more precise, there is a time awareness in popularity measures and temporal queries can be constructed. All of this leads to tags filtering pages by 'user impressions, sentiment characteristics or controversy levels.' They also conducted several analytical studies between PageRank and SBRank to support their approach. They concluded that a hybrid enhanced Web search is possible and provides advantages (Source: ACM Digital Library; http://delivery.acm.org/10.1145/1260000/1255198/p107-yanbe.pdf?key1=1255198&key2=9179541021&coll=GUIDE&dl=GUIDE&CFID=51957233&CFTOKEN=17577367)	Searching Social bookmarking Ranking User generated tags

Search Methodology

491 Articles reviewed/40 were useful

Name of site	Records Retrieved	Search term	Usable records
CiteULike	127	delicious	13 good
Digital Dissertations	0	Delicious tagging	0
	0	Del.icio.us	0
	0	Del.icio.us tagging	0
	0	Delicious social bookmarking	0
	1	Social bookmarking	
Education Full Text:	9	del.icio.us	descriptive
	14	Bookmarks websites	
	1	Bookmarks/websites internet resources	
ERIC	5	Social bookmarking	descriptive
	0	Del.icio.us	0
ERIC via EBSCO Host	26	Delicious	0
	0	Delicious tagging	0
	9	Social software	0
	4	Social bookmarking	0
	0	Bookmarks website	0
Google Scholar	11,200 viewed 276	Delicious tagging	27 good
http://reference.igi-online.com/search/results.asp	0	delicious	0
	0	Social bookmarking	0
	0	Delicious tagging	0
	0	Del.icio.us	0
	0	Del.icio.us tagging	0
Psychology & Behavioral Science collection	19	Del.icio.us	0

Appendix B: Facebook (**www.facebook.com**)

Researcher: Melynda Fitt <melynda.fitt@gmail.com>

Literature Review

Reference	Synopsis/Findings	Themes
Acquisti, A., & Gross, R. (2006). Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In *Privacy: 6th International Workshop, PET 2006, Cambridge, UK, June 28-30, 2006, Revised Selected Papers* (Vol. 4258/2006, pp. 36-58). Berling/Heidelberg: Spring.	The authors found that age is a strong predictor of Facebook membership. Non-members tend to be older (a mean of 30 years versus a mean of 21) but their age is also more broadly distributed (sd 8.840476 vs. sd 2.08514). "Privacy policy" was on average considered a highly important issue in the public debate by our respondents (mean on the 7-point Likert scale: 5.411, where 1 is "Not important at all" and 7 is "very important"; sd: 1.393795). Authors found additional evidence that the sensitivity towards privacy is stronger among non-members than members. Members claim that the FB is very useful to them for learning about and finding classmates (4.93 mean on a 7-point Likert scale) and for making it more convenient for people to get in touch with them (4.92), but deny any usefulness for other activities. detected little or no relation between participants' reported privacy attitudes and their likelihood of providing certain information, even when controlling, separately, for male and female members. (Survey n=294, Data mining n=7000: Likert scale survey analysis and data mining; 7000 profiles mined before & after survey; survey respondents were volunteers found through online university sources.) (Source:http://64.233.179.104/scholar?hl=en&lr=&q=cache:qfny3jN33rsJ:petworkshop.org/2006/preproc/preproc_03.pdf+; retrieved January 25, 2008)	Membership Privacy Social interactions

Reference	Synopsis/Findings	Themes
Andon, S. P. (2006). *Evaluating Computer-Mediated Communication on the University Campus: The Impact of Facebook.com on the Development of Romantic Relationships.* Florida State University, Tallahassee Florida.	Overall, the author found that heavier users of Facebook report higher levels of interpersonal attraction, attributional confidence, and perceived similarity with relational target that suggests some of the traditional purposes of face-to-face communication are being achieved via Facebook. Apparently, Facebook is being used in conjunction with interpersonal relationships offline. Nearly 96% (95.9) of participants were members of Facebook. At a significance level of p<.001, there was a positive correlation between these Facebook usage and confidence measures. The reported r-value for the analysis was r=.297, showing a slightly positive correlation between Facebook use and physical and social attraction of the relational target at a significance level of p<.001. At a significance level of p<.001, the r-value for the correlation was r = .268, showing a positive correlation between amount of Facebook use and perceived similarity of relational target. Factors such as age, sex, relationship status, Internet produced no significant results when considering possible associations to the scales used. However, both race and Greek membership were found to indicate some relationships. (n=252 "valid" respondents: 4-page survey given to non-random sample of undergraduates from FSU required courses; were given extra credit for participating.) (Source:http://64.233.179.104/scholar?hl=en&lr=&q=cache:qWpYMGxeKtMJ:etd.lib.fsu.edu/theses/available/etd-01052007-170005/+; retrieved January 24, 2008)	Users Traditional communication Offline to online communication Relationships

Reference	Synopsis/Findings	Themes
Boogart, M. R. V. (2006). *Uncovering the social impacts of Facebook on a college campus.* Kansas State University, Manhattan, Kansas.	Students with a lower GPA tend to spend more time using Facebook. Though this is not a causal relationship, there is a significant relationship between the two variables (p = .000). Other than this finding, there were no other negative effects. Being connected to the college campus was the positive implication around which all other findings revolved. Facebook was used by students of color and students who identify as non-heterosexual as a tool to make social connections they could not make in person. Students who use Facebook more than one hour daily report feeling 9.3% (p = .002) more connected to campus that those who use it less than 30 minutes a day. *Researcher Comment: While "just" a Master's thesis, this research was well designed and reported. This thesis is actually cited repeatedly in the other literature appearing after it's publication date.* (n=2776: Online survey; participating institutions randomly selected from listings of Carnegie ratings &16,667 potential subjects were sent an email invitation to participate – 18% response rate.) (Source: http://krex.k-state.edu/dspace/handle/2097/181; retrieved January 8, 2008)	Users Social connections
Decker (2006). Online self-reported information: Facebook a hiring tool for businesses.	While not focusing on the use of Facebook in academics, this thesis has some excellent models of perception management that could be easily adapted to the educational realm. The author found that from the companies sampled, 30% of them use Facebook to lookup potential employee profiles. He states that the results also suggest that the use of Facebook as a background-checking tool will most likely increase in the future. Of the respondents who did not know about Facebook, 46.4% said they would consider using it now that they know it is available. The author also noted that roughly 10% of companies surveyed thought the use of Facebook to lookup potential employees to be unethical. (n=40 total; Used Facebook n=12 Not used Facebook n=28: Likert scale self-report questionnaire given to 100 companies at a university job fair; 40% response rate.) (Source: http://proquest.umi.com/pqdweb?did=1232407651&sid=2&Fmt=2&clientId=1652&RQT=309&VName=PQD&cfc=1)	Workplace use Background checking

Reference	Synopsis/Findings	Themes
DiMicco, J. M., & Millen, D. R. (2007). *Identity management: multiple presentations of self in Facebook*. Paper presented at the Proceedings of the 2007 international ACM conference on Supporting group work.	From the authors' preliminary exploration, they identified signs pointing towards the use of Facebook in the workplace and it becoming part of the workday routine of young hires. The authors identify two main benefits of Facebook in the workplace, the first being that it offers workers improved ways of keeping in touch with colleagues. The second benefit is that it allows users to learn about new employees and team members that may lead to finding coworkers with specific skills that are needed for a project. However, there are significant difficulties involved with using one site simultaneously for both professional and non-professional purposes and efforts should be directed to alleviate some of these concerns so there is no unintended leakage between the corporate and social personas the worked maintains in Facebook. (n=68: Initial interviews (no *n* reported); subsequent "examination" of 68 user profiles using Facebook's feature of randomly viewing profiles within a specific network.) (Source: http://portal.acm.org/citation.cfm?id=1316624.1316682#; retrieved January 25, 2008)	Workplace use Professional use Non-professional use Social connections Networking
Dwyer, C., Hiltz, S. R., & Passerini, K. (2007). *Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace*. Paper presented at the Americas Conference on Information Systems.	The authors found that users of Facebook and users of MySpace both expressed similar levels of concern regarding internet privacy. However, Facebook members were more willing to include identifying information in their profiled and were more trusting of the site and its members overall. Despite this, MySpace members were more active in developing new relationships through the social networking software. The author conclude that not enough is know about the interaction of trust and privacy to develop a behavioral model in online social activities. They call for further research to help researchers understand the development of relationships and the reason for the behavioral differences on the different sites. (n=117: Includes 69 Facebook members (29 male and 40 female) and 48 MySpace members (29 males and 19 females), Self-report Likert scale survey.)	Privacy Trust Relationships Social connections

Reference	Synopsis/Findings	Themes
Ellison, N. B., Steinfield, C., & Lampe, C. (2006). *Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook*. Paper presented at the Annual Conference of the International Communication Association.	The authors found that 94% of the students surveyed were members of Facebook and that neither gender, ethnicity, nor income related to Facebook use. However, older students and those that had been at school longer were less likely to use Facebook and on-campus students were more likely than off campus to use Facebook (the authors point out this may be due to the fact that freshmen are required to live on campus). They found that all of the sorority and fraternity members in their sample were Facebook members. Students reported spending 10-30 minutes on average each day using Facebook and had between 150 and 200 friends on the system. They were significantly more likely to use Facebook for fun than for gathering information & spend more time responding to someone with whom they have an offline connection. In regards to social capital and the forms reported by the students, the authors say "Although we cannot say which precedes the other, Facebook does appear to play an important role in the process by which students form and maintain social capital, with usage associated with all three kinds of social capital included in our instrument." The authors believe that Facebook constitutes a newer form of socializing in a virtual way in which connections are initially made off-line and then migrated Facebook. (n=286: 800 survey invitations sent out; participants got $5 for completing survey; 35.8% response rate.)	User characteristics User behavior Online vs. offline connections Social connections

Reference	Synopsis/Findings	Themes
Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook "Friends:" Social Capital and College Students Use of Online Social Network Sites. [Research]. *Journal of Computer-Mediated Communication, 12*(4), 1143-1168.	The authors found a positive relationship between certain kinds of Facebook use and the maintenance and creation of social capital. While not being able to say which precedes the other, Facebook appears to play an important role in the process of forming and maintaining social capital. When authors compared members vs. nonmembers, they saw no real difference in demographics, with the exception of class year and age (which is strongly correlated with class year). This suggests that "Facebook has broad appeal, does not exclude particular social groups, and has not had a noticeable effect on participants' grades." The authors found that most students used Facebook to keep in touch with old friends as well as to maintain or intensify relationships in offline settings, such as a shared class, dormitory, etc. The use of Facebook "intensity predicted increased levels of maintained social capital, which assessed the extent to which participants could rely on high school acquaintances to do small favors." (n=30,773: From target network, data was collected using automated scripts the downloaded information from profiles and saved it in an offline database. Only profiles that were not designated "friends only" could be downloaded.) (Source: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1083-6101.2007.00367.x; retrieved January 21, 2008)	Social connections User characteristics Relationships
Faith, K. L. (2007). *Face-Time: The Construction of Identity on Facebook.* Miami University, Oxford, Ohio USA.	The author found that the majority of Facebook users see their ' "real life" and "virtual life" as being integrally linked." Users also seem to be aware of the performance factors involved with their Facebook identity. The author found that Facebook users frequently think about their uncertain audience and how their audience might interpret their actions. (n=8: Interviews; examination of Facebook profiles.) (Source: http://www.ohiolink.edu/etd/view.cgi?acc_num=muhonors1178214020)	Offline and online relationships Reputation

Reference	Synopsis/Findings	Themes
Golder, S., Wilkinson, D. M., & Huberman, B. A. (2006). Rhythms of social interaction: messaging within a massive online network. *Computers and Society*, 16.	The authors looked at patterns and trends among 362 million anonymized messages and "pokes" sent by 4.2 million Facebook users. They found that the college student "weekend" is clearly visible in the data with the students using a great deal less than during other times of the week. Presumably, this is because most of them spend the weekend socializing in other ways. Seasonal variation in same- school/different-school messaging confirmed the authors' belief that messaging is used in support of geographically distant relationships. They found that messages within a student's school are higher precisely at the times students are not at school suggesting students are using Facebook to communicate with others who may not be geographically close. The authors also found that friends sent most messages although most friend pairs do not exchange messages. The authors propose that messaging is reliable measure of Facebook activity. (n=362,000,000 messages: Analyzed the fully-anonymized headers of 362 million messages exchanged by 4.2 million users of Facebook, an online social network of college students, during a 26-month interval.) (Source: http://arxiv.org/abs/cs/0611137v1)	User behavior Online and offline relationships Geographic communication Messaging
Gosling, S. D., Gaddis, S., & Vazire, S. (2007). *Personality Impression Based on Facebook Profiles*. Paper presented at the International Conference on Weblogs and Social Media.	The authors found that (a) there is some consensus for all Big Five dimensions and Facebook-based personality impressions, especially for extroversion; (b) except for Emotional Stability, the impressions are fairly accurate; (c) observers are only aware of how the seen in terms of Extraversion; (d) and in regards to the domains of Emotional Stability and Openness to Experience, profile authors engaged in some self-enhancement. The authors content "the data presented here suggest that the online social networking websites are, in fact, a relevant and valid means of communicating personality." They urge future research to focus on the social cuing that occurs in online social networks and the real-world effects of impressions gleaned from online social networks. (n=133: Participants rated themselves and 4 friends; 8 months later, they were completed similar measures regarding their perceptions of themselves and their friends.) (Source: http://www.icwsm.org/papers/paper30.html; retrieved January 25, 2008)	Social interaction

Reference	Synopsis/Findings	Themes
Gross, R. & Acquisti, A. (2005). *Information revelation and privacy in online social networks*. Paper presented at the Proceedings of the 2005 ACM workshop on Privacy in the electronic society.	The authors found that users in the target network provided an "astonishing amount" of information: 90.8% of profiles contain an image, 87.8% of users reveal their birth date, 39.9% list a phone number (including 28.8% of profiles that contain a cell phone number), and 50.8% list their current residence. Users also disclose their dating preferences, relationship status, political views, and other interests. Most types of information are shared equally by either male or female, except for phone numbers. 47.1% male users vs. 28.9% female users provide a phone number. Only about 8% of the user names were obviously fake & only 3% of users chose only to disclose their first name. 98.5% of the profiles that provide a birth date provide fully identifiable information, including day, month & year. 61% of all profiles had images that could lead to direct identification. Only 3 users changed limited their profile visibility (0.06%) to only those at the home institution. Authors provide data related to re-identifying a person based on the information provided in their Facebook profiles. (n=4540: Search for male and female profiles in target university's Facebook network & extracted profile data.) (Source: http://portal.acm.org/citation.cfm?doid=1102199.1102214)	Privacy Information sharing Fake vs. real identity
Grude, A., Scholl, M., & Thompson, R. (2006). Privacy on Facebook. Unpublished Research Report.	The authors found from their survey of 60 profiles that the type of information in questions affects the users' willingness to share that information. They found that the categories most often shared were email addresses, birthdays, and undergraduate schools. Users indicated there were three main reasons they use Facebook, (1) ease of use; (2) can maintain old connections and make new ones; (3) the tiered privacy satisfies the privacy requirement in the mind of many users. Users did not express concern over privacy issues or loss of personal information with Facebook, even when a breech was reported. (Interviews n=9, Profile survey n=60: Interviews, profile surveys.) (Source: http://www.amygrude.com/documents/689.pdf; retrieved January 25, 2008)	Information shared User interests Privacy

Reference	Synopsis/Findings	Themes
Hewitt, A., & Forte, A. (2006). *Crossing boundaries: Identity management and student/faculty relationships on the Facebook. Paper presented at CSCW, Banff, Alberta, Canada.* Paper presented at the Computer Supported Cooperative Work.	The authors found there was no significant difference on students' ratings of professors on Facebook as opposed to those not on Facebook. Some students said that the professor's interaction in Facebook was a positive thing and none reported it as being negative. However, 33% of the respondents raised issues regarding privacy and their professor's presence. Men were more comfortable with professor being on Facebook than women. Those who responded positively to the professor's presence indicated that having an alternative means of communication allowed them to get to know their professors better. (n=106: Surveys given to volunteers in 2 undergrad classrooms where professors had Facebook profiles for at least 6 months.)	Educational use Privacy Social communication
Jones, H., & Soltern, J. H. (2005). Facebook: Threats to Privacy. Unpublished Research Report. Project MAC: MIT Project on Mathematics and Computing.	The authors posit that privacy on Facebook is undermined by three principal factors: that users typically disclose too much, that Facebook doesn't take adequate steps to protect the privacy of the user and that other people can actively seek out user information using Facebook & exploiting security holes. *Researcher Comment: Fascinating reading – I highly recommend this article for a great overview of Facebook usage.* (Survey n=419, Spider "crawl" n=70311 (72.3% of all profiles at target institutions): Survey of MIT students on the use of Facebook's features; created a spider that "crawls" and indexes Facebook, attempting to download every single profile at a target school.)	Privacy Information sharing
Lampe, C., Ellison, N. B., & Steinfield, C. (2006). *A face(book) in the crowd: social Searching vs. social browsing.* Paper presented at the Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work.	Students use Facebook to find information about people they have an offline connection with. They do not expect their profiles to be accessed by professors or faculty, but mainly by peers. The participants in this study are not using Facebook to make new connections, but to strengthen existing ones. This is contrary to the popular view that online social networks are typically used to make new connections with people not known offline. Participants also felt that their profiles were an accurate and positive representation of themselves. (Study 1 n= 1440, Study 2 n= 1085: Two voluntary surveys of first-year students at Michigan State University (MSU); web-based delivery; Response rate: Study 1 = 20%; Study 2 = 15%.) (Source: http://portal.acm.org/citation.cfm?doid=1180875.1180901; retrieved January 17, 2008)	Offline/online communication User characteristics Relationships Social connections

Reference	Synopsis/Findings	Themes
Lampe, C., Ellison, N. B., & Steinfield, C. (2007). *A familiar face(book): profile elements as signals in an online social network*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.	The authors found that there is a positive relationship between how many fields are populated and the number of friends a user will have listed in Facebook. However, the amount of information in the open-ended fields does not have a relationship to the number of friend. Even when the authors controlled for gender, time on the system, user status in the community, and the regency of updating, there is a significant association with the types of profile information and the number of friends. (n=30,773: From target network, data was collected using automated scripts the downloaded information from profiles and saved it in an offline database. Only profiles that were not designated "friends only" could be downloaded.) (Source: http://doi.acm.org/10.1145/1240624.1240695; retrieved January 5, 2008)	User characteristics Social connections
Lorenzetti, J. P. (2007). Do You Really Understand Your Students? Answers About Student Preferences in Technology. [Feature Article]. *Distance Education Report, 11*(7), 5-6, 8.	The author found that 98 percent of participants owned a computer, 96 percent owned cell phones, 63 percent owned a digital camera, 53 percent owned a flash drive, over 99 percent of all students used their computers at least several times a week, 68 percent of students checked e-mail daily and 22 percent checked at least several times a week, and 88 percent of student had an account with either Facebook or MySpace. The author then offers suggestions for administration and faculty. (Case study.) (Source:http://vnweb.hwwilsonweb.com/hww/results/results_common.jhtml;hwwilsonid=G32BDV0TAEVHFQA3DIMSFGGADUNGIIV0)	User characteristics Professional use

Reference	Synopsis/Findings	Themes
Mack, D., Behler, A., Roberts, B., & Rimland, E. (2007). Reaching Students with Facebook: Data and Best Practices. Electronic Journal of Academic and Special Librarianship, 8(2).	After tracking all reference questions by method of contact over the course of a semester, the librarian/research found that all Facebook questions were from undergraduates & all of the phone inquiries were received from faculty and community patrons. Email was used by students, faculty, & community patrons. During Fall semester 2006, a librarian at Penn State documented the number of reference and research questions he received over the course of the term. These statistics include transactions by email, telephone, instant messenger, Facebook, and in person during office hours.) (n=441 different questions.) (Source: http://southernlibrarianship.icaap.org/content/v08n02/mack_d01.html)	Library User characteristics Professional use Non-professional use
Mazer, J. P., Murphy, R. E., & Simonds, C. J. (2007). I'll See You on "Facebook": The Effects of Computer-Mediated Teacher Self-Disclosure on Student Motivation, Affective Learning, and Classroom Climate. [Research; Peer reviewed]. *Communication Education, 56*(1), 17.	Excellent article! The authors found that high teacher self-disclosures as operationalized in their study may lead to a more comfortable classroom climate and higher levels of student motivations and affective learning. The small effect sizes the authors found, however, prevent them from making broad generalizations with regards to faculty use of Facebook and potential impact on students learner and classroom culture. (n=133: Experimental study; random assignment to one of three treatment conditions; questionnaire administered at end of treatment. Participants were volunteers from a basic communications course & given extra credit for participating.) (Source: http://www.informaworld.com/smpp/content?content=10.1080/0363452060100 9710; retrieved January 14, 2008)	Professional use Educational use Trust User incentive Classroom culture

Reference	Synopsis/Findings	Themes
Medina, R. & Gabriella, M. (?). A Look at the Factors influencing the Utilization and Enjoyment of Computer-based Social Network.	This study examined the association of university students' usage of Facebook and the factors influencing the enjoyment of computer-based social networks. The factors that were investigated in this study were categorized as factors related to an escape or entertainment dimension and also to the enhancement and maintenance of interpersonal relationships through Facebook. Surveys were distributed to a random sample of students at a mid-sized southeastern university. Collected data from these surveys were used to first examine relationships. Significant correlations were found between the escape dimension and the students' usage of Facebook. Overall, the results suggest that Facebook usage by the students was more associated with a need for entertainment and escape from reality.	Non-professional use Enjoyment Entertainment Relationships
Miller, E. M. (2007). *A phenomenological exploration of residential students' use of online social networking sites.* California State University, Long Beach, California.	The author reports that the student participants viewed participation in online social networks as a natural occurrence. Most reported that the students feel technology and their ability to use it is part of their overall identity. The author found that students used Facebook to maintain connections to those that they don't have a daily connection with and others they don't see often. They also used the online social networking sites to gain knowledge about people they met or were going to meet. Students described using the OSN as a "home-base" of sorts that wasn't dependent on location and that it allowed them a way to stay connected with their community of peers regardless of geographic boundaries. (n=14: Phenomenological study of 2nd semester freshmen living on campus.) (Source:http://proquest.umi.com/pqdweb?did=1407497981&sid=1&Fmt=2&clientId=1652&RQT=309&VName=PQD)	Non-professional use Networking Offline to online relationships

Reference	Synopsis/Findings	Themes
Nyland, R. & Near, C. (2007). *Jesus is My Friend: Religiosity as a Mediating Factor in Internet Social Networking Use*. Paper presented at AEJMC Midwinter Conference, 2007.	In this interesting & well thought out article, the authors conclude "while religiosity may not be a good predictor of overall use of social networking tools, it is a significant mediating factor in looking at the individual purposes of use, and those uses may be used to enhance religious experiences and community." They authors found that there is no relationship between a respondents religiosity and their activity on social networking sites. Though the more religious a respondent was the more likely it was they used the social networking sites to maintain relationships, religious people also used the networks to find new relationships as well. For religious women, authors found a relationship between religiosity and the use of social networking to keep updated on social events. (n=184: Internet survey (less than a 5% response rate of the 5000 invitations sent out).)	Religious experience and community Relationships Social connections
Shea, K., & Wesley, J. (2006). How Social Networking Sites Affect Students, Career Services, and Employers. [Feature Article]. *NACE Journal, 66*(4), 26-32.	ABSTRACT: A study examined the role of social networking sites, such as Facebook and MySpace, in the job search of college students and the candidate screening process of employers. Participants were 124 college students and 64 companies. Findings revealed that the majority of student participants felt they would be comfortable with what an employer might find out about them through a Google search. Among student participants who thought employers might use search engines to screen potential candidates, few thought that employers would use social networking sites for the same purpose. However, some employer participants did use social networking sites to source and research candidates. Other findings and implications of the study are discussed. (College students n=124, Potential employers n=64)	Job screening Professional use

Reference	Synopsis/Findings	Themes
Strater, K., & Richter, H. (2007). *Examining privacy and disclosure in a social networking community*. Paper presented at the Proceedings of the 3rd symposium on Usable privacy and security.	The authors found that participants were careful to maintain appropriate disclosure in their personal profile information. However, participants were significantly less conscious about maintaining their friends list, posted pictures, and Wall posts. The authors report that participants underestimated the extent and activity of their social network and therefore remained at risk for over-disclosure and privacy invasions. (n=12: Participants completed demographic surveys and a personality inventory (NEO-FFI). Participants then reviewed and were interviewed regarding their own profile. Participants then viewed profiles of two fellow participants, and two introduced by the researcher. Participants evaluated these other profiles using selected questions from the NEO-FFI, and answered additional interview questions.) (Source: http://doi.acm.org/10.1145/1280680.1280706)	Privacy Information sharing
Tufekci, Z. (2008). Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites. *Bulletin of Science, Technology & Society, 28*(1), 16.	The authors found that 85.4% of respondents had a profile on a social networking site. 94.9% of respondents with Facebook profiles used their real names in their profile while 62.8% of MySpace users used their real name. Despite this, Facebook users were more likely to restrict their profiles to friends only. Also, respondents were just as likely as not to have a public profile when using their real name. To manage unwanted audiences, respondents adjusted profile visibility and used nicknames but didn't restrict the information within the profile itself. On Facebook men were more likely than women to use their real name on a public profile (p=.06). Racial differences in profile visibility were also significant. (n=704: Convenience sample from 8 different courses over 3 semesters; Likert-scale survey.)	Information sharing User characteristics

Reference	Synopsis/Findings	Themes
Vie, S. (2007). *Engaging others in online social networking sites: Rhetorical practices in MySpace and Facebook.* Unpublished Dissertation, The University of Arizona.	The author found that student users of social networking sites tend to encourage others to join for an number of reasons, primarily to increase their own friends number and to build up the friends network—they also encourage others to join to show allegiance to a particular group. The author also found that joining MySpace or Facebook also signals participation in a particular social group with Facebook being perceived as being used by more college students and MySpace was primarily the domain of high school students. Students perceived that privacy and security was higher in Facebook than MySpace and find it more "trustworthy." In regards to instructors, the author found that even when they are familiar with the technology, they aren't as likely to create or maintain a profile on either MySpace or Facebook. Many felt that it was a "student space" and expressed a desire to respect the students' personal identities. Instructors also reported feeling that the sites were a "waste of time." Overall, those instructors who did participate in the social networking sites were very suspicious of privacy and security issues. *Researcher Comment: Excellent study – has a great deal of important information in it but I wonder if it won't get much attention because it quite lengthy.* (n= 354 for survey, n=20 for interviews: Qualitative—used combination of surveys, interviews, site analyses, member checking, and self-reflection for triangulation; interviews were of both instructors and students using MySpace and/or Facebook.) (Source:http://proquest.umi.com/pqdweb?did=1288668741&sid=1&Fmt=2&clientId=1652&RQT=309&VName=PQD)	Networking User characteristics Privacy Trust Professional use Non-professional use Educational use

Reference	Synopsis/Findings	Themes
Walther, J. B., Van Der Heide, B., Kim, S.-Y., Westerman, D., & Tong, S. T. (2008). The Role of Friends, Appearance and Behavior on Evaluations of Individuals on Facebook: Are We Known by the Company We Keep? *Human Communication Research, 34*(1), 28-49.	In this experiment, the physical attractiveness of one's friends' photos, as seen in the Facebook wall postings presented on another individual's profile, had a significant effect on the physical attractiveness of the profile's owner. Perceptions of physical attractiveness did not reduce task competence attributions, an effect associated with evaluations of women in other, offline domains. Results showed that complimentary, pro-social statements by friends about profile owners improved the profile owner's social and task attractiveness, as well as the target's credibility. An unanticipated interaction effect involving the sex of the profile owner and the nature of the wall statements was obtained with respect to the effect of friends' comments on perceptions of the targets' physical attractiveness. The negative statements depicted normatively undesirable behavior, as they involved sexual innuendo and insinuated that the target person was drinking excessively the previous night. These statements raised the desirability of a man's appearance among the subject population in this study, whereas the residues of such behavior rendered the target physically unattractive when she is female. These results reflect what has come to be known as the sexual double standard when making social judgments or forming impressions of others. (n=389: Participants volunteered to take part in the research in exchange for extra credit or satisfaction of a research requirement in one of several communication or telecommunication courses. Each participant viewed one of the eight stimuli each containing a mock-up of a Facebook profile. Differences among stimuli reflected variations in (a) physically attractive or unattractive photos of ostensible wall posters and (b) positively or negatively valued content of the wall posting messages with respect to their description of the profile owner's behavior.) (Source: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1468-2958.2007.00312.x)	User characteristics Social connections User behavior

Reference	Synopsis/Findings	Themes
Watson, S. W., Smith, Z., & Driver, J. (2006). *Alcohol, Sex and Illegal Activities: An Analysis of Selected Facebook Central Photos in Fifty States*.	After examining 150 photos, authors found that "alcohol was included in approximately 9% of the central photos, some photos were sexually suggestive (2.7%), and one photo contained partial nudity (0.7%)." Authors concluded that their findings indicate that the incidence of negative photos in Facebook central photos is far less frequent than the media typically reports. (n=150: Used a checklist/inventory developed by researchers that included settings, actions, events & people as well as space for writing additional descriptors when needed.)	User characteristics

Search Methodology

Database	Search Terms	Articles Found	Empirical Studies
CiteULike	Facebook	65	8
	Facebook, research	1	1
	Social network, Facebook	11	8
	Online social network, Facebook	11	8
Digital Dissertations	Facebook	5	5
	Online social network, Facebook	1	1
Education Full Text	Facebook	34	3
	Online social network, Facebook	2	0
	Social network system, Facebook	2	0
	Social network, Facebook	19	0
ERIC via Ebsco Host	Facebook	7	0
	Social network, Facebook	5	0
	Online social network, Facebook	4	0
ERIC via US Dept. of Education	Facebook	11	2
	Online social network, Facebook	5	1
	Social network, Facebook	5	1
Google Scholar Advanced Search	"Facebook research"	7380	17
	"Online social network", Facebook	91	14
	"Social network", Facebook	610	14
ISI Web of Knowledge	Facebook	9	3
	Social network, Facebook	0	0
	Online social network	0	0
Library Literature	Facebook	20	1
	Online social network, Facebook	11	0
	Social network, Facebook	10	0
PsyINFO via EBSCO Host	Facebook	6	5

Appendix C: Wikipedia (www.wikipedia.org)

Researcher: Heather Leary heatherleary@gmail.com

Literature Review

Reference	Synopsis/Findings	Themes
Adler, B.T. & Alfaro, L. (2007). *A content-driven reputation system for Wikipedia*. In the proceedings for the International World Wide Web Conference, Banff, Alberta, Canada, May 8-12.	This study is looking at content-driven author reputation systems. Users are judged on how their contributions fair (stand the test of time and other contributors). Authors gain better reputations when their edits are preserved, and they lose reputation when edits are rolled back to previous edits. This way of tracking reputation could be used to flag new contributions from low-reputation authors or to allow only high-reputation authors to contribute to controversial or critical pages. This study implemented this system on the entire Italian and French Wikipedias. Results show that reputation has good predictive value; they found that low-reputation authors had a larger probability of having poor quality edits. Author reputation is important to Wikipedia because it guides the tracking of edited entries by novice or seasoned contributors. Limitations to using a reputation system like this is low-reputation contributors could just be new to the system and have not had a chance to build a reputation. Thus their edits are scrutinized more. The results show that a content-driven reputation is an effective tool for spotting bad contribution the moment they are introduced. (Tags: wikipedia; author reputation; reputation system) (Source: http://www2007.org/proceedings.html)	Reputation Reputation tracking User contributions Edit tracking

Reference	Synopsis/Findings	Themes
Anthony, D., Smith, S.W., Williamson, T. (2007). *The quality of open source production: Zealots and good Samaritans in the case of Wikipedia.* Retrieved February 19, 2008, from http://www.cs.dartmouth.edu/reports/abstracts/TR2007-606/.	There is a benefit to contributors of Wikipedia, a reputation within the community. Apparently this reputation is the reason they give to why the content on Wikipedia is of high quality and can match that of Britannica. The other motivation behind Wikipedians contributing is a sense of commitment to the community. Registered users are interested in establishing a reputation within the community since it is easier to track them that way, while those who submit anonymously usually have fewer submissions. This study looks at what the implications are for quality among contributors. First they identify two groups of contributors: 1) registered users with large numbers of contributions (zealots) 2) unregistered users who contribute once in a while. Those within group 2 can be three groups of people, those who are experts and contribute once in a while with high quality content (the good Samaritans), those who sees mistakes and holes and change them (more good Samaritans), and those who are not experts who contribute low quality content. The experts do not care about their reputation or the community, they care about the content for the topic they edit. Using the French and Dutch Wikipedias, samples were drawn form user lists (registered and anonymous) with n=7,058. Hypothesis: contributor motivations effect the quality of their contributions. They look at contributions by these people based on edit (content from contributor), previous (version of article previous to edit), and current (version of article when data was collected). They then measure the retention of the edit as contained in the current version. Results: overall registered users contribute more content than anonymous users. But anonymous users contribute higher quality content than registered users. Anonymous users quality is highest at lower levels of commitment and decreases as participation increases. Registered users have higher quality with more participation. The shorter a contribution for both types of users, the higher the quality. The finding that anonymous good Samaritans as well as experts who contribute high quality content suggests that open source production enables the exploitation of untapped productive resources. From the findings they suggest that for the majority of users it is the quantity as well as the quality of contributors that positively affects the quality of open source production.

(Tags: wikipedia, reputation, incentives, quality)

(Source: http://www.cs.dartmouth.edu/reports/abstracts/TR2007-606/) | Reputation
User contributions
Content quality
Community
Edit tracking |

Reference	Synopsis/Findings	Themes
Arazy, O., Morgan, W., Patterson, R. (2006). *Widsom of the crowds: Decentralized knowledge construction in Wikipedia*, paper presented at the 16th Annual Workshop on Information Technologies & Systems, Milwaukee, Wisconsin, December 9-10.	Building off of Surowiecki's (2005) *Wisdom of Crowds* (WoC), the authors develop a model of factors that determine wiki content quality. In this study they operationalize this framework and study Wikipedia. The main goal of the study is to determine if the WoC principles can indicate the quality of a Wikipedia entry. The main question to answer is to what extent the size and diversity of the group editing in Wikipedia impacts the quality of the entry? Methods: Using 42 Wikipedia articles (the same 42 that were used in the *Nature* study), Their quality measure is 1/number of errors. Using partial least squares (because it measure causal relations), the authors found that the number of edits is highly correlated with the number of authors; discussion page length and the number of edit wars is highly correlated, both significant at 0.01 level. A causal relationship between diversity and quality is established, and size has a significant effect on diversity. Here's a model to look at: size has a positive effect on diversity, diversity has a positive effect on quality. They report that 30% of the variance in article quality is explained by this model. These findings suggest that the WoC principles do determine the quality on Wikipedia and that the more people who edit entries the higher the quality of the entry. (Tags: wikipedia, content quality, wisdom of crowds) (Source: http://ssrn.com/abstract=1025624)	Content quality Users Diversity

Reference	Synopsis/Findings	Themes
Bellomi, F., & Bonato, R. (2005). *Network analysis for Wikipedia*. In proceedings of The International Wikimania Conference, Frankfurt, Germany, August 4-8.	The point of this article is to understand some of the high level structure in Wikipedia and to understand a little more its content, particularly hidden cultural biases. Using two web metrics, Hyperlinke-Induced Topic Selection (HITS) algorithm for page ranking web pages with common topics and PageRank algorithm for relevance, the entire English Wikipedia was were drawn from. The authors claim that network analysis gives a simple way of semantic measures. Results show that Wikipedia's internal references form a connected graph. You can get anywhere in the encyclopedia from anywhere. To understand the general structure, they use relevancy metrics to show which classes of concepts are relevant in global ranking. By isolating some of these classes extract a topic-specific ranking to work on identifying cultural biases. The HITS authority shows that geographical space and time (spans and events) are the main organizing categories for Wikipedia. The PageRank metric shows an overwhelming dominance of concepts tied to religion. This shows that Wikipedia has biases. Namely that it is based on Western culture and history. They are interested in looking at this same concept for all of the Wikipeida resources. (Tags: wikipedia, metrics, page ranking, cultural bias) (Source: http://www.fran.it/blog/2005/08/network-analisis-for-wikipedia.html)	Structure Content Cultural biases Page ranking

Reference	Synopsis/Findings	Themes
Bragues, G. (2007). *Wiki-Philosophizing in a marketplace of ideas: Evaluating Wikipedia's entries on seven great minds*. Accessed February 22, 2008 from http://ssrn.com/abstract=978177.	This study compares the information found about topics from seven philosophers in four academically respected works with the same information found in Wikipedia. The study's objectives are to have a better understanding of collaborative process (Wikipedia) stands up against traditional select individual expertise. The author uses Wikipedia to test if the ideas of the encyclopedia can be trusted as a social mechanism for producing valid information. Methodology: content analysis of topics of the entries in Wikipedia and four respected works; to analyze the quality of the respected works the author looked at where the summaries of these books agreed on information; a consensus was defined when themes were found in 3 of the 4 works. In Wikipedia quality is determined by it's ability to record facts and avoid expressions of opinions. After determining the quality and themes in the four respected works, Wikipedia entries were analyzed from downloads made in January 2007. Results: Wikipedia captured 51% of the themes and information found within the expert consensus. The study was unable to uncover any outright errors, what it found was more information was missing. Wikipedia has the potential to generate knowledge and cannot be dismissed. But, it still has a ways to go before proving it is better than knowledge that comes from the select few experts. (Tags: wikipedia, content quality) (Source: http://ssrn.com/abstract=978177)	Comparison Collaboration Expertise vs. novice Content quality Social interactions

Reference	Synopsis/Findings	Themes
Braun, S. & Schmidt, A. (2007). *Wikis as technology fostering knowledge maturing: What we can learn from Wikipedia*. In proceedings from the International Conference on Knowledge Management, Graz, Austria, September 3-5.	This study takes a qualitative and quantitative approach to studying Wikipedia entries for knowledge maturation. Maturing is defined as artefact layer, knowledge layer, and social layer. In looking at what Wikipedia has in place for these three topics, it is obvious that it is based on artifacts. This is shown by the number of wiki pages present in the wiki. This is also seen in Wikipedia's assigning 'tags' to certain articles (featured article, good article, need review). It does have in place a few social layers, for example discussion pages and watchlists. These allow the community to communicate and establish themselves within it. In the quantitative part of the study the authors analyze the maturity of the artefact layer. Being able to look at the history of an entry is very helpful here. Maturity is measure by readability, presentation and format, and linkage. From the German Wikipedia dump XML dump on January 24, 2007 articles and words were examined. The mean and median results of the analysis correlate the maturing process. The number of headlines and images per article increases from stub to featured articles, as well as linkage. The study concludes that Wikipedia does in fact show maturation and that social processes have brought forth the ability to foster the maturity. (Tags: wikipedia, knowledge maturity)	User tags Content quality Readability Presentation Linkage Social interactions Community communication

Reference	Synopsis/Findings	Themes
Bryant, S.L., Forte, A., Bruckman, A. (2005). *Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia.* In proceedings of the International ACM SIGGROUP Conference on supporting group work, Sanibel Island, Florida, November 6-9.	This qualitative study uses peripheral participation to understand participation as an adaptable process. Using ideas from activity theory, the authors describe how activity in Wikipedia stands in contrast to traditional publishing and suggests a new paradigm for collaborative systems. Legitimate peripheral participation is a description of how newbies become members in communities of practice. This study examines how Wikipedia users motivations and perceptions of their roles change as they become less peripheral and more engaged in the community. Activity theory describes activity in socio-technical systems with six elements: object, subject, community, division of labor, tools, and rules. Methods: A purposeful sample was drawn from the Wikipedia community through communication channels used by active members. Interviews were conducted with n=9. The interviews were conducted to gather qualitative data about why these participants contribute to Wikipedia, how they got started, how they perceive their role, and how their perception of Wikipedia and their role has changed over time. Participants averaged activity was 14 months with a duration of 2 months to 2.5 years. All reported daily or almost daily activity on the site. Findings: As participants move from peripheral to engaged activity on Wikipedia, their activity changes as well. They report moving from a consumer to a creator, their goals for participating in the community change; novices edit what they know, expert Wikipedians broaden their goals of just editing to becoming more concerned with the overall improvement of the encyclopedia and the community. Transformation of the tool use is another step; novices come to Wikipedia usually through search engines and see the ease of editing, as expert Wikipedians they come to understand there are more tools to use that support them in their efforts on the site. Perceptions of community, rules, and division of labor evolve as well; novices do not talk or participate in the community, expert Wikipedians view themselves as part of the community where they give and get from it. (Tags: participation, wikipedia, collaborative systems) (Source:http://portal.acm.org/toc.cfm?id=1099203&coll=Portal&dl=ACM&type=proceeding&idx=SERIES376&part=series&WantType=Proceedings&title=GROUP&CFID=16553271&CFTOKEN=86760279)	User contributions Traditional publishing Collaboration Participation Community Users

Reference	Synopsis/Findings	Themes
Bunescu, R.C. (2007). *Learning for information extraction: From named entity recognition and disambiguation to relation extraction.* Unpublished Computer Science dissertation, University of Texas at Austin, Austin, Texas, United States.	This work focuses on creating extracting information from textual entities to represent that information in a structured format. The main goal is to design extraction models that obtain improved performance by using evidence that hasn't been explored before. Wikipedia (as it hasn't been used before for this purpose) was chosen for the study. The author is interested in looking at methods that automatically induce an extraction model by training on a dataset of manually labeled examples. Using named entity recognition (extraction task concerned with finding textual mentions of entities in a category), they use Wikipedia as a repository of named entities to study their extraction and compare the Wikipedia taxonomy with their ranking approach to disambiguation. They are looking for relevant relationships between the entities named in the text. Basically making sure there is authority control in the text. They find that correlations between their query context and categories from Wikipedia's taxonomy provide useful information for disambiguating named entities. (Tags: wikipedia, extraction, taxonomy)	Extraction model Taxonomy Ranking Authority control Content relationships

Reference	Synopsis/Findings	Themes
Capocci, A., Servedio, V. D. P., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S., Caldarelli, G. (2006). *Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia*. Physical Review E, 74(3).	This article presents the properties and growth of Wikipedia visually. Explaining the graph: it has vertices (entries in Wikipedia) and edges (hyperlinks). The analysis shows a topological graph being in a bow-tie shape (much like the WWW). Looking at six Wikipedia's (Portuguese, Italian, Spanish, French, German, English) with vertices ranging between 8,000 to over 300,00 and edges between 51,000 to over 5,000,000. A large number of incoming and outgoing edges (hyperlinks) increase the probability of acquiring new edges. In analyzing the bow-tie graph, Wikipedia shows a large interconnection since the vertices (entries) all go to each other. In taking a snapshot of the edges deletion rate, it shows that edge deletion is rare and fast. The shape and correlations between the in- and out-components in Wikipedia show it follows a preferential attachment rule. Wikipedia's growth can be described using local rules like preferential attachment. (Tags: wikipedia, growth)	Content Content growth Linkage

Reference	Synopsis/Findings	Themes
Carillo, K.D. (2006). Creating high quality products in open content virtual communities: A functional group perspective based on the time, interaction, and performance theory. Unpublished thesis, Concordia University, Montreal, Quebec, Canada.	This thesis looks at the quality of products in open content communities like wikipedia. Looks into what makes open content communities successful through the lens of groups. Describes the input-process-output factors that the research is based on (see figure 1 page 14). It looks at the three group factors of input, process, and output in terms of groups (see figure 2 page 24). Hypothesizes that the study of the Wikipedia community will provide a clear understanding of open content group functioning in terms of time, interaction, performance theory. The model tests the relationship between input variables: organizational (Organizational Support), group-related (Group Size, Shared Experience and Group Heterogeneity) and individual (Member Competency and Member Activeness); process variables: group production, group well-being and member support: and an output variable: group effectiveness. The study looked at featured articles on Wikipedia with a final count of 10,000 being looked at. The final count included featured articles and a random selection of other articles that were over 5,000 characters. A quantitative study using partial least squares was conducted. The results show that organizational support is positively related to group well-being. In Wikipedia this is supported through peer review requests. Group production was not affected by organizational support, but organizational support plays a role in group effectiveness. Organizational support seems to be more about quality than quantity for open communities. The influence of group size is essential to open content projects. Group size has a synergistic effect on group activities. Shared experience was the strongest predictor of group processes and effectiveness, but it is not related to group well-being. Group heterogeneity did not have any influence on group well-being and member support functions. This finding is probably due to the high degree of heterogeneity among group members. Member competency was found to not be related to any group process variables. Administrators are more likely to be present in articles that have had conflicts. A high presence of administrators on an article improved quality. Need further research in this area. Open content group production is stimulated by large groups, and group well-being is affected by large groups, active members, and community support. (Tags: quality, open content, wikipedia, group factors, input-process-output)	Content quality Community Peer review Group characteristics User characteristics

Reference	Synopsis/Findings	Themes
Cedergren, M. (2003). *Open content and value creation*. First Monday, 8(8).	This paper describes three open content projects: Open Directory Project, Wikipedia, and Rick Prelinger's archive. The author creates value models for each project and then compares the three to draw conclusions about open content. Open content is defined as content made for not-for-profit with the intention of making the content improved and distributed to others. Distribution over the internet is an important cornerstone to open content. Value creation issues are essential for open content and these issues are looked at with this article. The general questions to be answered are: What value is exchanged between open content parties? What is the "payment" for using open content? What does a value model look like? The author considers value chains and how they can be used to help improve online content. Using the e3value method (a method to model value chains), the author looks at the three projects listed above. This method allows you to trace value changes in digital content. Results: Value chains are not described by a single scenario path, the needs of the customer are not equivalent to what is offered, producers and creators decide what to offer based on what they want to offer; so the producers are controlling the value chain, which makes it less linear. After interviewing one suitable person for each project (this person plays a dominant role in the value chain), the author describes driving forces behind open content; stimulating to work together, important to learn new stuff, feedback potential, intrinsic motivation, altruism, no interest in media business, publicity potential, end user benefits, interest in publishing facts, share common goals with a community, work with that community. In using value chains to analyze these three projects, the author has found many driving forces behind users using, changing, and distributing open content. (Tags: wikipedia, online encyclopedias, value model, driving forces) (Source: http://www.firstmonday.org/issues/issue8_8/cedergren/index.html)	User incentive User benefit Open content Content quality Participation Community

Reference	Synopsis/Findings	Themes
Chesney, T. (2006). *An empirical examination of Wikipedia's credibility.* First Monday, 11(11).	This study examines the credibility of Wikipedia looking at the author, article, and the site as a whole. The author sets up two conditions, one where participants (research staff and phd students) are given an article in their content area (thus making them an expert on the information), the other they are given a random article. They are asked to read an article and evaluate its credibility, the credibility of the author, and the credibility of Wikipedia as a whole. There were no differences found between the two groups in terms of their perceived credibility of the author or Wikipedia as a whole. There was a difference between the groups in terms of the perceived credibility of the articles. The experts found the articles to be more credible than the non-experts, suggesting that Wikipedia's accuracy is high. (Tags: wikipedia, credibility) (Source: http://www.firstmonday.org/ISSUES/issue11_11/chesney/)	User credibility Content quality Expert vs. novice

Reference	Synopsis/Findings	Themes
Cosley, D.R. (2006). Helping hands: Design for member-maintained online communities. Unpublished dissertation, University of Minnesota, Twin Cities, Minnesota.	This thesis explores the potential, problems, and design of member-maintained communities. It investigates how to design communities so members will participate for a better community. As part of the thesis the author explores using intelligent task routing for improving community contributions in Wikipedia. Although there are many articles that need reviewing and input (as seen on Wikipedia's community portal page), it is difficult to match users interests with needs in the community. In work earlier in the thesis, the author found that knowing a little bit about the users will help in encouraging them to contribute. Suggests that if intelligent task routing helps people find more articles to contribute to, they will become experienced contributors faster. I terms of recommending articles to Wikipedia users, they were not happy to have automated recommending that were dependent on articles they had made edits to. It ended up being that most of the recommendations were ones the user had created in the first place. It made them very unhappy and not wanting to contribute. They wanted to see more and different recommendations. The author wanted to use SuggestBot to welcome new members in the Wikipedia community to suggest more articles that they could edit to increase their editing contributions so they could become experienced contributors. Initial work found that users did not edit suggested articles, and reasons for this behavior included the note taking them to the page with suggestions was very small and not easily noticeable, wanted to use a more experimental design so users were not asked to participate-it was a random assignment and probably seen as spam. SuggestBot was redesigned so that only articles that needed work done on them were suggested, suggestions went directly to the users page, and allowed users to request suggestions rather than automate them. Based on Wikipedia users behavior, they find Suggestbot useful. Finding: users are editing more and becoming experienced contributors and do so more often when suggested editing opportunities are within their interests. (Tags: member maintained communities; online communities; intelligent task routing)	Community User control Recommenders

Reference	Synopsis/Findings	Themes
Devgan, L., Powe, N., Blakey, B., Makary, M. (2007). *Wiki-surgery? Internal validity of Wikipedia as a medical and surgical reference.* Journal of the American College of Surgeons, 205(3), S76-77.	Little is known about the accuracy of health related content in Wikipedia. It is now popular for learning about medical and surgical problems. This study identified entries with commonly performed procedures in the United States. The articles were evaluated by two independent reviewers to determine quantitative metrics on the quality. They looked at number of edits, unique editors, references, accuracy, completeness, discussion of risks, and suitability for patients. Looking for 39 procedures an n=35 were identified on Wikipedia with n=30 of those 35 deemed appropriate for patients. All 35 presented accurate content, but n=22 were free of critical omission. Many addressed procedure indications (n=34), but only an n=22 discussed risks. There are a correlation between an entries quality and the number of edits performed on it. They conclude that Wikipedia is an accurate but incomplete medical reference for patients. (Tags: wikipedia, medicine, surgery, accuracy)	Content quality

Reference	Synopsis/Findings	Themes
Ernigh, W., Herring, S.C. (2005). *Collaborative authoring in the web: A genre analysis of online encyclopedias*. In proceedings of the Thirty-Eighth Hawai'i International Conference on System Sciences, Los Alamitos: IEEE Press.	This study tries to fill the gap of understanding the content quality between Wikipedia and Everything2, along with determining how the different mechanisms for promoting quality on these two sites gives rise to their features. Data for this study comes from nodes or entries on the sites. In 2004, a random list of nodes were gathered from both sites, from 100 generated the list was pared down to 15 nodes from each site that contained 100 words or more. These ranged in content from people to places to things. Also included for analysis were talk pages from Wikipedia that corresponded to the 15 nodes (they are technically part of Wikipedia) and entries from the online 6th edition of the Columbia Encyclopedia (for comparison with user-created and expert-created content). That brought them to a total n=49 nodes. Quantitative analysis was done on the nodes measuring content variability in the degree of formality of language. This was done because it has been validated in other studies and is a good predictor of genre. For analysis purposes, a frequency count was conducted on the nodes words and parts of words to indicate formality of the words. Factor analysis was then done on these frequency counts to identify different genres following work done by Biber, 1998. An ANOVA and regression was then conducted on the factor scores against the source and node variables to look for correlations. Qualitative observations on the kinds of information in the nodes were also used to enrich the quantitative analysis for better characterization of the content. Findings: averages show differences in entry and word length; Arranged along a continuum Wikipedia discussion and Everything2 have the longest entries with the smallest words, Columbia Encyclopedia has the shortest entries and the longest words, with Wikipedia in between. So Wikipedia discussion and Everything2 are informal, Wikipedia being more formal, and Columbia Encyclopedia the most formal in language. Qualitative observations support this finding that there are content presentation differences between the sites. Wikiedia entries are homogenous and present a neutral point of view on the topic. Results: This study has shown through a comparison of sites a continuum of formalization and standardization. Wikipedia is different from Everything2 in that it is more formal and standardized. Wikipedia is statistically indistinguishable from the print encyclopedia in terms of formality features. Wikipedia and Everything2 are similar in many ways and very different in others. To classify them in genres the authors suggest that they have structural and functional similarities (aim to be repositories of knowledge, are available online, contents are searchable, use hyperlinks in the entries, created by multiple non-expert authors who are part of a community of practice, and are consulted by internet users. They are different in editorial control, guidelines, and entry styles. So, they are both members of the 'online knowledge repository' genre. Wikipedia is in the sub-genre of wikis and collaboratively-authored documentations (like FAQ). Everything2 is in the sub-genre of collaborative content systems like Slashdot or Fark where reputation metrics are used. Why Wikipedia is mostly indistinguishable from the Columbia Encyclopedia is considered: 1) Wikipedia users have norms and expectations of what they think an online encyclopedia should be like 2) the norms are enforced through the dedicated community of	Content quality Structure Content characteristics

Reference	Synopsis/Findings	Themes
Forte, A. & Bruckman, A. (2005). *Why do people write for Wikipedia? Incentives to contribute to open-content publishing*. GROUP 05 workshop: Sustaining community: The role of design and incentive mechanisms in online systems.	This article takes a qualitative approach to understanding why people contribute to Wikipedia. Two rounds of interviews were conducted with n=22 writers in fall 2004 and spring 2005. Contributors seek to collaboratively publish true facts, just like scientists feel about their field of study. Using established models of incentive systems in the scientific publishing community; the study seeks to understand how open-content publishing fits into the above model. Giving credit exists in Wikipedia and empowers individuals in the community. But the nature of the writing, the wiki technology, and the community values bring important and different aspects to the established incentive systems. The biggest difference being the indirect attribution of authorship. Although identity of authors is somewhat hidden, among the contributors is a sense of recognition. Users often claim ownership on their user pages, even though the culture forbids authorship claims. Administrators influence the character of Wikipedia, but they are voted in as an administrator, and thus require credibility in the community to be such. Whether or not a name should be made for oneself is debatable in the Wikipedia community. That is a stark contrast to the scientific community. Implications: online communities must structure participants contributions to sustain involvement. A need to invest in the community needs to be apparent. (Tags: wikipedia, incentives) (Source:http://www.google.com/search?hl=en&q=why+do+people+write+for+wikipedia%3F+incentives+to+&btnG=Google+Search)	User contributions User Incentives Community Reputation

Reference	Synopsis/Findings	Themes
Giles, J. (2005). *Internet encyclopedias go head to head*. Nature, 438(7070), 900-901.	This is a special report done by Nature comparing Wikipedia to the Encyclopedia Britannica. The main question asked here, how do users know that Wikipedia entries are as accurate as established sources like Britannica? This study was done by peer review and found that high profile examples of inaccurate information are more the exception than the rule for Wikipedia. Entries were chosen from both encyclopedias on a broad range of scientific disciplines and sent to an expert for review. It was a blind study, so reviewers did not know which encyclopedia the content was coming from. There were a total of 42 reviews sent back. Many factual errors, omissions or misleading statements were found in both. Several reviewers noted that the quality of the writing or readability in different articles was low, but a stong point for Wikipedia is the quickness that this can be changed and improved (which was not considered in this study). Having experts be part of the editing process would boost the reliability of entries in Wikipedia. People need to understand that print encyclopedias are not perfect and will have flaws. (Tags: wikipedia; Britannica; accuracy) (Source: http://www.nature.com/nature/journal/v438/n7070/full/438900a.html)	Comparison Content quality Accuracy

Reference	Synopsis/Findings	Themes
Halavais, A. & Lackaff, D. (2008). *An analysis of topical coverage in Wikipedia*. Journal of Computer-Mediated Communication, 13(2), 429-440.	This study presents efforts to map the diversity of content (topical scope and coverage) in Wikipedia to better understand is representation as collected knowledge. This study analyzes the distribution of Wikipedia on two levels: first at an overall level using full articles in the English Wikipedia, and secondly at article level in three specific academic fields. Method: Using a random sample of n=3000 articles from the English Wikipedia in spring 2006, the articles were classified according to Library of Congress categories. Two coders looked at length of the article and the number of edits. Compared to books in print, Wikipedia shows a variety of topics with a variety of lengths. To give an idea of article depth, another analysis was conducted on the length of the articles and the number of edits they have had. It is known that articles with more edits usually are higher in quality. The authors then compared Wikipedia with other reference sources, the Encyclopedia of Linguistics, New Princeton Encyclopedia of Poetry and Poetics, and Encyclopedia of Physics. These were compared by looking at headwords for topical coverage. Each headword used was mapped through Google to Wikipedia and vice versa. There were a considerable number of topics that were not found in Wikipedia, showing that its topical coverage is more limited than print encyclopedias. As Wikipedia's content expands according to the users, this is not surprising. Although it lacks the structure depth that other encyclopedia's have, Wikipedia has the potential due to its unique construction to expand to be larger and more comprehensive than other encyclopedia. (Tags: wikipedia, topical coverage, scope) (Source: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1083-6101.2008.00403.x)	Diversity Content scope Content quality

Reference	Synopsis/Findings	Themes
Hammwohner, R. (2007). *Interlingual aspects of Wikipedia's quality*. In proceedings of the International Conference on Information Quality, Cambridge, MA, November 9-11.	This article takes a look at the information quality of Wikipedia. Using some established categories of quality (AIMQ, Wikipedia, Crawford, Stvillia) the study compares quality among different language Wikipedias. Categories considered: intrinsic information quality, contextual information quality, representational information quality, accessibility information quality, scope, format, uniqueness, authority, accuracy, currency, indexing, and relevance. This article wants to know if the current models work for more than the English Wikipedia. This research is important for improving retrieval quality in Wikipedia. The focus here is on finding information patterns that can be picked up by bibliometrics or webometrics. This study focuses on interlanguage differences. Data is extracted from the English, German, French, and Italian Wikipedias. General features, reliability of links, and quality issues with structure and the category system. Looking at all featured articles plus n=1100 random articles, the comparison of the Wikipedias shows that the differences between standard and featured articles is more than the differences between languages. So, theoretically the process of model construction proposed by Stvilia et al can apply to all languages in Wikipedia. The presence of interlanguage links appears to be an indicator of information quality since featured articles have more of them than standard articles. Looking at consistency and usability in articles and categories assigned to them shows with n=463 articles shows that more complex category systems are harder to use and are more susceptible to inconsistencies. The simple methods used in the pilot studies here suggest that the methods are able to detect and solve quality problems in Wikipedia. (Tags: wikipedia, quality)	Content quality Structure Accuracy Authority Format Searching

Reference	Synopsis/Findings	Themes
Head, A.J. (2007). *Beyond Google: How do students conduct academic research?* First Monday, 12(8).	This study looks at how students use the internet and library when conducting research. Although it is not a study specifically on Wikipedia, the findings within it and the year it was done reflect some concerns that academic students still have with the online encyclopedia. It does have limitations as it may not be generalizable to all students in higher education, but is still something to consider when looking at needing to disseminate information about the quality and reliability of Wikipedia. The study was conducted from January to May 2007. Students majoring in humanities and social sciences at a Saint Mary's College of California were interviewed (n=13) and surveyed (n=178) about their process of conducting research. Content analysis of 30 research assignments over the last two years was also conducted with a reliability coefficient of .928176. Results show that most students first went to the library website and as need then used Yahoo!, Google, and Wikipedia. Reasons for doing so were not considering the results their reliable, and that the source wouldn't be an acceptable resource to cite. (Tags: research, use of internet resources) (Source: http://www.firstmonday.org/issues/issue12_8/head/index.html)	Library Content quality Accuracy
Hepp, M., Siorpaes, K., Bachlechner, D. (2007). *Harvesting wiki consensus: Using Wikipedia entries as vocabulary for knowledge management*. IEEE Internet Computing, 11(5), 54-65.	This study aims at using Wikipedia entries with Dublin Core attributes as a vocabulary for knowledge management. The study looked at whether Wikipedia's URI's changed over time in meaning; then they assessed Wikipedia's current content ontological nature. The study used a sample of n=150 pages from the English Wikipedia, all retrieved before January 13, 2007. Testing URI's stability was done through comparison of versions over a historical time period of first version to current; they found that even though Wikipedia is a changing place, a stable community consensus exists for the meanings of most URI's. Thus, open communities seem to be able to achieve a consensus about named conceptual entities as lightweight ontological agreements in an unsupervised fashion. (Tags: wikipedia, categories, knowledge management)	Ontology Community

Reference	Synopsis/Findings	Themes
Holloway, T., Bozicevic, M., Borner, K. (2007). *Analyzing and visualizing the semantic coverage of Wikipedia and its authors*. Complexity, 12(3), 30-40.	This study maps out the categories of the English Wikipedia. Pulling data from the November 5, 2005 data dump, the authors looked at the cur and categorylinks tables. There is an obvious increase in articles, categories, and contributors over time. The study created a category base map based on the co-occurrence of categories assigned to articles. From that map major semantic topics, last edit time, and topic coverage of major authors are revealed. The maps reveal that when co-occurrence of categories is considered as a measure for category similarity, categories cluster naturally revealing the content coverage of Wikipedia. It also shows that the category structure is well maintained. (Tags: wikipedia, semanitcs)	Semantic metadata Content Ontology Structure

Reference	Synopsis/Findings	Themes
Johnson, B.K. (2007). *Wikipedia as collective action: Personal incentives and enabling structures*. Unpublished Masters of Art thesis, Michigan State University, East Lansing, Michigan.	This thesis looks at why people contribute to Wikipedia and how the challenges the online encyclopedia faces with this community are solved. A convenience sample of n=27 editors was used. The sample were of higher performing editors who could shed light on the inner workings of Wikipedia. All contributors are editors of the English Wikipedia. A qualitative constant comparative method was used to analyze the phone and email interviews conducted. The study aimed to understand the community of contributors and what technological and social structures exist to aid their ability to maintain the encyclopedia. Results: editors gradually become Wikipedian, they move from being peripheral users to engaged editors. For a user to move from peripheral user to editing a Wikipedia article was predicated by interest in the topic. There is little cost involved to the user to actually edit articles. Sharing personal knowledge is a way of advocating personal interests so the quality of information is higher. Users began editing to correct or add to the information and found that being able to freely do so was encouraging. Participants also noted that finding sources to cite and doing research in an area of interest was another way to learn more about the topic. Other reasons for editing: ease of use, notoriety of the site, developing writing skills, feelings of obligation to reciprocate information on the site. A universal response was that editing behavior was driven by personal interest. Most users helped with vandalism clean up or other maintenance behaviors. Reputation is an important part of editing and ownership to the content. Editors get to know each other and interact through behind the scenes discussions and user talk pages. User pages reflected a sense of identity and reputation for the users. Positive reinforcement for editors is found in feature articles and other informal congratulations among the community. These though are very motivating and encouraging to editors. Socializing through the community is positive, and there is a consensus that a community must exist to support Wikipedia, but that the community shouldn't come first. Socializing comes mostly in the form of discussion of articles and policies. Discussions take place on talk pages and the Wikipedia namespace. There are also numerous mailing lists users can be part of. The community based around writing articles was named as an enabling characteristic of the encyclopedia. Many describe that avoiding conflicts is best, but it is inevitable to have disputes over some content. The transparency of Wikipedia is another great attribute to the site. Self satisfaction was named as a reason to continue editing and using the site. They were quick to stress the encyclopedia is a work in progress. One strong weakness was in content of biographies of living people. Concerns with the future of Wikipedia were worries about lawsuits of the Wikimedia Foundation. Many noted that Wikipedia works because of many factors that all came together at the right time. (Tags: wikipedia, user motivation, online community, organization)	User contributions Community User incentive Reputation Content quality

Reference	Synopsis/Findings	Themes
Kittur, A., Chi, E., Pendleton, B.A., Suh, B., Mytkowicz, T. (2007). *Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie*. Viktoria Institute. Retrieved on 2008-02-15.	This study looks at the history of Wikipedia participation. It considers the elite vs. common user and who contributes the most. Data is collected using a history dump from July 2, 2006. Administrator influence was considered first, as they are a small elite group. Looking at the history, administrator influence was calculated by taking the number of edits made that month by administrators divided by the total number of edits. There was a high influence of administrator edits (peaking at 59%) from 2002 to 2004, but it began to decline (low point of 10%) into 2006. After looking at several ideas why there was a decline, the data show that an increase in non-administrator edits supports the decline. Through more analyses of user groups the data show that there was a shift from elite users to common users in the editing of content. The elite editors didn't decline, but the commoners rose in numbers. This shift is also reflected in the site del.icio.us. This shift can be explained through the facts that early adopters are usually elite users. After they have refined things, novice users become the primary users of a system. Wikipedia is different from technology products because the product itself changes as a result of use. Thus, collaborative products have dynamic social systems that are fixed on the product. (Tags: wikipedia, history, users) (Source: http://en.wikipedia.org/wiki/Criticism_of_Wikipedia#See_also)	Participation User contributions User characteristics
Korflatis, N., Poulos, M., Bokos, G. (2006). *Evaluating authoritative sources using social networks: an insight from Wikipedia*. Online Information Review, 30(3), 252-262.	This study evaluates the quality of Wikipedia articles from a social networking perspective. It aims to provide a way of looking at the quality of the articles for further evaluation of the encyclopedia as it continues to grow. Using n=10 articles collected through a robot crawler, the number of contributors and their inter-relations (the amount of edits to the text) were analyzed. Through the analysis, it shows that contributors with higher inter-relation on the same topic have higher authority. They have done more research on the material they have contributed and have a higher degree of interest in that content. Contributors with lower inter-relations are more likely to have their edits rejected. (Topics: wikipedia, quality, social networks)	Content quality Authority

Reference	Synopsis/Findings	Themes
Kuznetsov, S. (2006). *Motivations of contributors to Wikipedia.* ACM SIGCAS Computers and Society, 36(2).	This study aims to show why people are motivated to contribute to Wikipedia. Three stages are used to gather information: empirical investigation, conceptual investigation, and technical investigation. The empirical investigation looked at two surveys of other social open source software and interviews with New York University students about Wikipedia. The two surveys outside of Wikiepdia show people who contribute to online open source sites do so to gain more knowledge for themselves about the software and to share their knowledge with others. The survey of NY University students consisted of n=102 students enrolled at the school. The majority of the students responded that they would contribute to Wikipedia and that half of those said they felt they had new information to contribute. More than 81% of those surveyed indicated they would make edits or corrections to Wikipedia, and 16% said they would contact a publisher of a print encyclopedia to report errors to be changed. Willingness to edit Wikipedia correlated with their use of the encyclopedia, 50% of weekly users would willing to add information. They indicated reasons for contributing to Wikipedia would be to educate others and raise awareness on information, to make a difference, and to give back to the Wikipedia community. Those unwilling to contribute cited lack of time as the primary reason for not contributing. (Topics: wikipedia, contributors) (Source: http://portal.acm.org/citation.cfm?id=1215942.1215943)	User incentive Community User contributions User characteristics

Reference	Synopsis/Findings	Themes
Lally, A.M. & Dunford, C.E. (2007). Using Wikipedia to extend digital collections. *D-Lib Magazine*, 13(5/6).	The University of Washington Libraries extended their digital initiatives to include adding links to Wikipedia about their collections. They acknowledge that patrons begin their searching outside of the library resource and wanted to know if using Wikipedia as a route into their collections would indeed increase their usage. After identifying the collections to start with, they searched Wikipedia for articles that already included similar information. When one was found an overview and link would be added. When one was not found, a new article on Wikipedia would be created. At first they were adding their information as an unregistered user, but were quickly flagged by IP address on the User talk page on Wikipedia. Wikipedia people monitoring pages encouraged them to register, which allows them the user to monitor changes on pages of interest (allowing them to have a heads up on maintenance of their additions). They found that any changes were edited and corrected by others before they could get to them. Looking back at their statistics over a years time, from when they began adding information to Wikipedia, the Library discovered that there were people coming to their collections through the links on Wikipedia. Graphs and screen shots are supplied. (Tags: wikipedia, digital collections, external links) (Source: http://www.dlib.org/dlib/may07/lally/05lally.html)	Digital library Connecting Professional use
Levin, C.M. (2006). *Venturing into the land of Wiki: Experiencing the people's information revolution*. Logos, 17(2), 99-103.	This is a small comparison study of the online Encyclopedia Britannica vs. Wikipedia on contemporary topics. The contemporary topics (n=14) were chosen from those that had recently appears in the news and in the "Year in Ideas" from the New York Times (December 2005 issue). The coverage of these topics in the two encyclopedias showed Britannica at 57% (8 out of 14) and Wikipedia at 93% (13 out of 14). This shows that even though the content quality of Wikipedia worries people, Wikipedia has an upper hand on Britannica in currency of entries due to the nature of the encyclopedia. (Tags: wikipedia, content coverage, contemporary topics)	Comparison Content quality Currency

Reference	Synopsis/Findings	Themes
Lih, A. (2004). *Wikipedia as participatory journalism: reliable sources? metric for evaluating collaborative media as a news resource.* In proceedings of the International Symposium on Online Journalism, Austin, Texas, April 16-17.	This study examines the growth of Wikipedia and analyzes the technology and community policies that have helped it to grow. It also analyzes Wikipedia articles that have been used in the media and looks at trends in using Wikipedia as a source. Data for the study was gathered from the English Wikipedia from January 2003 to March 2004. Looking at the edit history of the articles, reputation and quality are established from rigor and diversity domains. A comparison of n=333 benchmark subjects were used to make sense of trends and edit changes to understand the evolution of articles. They show a non-linear relationship between diversity and rigor. Once these were established the author analyzed cited articles in the news from January 2003 to March 2004, with n=113. When compared before the news citation, the distribution between rigor and diversity is much like the benchmark distribution, non-linear. After the news citations the number of articles that improved doubled. This growth can be linked back to news events on the topics or specific mention of the Wikipedia article in the press. This study shows that benchmarking of articles relying only on metadata and not content, can show the reputation of the entry. It shows the encyclopedia as a "working draft of history" and current news, something that print encyclopedias cannot do quickly. It shows a participatory side to journalism that is working when more traffic and users make their mark on the content. (Tags: wikipedia, growth) (Source: http://online.journalism.utexas.edu/papers.php?year=2004)	Growth Reputation Content quality Currency

Reference	Synopsis/Findings	Themes
Lorenzen, M. (2006). *Vandals, Administrators, and Sockpuppets, Oh My! An Ethnographic Study of Wikipedia's Handling of Problem Behavior*. MLA Forum, 5(2).	This is an ethnographic study looking at how Wikipedia identifies and corrects problem behavior on the site. The study looked at a single abuse detection page used by Wikipedians. It also examined how often behavior that was reported as vandalism was actually vandalism and how often users were banned. For several months the researcher visited the vandalism page and noted activity there. The researcher was completely invisible to the community, making no edits or comments. Logs of edits were also examined. The study found that not all reported vandalism is actually vandalism (n=16). Wikipedia has specific rules to follow when reporting vandalism, and with this low n and the large number of edits done each day, it looks like most people follow the correct procedures in reporting vandals. Bans (n=39) of users was also done while the researcher viewed the happenings of the site. The vandalism page also has a place for reporting sockpuppets (people creating more than one username for various reasons). There were n=30 reports of sockpuppets during this examination. This study shows that Wikipedia has a fairly robust way of tracking and dealing with vandalism, even though this one page is only one way the site deals with such matters. The Wikipedia community works hard to keep vandals out and the content clean and good quality. (Tags: wikipedia, problem behavior) (Source: http://www.mlaforum.org/volumeV/issue2/article2.html)	Editiorial control Editing tracking Content quality Community

Reference	Synopsis/Findings	Themes
Luyt, B., Kwek, W.T., Sim, J.W., York, P. (2007). *Evaluating the comprehensiveness of Wikipedia: The case of biochemistry*. In Asian Digital Libraries. Looking back10 years and forging new frontiers (p. 512-513). Springer Berlin/Heidelberg.	This study compares Wikipedia content scope in biochemistry with the same content found in the online version of the Encyclopedia Britannica. The data was benchmarked with an undergraduate textbook recommended by instructors at Nanyang Technological University. A checklist of keywords and concepts were taken from the textbook and applied to both online encyclopedias. One chapter was randomly chosen from each of the five sections of the textbook. The concepts were searched on the biochemistry page and if not found there they were searched for on the entire encyclopedia site. The number of concepts found in Wikipedia and Britannica were almost the same (23 vs. 22). A large number of concepts on separate pages were found as well, 33% for Wikipedia and 14% for Britannica. Concepts not found on the sites were 19% for Wikipedia and 33% for Britannica. A chi-square showed that the level of comprehensiveness in content coverage of Wikipedia was not higher than Britannica ($x^2=2.88$, $p>0.05$); the breath coverage of Wikipedia was higher than Britannica ($x^2=17.62$, $p>0.001$). Both references are similar in scope, and Wikipedia actually covers a larger number of concepts. (Tags: wikipedia, biochemistry, content scope)	Content quality Comparison Scope
McGuinness, D.L., Zeng, H., da Silva, P.P., Ding, L., Narayanan, D., Bhaowal, M. (2006). *Investigations into trust for collaborative information repositories: A Wikipedia case study*. In proceedings of the International World Wide Web Conference, Edinburgh, Scotland, May 22-26.	This study reports on an algorithm for reporting trust that is grounded in Wikipedia. Data was used from the main Wikipedia dump in December 2005. Using a random sample of n=50 articles in each area: featured, normal, and clean-up articles in the category of geography. Featured article had the highest link-ratio and clean-up the lowest. Articles with a link-ratio over 0.30 are deemed trustworthy. The featured articles had a link-ratio of 0.34, while the normal were 0.26 and the clean-up 0.21. Link-ratio value also depends on the links in the article. More links are made to less known definitions of words. Co-reference (names spelled out, instead of an acronym) also plays a part in link-ratio value. Acronyms do not help link-ratio value. Along with page rank, link-ratio can add to determining the quality of Wikipedia entries. (Tags: wikipedia, quality, trust) (Source: http://www.l3s.de/~olmedilla/events/MTW06_papers/programme.html)	Content quality Accuracy

Reference	Synopsis/Findings	Themes
Morrissette, R. (2007). *What do they know?* Knowledge Quest, 35(5), 14-17.	This article focuses on how to teach students to discern good and bad information on the internet. It also reports findings from a survey done of students and staff at an adult senior high school. With n=132 for students and n=34 for staff, the author found that staff used the internet more than students, that they used it for finding information and research, 40% of the students identified Wikipedia as the most reliable website while only 12% of the staff did so. The author believes the numbers are so different because of a hightened awareness among educators on the reliability of Wikipedia. This article although not a full blown research project, gives an insight into where Wikipedia stands with educators. (Tags: wikipedia; information quality; reliability; educators)	Content quality Reliability
Nielsen, F. A. (2007). Scientific citations in Wikipedia. http://arxiv.org/pdf/0705.2106	The work examines the outbound links from Wikipedia articles to articles in scientific journals with a comparison against journal statistics from Journal Citation Reports such as impact factors. The results show an increasing use of structured citation markup and good agreement with the citation pattern seen in the scientific literature though with a slight tendency to cite articles in high-impact journals such as Nature and Science. Wikipedia citation numbers showed high correlation with the JCR's numbers for the total number of citations to a journal. Wikipedia citation numbers correlated less with JCR impact factor and the JCR's measure of numbers of articles in a journal. These results increase confidence in Wikipedia as an good information organizer for science in general. (Tags: wikipedia, impact factor, statistics, citation) (Source: http://arxiv.org/pdf/0705.2106, pdf)	Impact Linkage Comparison

Reference	Synopsis/Findings	Themes
Nov, O. (2007). *What motivates Wikipedians?* Communications of the ACM, 50(11), 60-64.	This study explains and reports the results of a survey done with Wikipedians to explain their motivations for contributing to the site. Contribution was measured in hours per week spent contributing, motivation was measured by a volunteering motivations scale. Only valid responses from the Wikipedians emailed and asked to participate (n=151) were used. The average level of contribution was 8.27 hours per week. Top motivations were fun and ideology, where social, career, and protective were not strong motivations for contributing. Each of the six motivation levels correlated with contribution level. Contribution level did not correlate with ideology and social motivations. Why these correlations exist warrant more research to be done. But, recruiting of people to continue monitoring and editing on Wikipedia should focus on the motivations (fun, ideology, values, understanding, enhancement, protective, career, social). (Tags: wikipedia, wikipedians, motivation) (Source:http://portal.acm.org/citation.cfm?id=1297797.1297798&coll=GUIDE&dl=ACM&CFID=56478366&CFTOKEN=53048126)	User incentive Community User contributions
Ortega, F. & Gonzalez-Barahona, J.M. (2007). *Quantitative analysis of the Wikipedia community of users.* In proceedings of the International Symposium on Wikis, Quebec, Canada, October 21-23.	This article reports on a similar study to Kittur. Using Kittur's methodology, the authors analyze data from a more recent Wikipedia dump (November 2006 and April 2007). In relation to the previous study, there is a new peak at the beginning of the data, but that the percentage of edits still decreases. This data shows a stabilizing trend of the number of edits done by administrators. The data also shows what Kittur found with those who contribute small numbers of edits, they are still indeed the mass numbers of contributors. This reproduction shows a stable rate of all the factors reported in the Kittur study. This study extended the analysis to non-English Wikipedias. The other Wikipedias also show an increase in the number of edits, but many show a stable number of edits by administrators. This may be due to the differences in selecting administrators. Overall the history of the Wikipedia community is showing an increase in the number of editors and in the English version these editors are not all administrators. The community is growing and thriving. Quantitative statistics can explain the patterns of the users. (Tags: wikipedia, contributors) (Source: http://www.wikisym.org/ws2007/proceedings.html)	User contributions Community Edit tracking User behavior

Reference	Synopsis/Findings	Themes
Pfeil, U., Zaphiris, P., Ang, C.S. (2006). *Cultural Differences in Collaborative Authoring of Wikipedia*. Journal of Computer-Mediated Communication, 12(1), 88-113.	This article explores the relationship of national culture with cultural influences on Wikipedia. Through the use of content analysis, the behavior of wikipedians in the non-English Wikipedia are explored. The focus is placed on looking at the patterns of change on the site and the cultural backgrounds of the contributors. Online cultural preferences have been studied, so this study expands that to wikis. The authors base the analysis on cultural ideas from Hofstede (individualism index, power distance index, masculinity index, uncertainty avoidance index). Using the sequence of the Wikipedia pages, the contributions of the participants to these pages were investigated. They were sorted into categories and correlated with Hofstede's four dimensions. Four Wikipedias were studied: French, German, Japanese, and Dutch. The study assumes that language is highly correlated with culture. After conducting correlations between the data found and Hofstede's dimensions, it shows that cultural differences in the physical world also exist in the virtual world. (Tags: wikipedia, culture)	User behavior User characteristics Online content Physical content

Reference	Synopsis/Findings	Themes
Purdy, J.P. (2006). *Digital archives and the turn to design*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Illinois.	This dissertation examines three different online areas that act as archives of information; JSTOR, Wikipedia, and Turnitin. The point of the document is to look at these archives for their usefulness in historical research. The author argues as well that these archives drive and shape writing and research practices. In terms of Wikipedia, the author shows that through the history pages (or versions of the wiki) analysis of the content quality and the trustworthiness of the authors can be analyzed. In looking at the technology, the author studies the functionality of it and the design approach it takes for the users. Wikipedia has an abundant amount of data that is freely available to users and researchers. It makes for a great place to conduct research on online activity. It also is a nice place to practice writing. Even though there are not set standards by archivists or experts, the community values good authors. There are still concerns of bias, especially just in the articles that are selected to be added or edited. It makes some topics especially robust and others not. There are of course problems that are still not worked out. When someone cites a Wikipedia page and gives the date of the version they refer to, users still are taken to the newest version and may not see the version the author intended them to see. As Wikipedia does not follow scholarly conventions citations are a challenge. So, Wikipedia relies on large numbers of contributors and editors to keep the quality high, so scholarly experts are not as necessary. Although this is a draw back to some, Wikipedia has built the request for scholarly work into its design. Users who write are asked to reflect on their writing, it's a self-reflexive way of writing. Wikis can serve as tools for the writing process, and Wikipedia does that well. It is a great place for collaboration too. (Tags: wikipedia, digital archive, research data, online community)	Community Content quality Reliability User contributions User credibility
Rafaeli, S., Hayat, T., Ariel, Y. (2005). *Wikipedians' sense of community, motivations, and knowledge building: a cross-cultural study*. In proceedings of the International Wikimedia Conference, Frankfurt, Germany, August 4-8.	This study looks at why Wikipedians contribute to Wikipedia. Data for this study was collected through an online questionnaire to English and Hebrew Wikipedia contributors. The sample involved n=120 active editors. The survey asked about participants perceptions of Wikipedia and the community. Motivation was measured on 8 items that are possible motivations. The data shows the strongest motivators are cognitive (learning new things and sharing knowledge). The survey data also shows how the contributors feel about being part of the community, that they are part of it. (Tags: wikipedia, contributors, community, motivations) (Source: http://meta.wikimedia.org/wiki/Transwiki:Wikimania05/Paper-YA1)	User contributions Community User incentive

Reference	Synopsis/Findings	Themes
Schroer, J. & Hertel, G. (2007). *Voluntary engagement in an open web-based encyclopedia: Wikipedians, and why they do it*. Retrieved February 18, 2008.	This study aims to determine why Wikipedians contribute to the online encyclopedia. Using an online survey to German contributors of the German Wikipedia (n=106), participants are asked about motivation, engagement, and satisfaction as users. Using two models, Klandermans Model of social movement participation and Job Characteristic Model, to explain reasons for contributions the data from the survey was analyzed. The results show satisfaction with engagement of Wikipedia is a balance of costs and benefits in the identity of the community and task characteristics. Engagement was correlated positively with intrinsic motivation, enjoyment, and information sharing. (Tags: wikipedia, contributors, engagement, community) (Source: http://www.abo.psychologie.uni-wuerzburg.de/virtualcollaboration/)	User contributions User incentive Community
Spoerri, A. (2007). *What is popular on Wikipedia and why*? First Monday, 12(4).	This study aims to know what topics are most popular on Wikipedia and why. It is known that the users drive the content of the encyclopedia. The tool WikiCharts was used to find the top 100 visited pages of the encyclopedia during the months of September 2006 to January 2007. The pages were categorized to identify the major topics of interest from the five month span. The data then showed that these topics correlated with the top searches on the Web. Wikipedia titles were searched in major search engines to determine their position in the top 10 results of the search. This showed that 87% of the Wikipedia pages are in the top 3 results in Google, with 72% for Yahoo and MSN. This study shows how search engines fuel the growth of Wikipedia, especially Google. Another article by this author in the same issue shows this data visually: Spoerri, A. (2007). *Visualizing the Overlap between the 100 Most Visited Pages on Wikipedia for September 2006 to January 2007*. First Monday, 12(4). http://www.firstmonday.org/issues/issue12_4/spoerri/index.html) (Tags: wikipedia, popular articles, topics, popularity) (Source: http://www.firstmonday.org/issues/issue12_4/spoerri2/index.html)	Content popularity Search engine Comparison

Reference	Synopsis/Findings	Themes
Stein, K. & Hess, C. (2007). *Does it matter who contributes: a study on featured articles in the German Wikipedia*. In proceedings of the Conference on Hypertext and Hypermedia, Manchester, UK, September 10-12.	This study is concerned with the quality of the German Wikipedia featured articles. These articles are in three categories: excellent, worth reading, and other. It explores whether the number of contributors makes a difference in quality or who makes the edits makes a difference. The data comes from the April 2, 2007 dump. To measure quality in the articles, the number of edits and the reputation of the authors making edits. The data show that excellent pages are have the highest ratings according to the measures indicated above. Worth reading are next with a much higher rating than other pages. According to the data, it looks like when a page is nominated to be an excellent or worth reading status a group of authors do edits to improve the page. As the data shows, there is a relationship between the quality of the article and who does editing on it. It also shows that there is a relationship in quality with the number of edits to a page. In conclusion, the data shows that both who does editing and the higher number of edits contribute together to the quality of featured articles. (Tags: wikipedia, contributors, quality)	Content quality Reputation Page ranking User contributions User characteristics
Stvilia, B., Twidale, M.B., Smith, L.C., Gasser, L. (2007). *Information quality work organization in Wikipedia*. Journal of the American Society for Information Science and Technology, accessed February 18, 2008.	This study is a follow-up to the 2005 study conducted by the same author. Looking to define information quality in a practicing community, Wikipedia supplies large amounts of data for creating categories that can inform other databases on quality. With a social order like the Wikipedia community, it is the perfect place to create and test quality. In Wikipedia, quality is assessed through collaborative models. This qualitative study helps to explain how a practicing community perceives quality. The Wikipedia community takes quality seriously, and through discussion pages and evolving policies how one community defines and improves quality are seen. In Wikipedia quality issues are tied very closely to the data itself, and especially important is that the data is open and easily accessible to the public, making research easier. (Tags: wikipedia, information quality) (Source: http://en.wikipedia.org/wiki/Criticism_of_Wikipedia#See_also)	Content quality Community

Reference	Synopsis/Findings	Themes
Stvilia, B., Twidale, M.B., Smith, L.C., Gasser, L. (2005). *Assessing information quality of a community-based encyclopedia*. In proceedings of the International Conference on Information Quality, Cambridge, MA, November 4-6.	This study is concerned with information quality, specifically in Wikipedia. Using a random sample of 834 articles from three 2005 data dumps, an analysis of the articles and edit history were examined. Also included were titles from featured articles and histories and discussion pages of the featured articles and random sample. From there an analysis of the random articles was conducted to create profiles of nineteen quality measures, which ended up being seven IQ metrics using factor analysis. When applying these measures, high and low quality articles are distinguishable and can be classified. These classifications show that feature articles have higher quality than other articles. These findings are corroborated by the quality the community gives to the articles. (Tags: wikipedia, information quality) (Source: http://mitiq.mit.edu/iciq/ICIQ/iqpapers.aspx?iciqyear=2005)	Content quality Community
Viegas, F.B., Wattenberg, M., Kriss, J., van Ham, F. (2007). *Talk before you type: Coordination in Wikipedia*. In proceedings of the Hawaii International Conference on System Sciences, Waikoloa, Big Island, Hawaii, January 3-6.	This article investigates the state of Wikipedia using the October 2005 English version data. Using the method of history flow (as used in previous research), the authors analyze the articles and edit histories. As with the previously analyzed data, vandalism was present, page size continued an upward trend in edits, and a drop in the frequency of edit wars. The analysis shows Wikipedia is resilient to vandals with fast repairs. The site has grown with the community evolving, mostly with the 20 different name spaces available for collaboration and policies. The site is becoming more diverse and complex. There are different pages serving distinct purposes for the community. The snapshot in this article shows how the encyclopedia and its community has grown and evolved over a few years time. The dynamics of it are what make it successful. Focusing on the talk pages to understand the changes in the site, and after coding a sample of pages, the study shows that a large amount of planning occurs on these pages. Some of the talk there is formal and policy driven. The happenings of the talk pages appear to play a crucial role in the behavior of the community. (Tags: wikipedia) (Source: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4076527)	Community Diversity

Reference	Synopsis/Findings	Themes
Viegas, F.B., Wattenberg, M., McKeon, M.M. (2007). *The hidden order of Wikipedia*. In Online Communities and Social Computing, Springer Berlin/Heidelberg.	This book chapter focuses on the order of online communities and the activities found within them. Since so many online communities do not have a specified structure or standard to follow, many people want to know how they work. Wikipedia is like other online communities where there is no set procedure, but the community has a procedural side that is well known to them and works for them. In looking at the featured articles process this study examines, through the theoretical framework of commons governance, how the process works in what seems like an unstructured community. Wikipedia has procedures and guidelines openly accessible to the public and other questions and answers can be found on talk pages. This study began with an investigation of these pages as well as interviews via email and phone to key players of the featured article process. The expectations of the featured article have changed over the years. Now it goes through a rigorous process of nomination and reviews. Templates (bits of wiki code) are used in the workflow of a featured article from nomination to featured article status. It is a rule bound process that has written policies, but is not a normal process. The differences include roles in the process that are filled by self-identified individuals (editors, reviewers, voters), and a non-hierarchical flow of information using the templates to indicate work being needed and other doing the work. This process uses collaboration and individual efforts based around quality assurance. The rules and policies in Wikipedia show a robust and thriving community that have changed over time. Wikipedia is an example of a self-governing institution that works. (Tags: wikipedia, order) (Source: http://www.google.com/search?hl=en&q=The+Hidden+Order+of+Wikipedia&btnG=Google+Search)	Community Structure User contributions Content quality

Reference	Synopsis/Findings	Themes
Viegas, F.B., Wattenberg, M., Dave, K. (2004). *Studying cooperation and conflict between authors with history flow visualizations*. In proceedings of Computer/Human Interaction Conference, Vienna, Austria, April 24-29.	This study uses a method called history flow to visualize the data from Wikipedia to gain a better understanding of patterns and activities of contributors on the site. Examining more than 70 different page histories up to the date of May 2003, revealed common patterns of collaboration and negotiation among editors. These patterns show some of the techniques that the community uses in to regulate behavior and to deal with disagreements on content. Some of these patterns include: vandalism and repair, anonymity versus names authorship, negotiation, and content stability. The patterns found show that Wikipedia has had great success in the community. Three thoughts that come from this analysis as to why Wikipedia has had success are it encourages community introspection, members watch each other, and the community must reach a consensus on content. There are many directions more research can go with Wikipedia. (Tags: wikipedia, visualization, history flow, contributors) (Source:http://www.google.com/search?hl=en&q=studying+cooperation+and+conflict+between+authors+with+history+flow+visualizations&btnG=Google+Search)	User behavior Community
Voss, J. (2006). *Collaborative thesaurus tagging the Wikipedia way.* Wikimetrics, 1(1).	This study examines the collaborative tagging system to classify articles on Wikipedia. Data was gathered from the English Wikipedia database January 2, 2006. The study measured and compared the structural properties of the data (articles and links). Along with the tagging in Wikipedia, the study compared classifications and other tagging systems like the Dewey Decimal System and del.ic.ious. Descriptors were grabbed from each article or post. The data show that Wikipedias category system is a thesaurus that has been collaboratively developed and is used to index the articles in the encyclopedia. Using collaborative tagging or category creation is new. Hierarchical systems have levels for top descriptors. In Wikipedia, these levels are normally distributed. Wikipedia uses collaborative tagging and hierarchical subject indexing in a unique way. (Tags: wikipedia, tagging) (Source: http://arxiv.org/abs/cs.IR/0604036)	Classification Tagging Structure Community

Reference	Synopsis/Findings	Themes
Voss, J. (2005). *Measuring Wikipedia*. In proceedings of the International Conference of the International Society for Scientometrics and Informetrics, Stockholm, Sweden, July 24-28.	This article gives an overview of the research opportunities in Wikipedia. It hands out basic statistics one Wikipedia through analysis of the fundamental components of the encyclopedia (articles, authors, edits, and links). It reports the characteristics that are found in Wikipedia and its architecture. It reports the growth of Wikipedia as exponential after an initial linear time, and that varies depending on the language looked at. Article size is lognormally distributed and the concepts covered in the articles have a thesaurus-like structure. There is potential for much more in-depth research and analysis of the content, authors, and structure of the encyclopedia open to anyone in the world who wants to tackle it. That it is freely available to anyone is a strong point of the wiki and should be appealing to many researchers. (Tags: wikipedia, research) (Source: http://eprints.rclis.org/archive/00003610/)	Research Structure Content

Reference	Synopsis/Findings	Themes
Wattenberg, M., Viegas, F.B., Hollenbach, K. (2007). *Visualizing activity on Wikipedia with chromograms.* In Human-Computer Interaction-Interact 2007, Springer Berlin/Heidelberg.	This book chapter reports on the study of examining how users of peer production systems use their time and what other online peer sites can learn from Wikipedia. The authors do this by looking at the activities of Wikipedia administrators. All active administrators as of October 2005 through August 2006 were selected, making n=509. To edit the huge amount of information in the edit histories, they use a visual technique called chromograms that display text in color coding schemes. This allows for analysis of patterns with visual ease. Each editor had a distinct chromogram activity, there were also many similar patterns. First, activity came in bursts. The bursts correlated with the type of edit they were doing; typo fixing or adding messages to pages. There were also calendar patterns: times of day and days of the week. There was also a pattern of switching between articles and tasks in a systematic way and in reactive duties. From the activities of the administrators, it can be learned that there is a large amount of diversity and activity by these people. They focus on a particular type of work and move through it in a systematic way. Other online communities can learn from the Wikipedia design by taking into account the flexibility of the technology and making it do what you want it do to, sorting mechanism or lists for activity help the system work, and the Wikipedia community has found a way to make the methods prototyped so they are using participatory design to run their system. (Tags: wikipedia, visual patterns, systematic strategies) (Source:http://www.google.com/search?hl=en&q=Visualizing+Activity+on+Wikipedia+with+Chromograms&btnG=Google+Search)	User characteristics User behavior

Reference	Synopsis/Findings	Themes
Wilkinson, D. M. & B. A. Huberman (2007). _Cooperation and quality in Wikipedia_. _International Symposium on Wikis_, Montreal, Quebec, Canada.	This study examines all 50 million edits made to the English Wikipedia articles. The analysis of the data includes showing the number of edits to an article and the number of people who edit the article. The data shows that the high quality articles are edited more, by more people, and with more cooperative behavior. Because the edited articles are driven by the interests of the community, the high quality articles are brought forward and are usually of high interest. This is important for Wikipedias future, because other sites have trouble sustaining this model. The communication and collaboration of the community is important to the sustainability of the article quality. Wikipedia has talk pages that are used for discussion of the articles. The differences between the featured and non-featured populations indicates that collaboration on an article is a good indicator of article quality. This research can also be seen as a journal article: Wilkinson, D.M. & Huberman, B. A. (2007). _Assessing the value of cooperation in Wikipedia_. First Monday, 12(4). http://www.firstmonday.org/issues/issue12_4/wilkinson/index.html) (Tags: wikipedia, quality) (Source: www.hpl.hp.com/research/idl/papers/wikipedia/wikipedia07.pdf)	Edit tracking Content quality Community Structure

Reference	Synopsis/Findings	Themes
Willinsky, J. (2007). *What open access research can do for Wikipedia*. First Monday, 12(3).	This study examines how much Wikipedia articles reference research and scholarship, and how available those citations are to users. The research assumes that Wikipedia's credibility and authority increases when users can easily access research and scholarship that is peer-reviewed. Using a sample of n=100 Wikipedia articles using the random button on the site. The demographics of the articles included stubs, clean-up, frequently edited. The study focused on the references in the entries under the headings sources, notes, related research, or external links. They were categorized as print or online, with the online links divided into freely available or not. Using a random subset of the above sample, an n=20, the study wanted to show to what degree the open access research was available through the above references in the articles. Looking for open access materials started as searches in Wikipedia and other search engines, but it was quickly determined that Google Scholar held all of the materials. The study then went back to Wikipedia and added links in the reference areas to the open access materials that were found. It is not definitive, but suggestive that the educational contribution that could come from open access research and scholarship could be a great contribution to Wikipedia. (Tags: wikipedia, research, open access) (Source: http://www.firstmonday.org/issues/issue12_3/willinsky/index.html)	Content quality Credibility Research Community User contributions Authority

Reference	Synopsis/Findings	Themes
Zeng, H., Alhossani, M. A., Ding, L., Fikes, R., McGuinness, D.L. (2006). *Computing trust from revision history*. In proceedings of the International Conference on Privacy, Security, and Trust, Ontario, Canada, October 30-November 1.	This paper uses the history of Wikipedia articles to determine the trustworthiness of the articles. A revision trust model may address these problems found with Wikipedia: article trust, fragment trust, and author trust. Using continuous notation from 0 (complete untrustworthiness) to 1 (complete trustworthiness), the articles and authors are evaluated. To evaluate the model, the English Wikipedia was examined. The sample consisted of 868 articles (40450 revisions). The data looked at feature articles, common articles, and clean-up articles. Authors are deemed trustworthy from the amount of edits they do that are not deleted. Feature articles have more revisions than any other group of articles, and clean-up articles have the lowest number of administrators editing them. To determine the trustworthiness of an article, the author and the number of revisions are most important. From the analysis, the authors believe Wikipedia is generally trustworthy because 1) most authors seem to have good intentions and 2) administrators have the responsibility and authority to settle disputes, prevent vandalism, and block inappropriate authors. Wikipedia has a good model set in place to create and keep trustworthy information and authors. (Tags: wikipedia, trust, quality) (Source: http://ebiquity.umbc.edu/paper/html/id/320/Computing-Trust-from-Revision-History)	Credibility Content quality Edit tracking User characteristics Structure
Zlatic, V., M. Bozicevic, et al. (2006). "Wikipedias: Collaborative web-based encyclopedias as complex networks." Physical Review E, 74(1).	This study looks at Wikipedia articles in different languages as complex networks. They view each article as a network with nodes (links) to corresponding articles. The sample consists of 30 of the largest non-English Wikipedias, with data from January 7, 2005. From the analysis of the articles and nodes, the authors found that many network characteristics are common in the Wikipedias studied. They include degree distribution, growth, topology, reciprocity, clustering, assortativity, path lengths, and triad significance profiles. All of these point to a unique growth pattern for Wikipedia, but that the pattern is universal for all of the different Wikipedias. The complex networks (different language Wikipedias) are in different stages of development. (Tags: wikipedia, other languages, complex networks)	Growth

Search Methodology

237 articles reviewed/54 fit inclusion criteria

Database/Journal	Search Terms	Records Retrieved
Chronicle of Higher Education (http://chronicle.com)	wikipedia	38
	Wiki(s), online encyclopedia	7
	wiki (internet)	2
	wiki, online communities	0
CiuteULike	wikipedia	479
	Wiki(s), online encyclopedia	495
	wiki (internet)	497
	wiki, online communities	496
	"wikipedia research"	2
	"wikipedia study"	1
	"wikipedia empirical"	1
	wikipedia empirical research	486
	"wikipedia empirical research"	0
	wikipedia analysis	493
	"wikipedia analysis"	2
	wikipedia measurement	493
	"wikipedia measurement"	0
	wikipedia quality	494
	"wikipedia quality"	36
Communications of the ACM	wikipedia	1345
	Wiki(s), online encyclopedia	171
	wiki (internet)	826
	wiki, online communities	611
	"wikipedia research"	5
	"wikipedia study"	1
	"wikipedia empirical"	1
	Wikipedia empirical research	187
Digital Dissertations	wikipedia	18
	Wiki(s), online encyclopedia	0
	wiki (internet)	0
	wiki, online communities	0
Education Full Text	wikipedia	54

	wikis, online encyclopedia	5
	wiki (internet)	54
	wiki, online communities	2
ERIC via Ebsco Host	wikipedia	12
	Wiki(s), online encyclopedia	2
	wiki (internet)	0
	wiki, online communities	2
ERIC via US Dept. of Education	wikipedia	14
	Wiki(s), online encyclopedia	2
	wiki (internet)	9
	wiki, online communities	1
First Monday (http://www.firstmonday.dk)	wikipedia	52
	Wiki(s), online encyclopedia	70
	wiki (internet)	44
	wiki, online communities	100
Google Scholar	wikipedia	95,300
	Wiki(s), online encyclopedia	5,960
	wiki (internet)	25,200
	wiki, online communities	12,600
	"wikipedia research"	26
	"wikipedia study"	9
	"wikipedia empirical"	1
	wikipedia empirical research	6,480
	"wikipedia empirical research"	0
	wikipedia analysis	24,400
	"wikipedia analysis"	5
	wikipedia measurement	6,850
	"wikipedia measurement"	1
	wikipedia quality	18,300
	"wikipedia quality"	42
Library Information Science & Technology Abstracts	wikipedia	333
	Wiki(s), online encyclopedia	43
	wiki (internet)	0
	wiki, online communities	12
Library Literature	wikipedia	60

	Wiki(s), online encyclopedia	2
	wiki (internet)	0
	wiki, online communities	1
Psych Info via EBSCOhost	wikipedia	9
	Wiki(s), online encyclopedia	0
	wiki (internet)	0
	wiki, online communities	0
Psychology and Behavioral Science Collections	wikipedia	16
	wiki(s), online encyclopedia	2
	wiki (internet)	0
	wiki, online communities	0
Web of Science	wikipedia	63
	wiki(s), online encyclopedia	0
	wiki (internet)	8
	wiki, online communities	0
Wikipedia (http://en.wikipedia.org/wiki/Criticism_of_Wikipedia#See_also)	wikipedia	145

Appendix D: Youtube (www.youtube.com)

Researcher: Sandie Waters

Literature Review

Reference	Synopsis/Findings	Themes
Cheng, X., Dale, C., & Liu, J. (2007). Understanding the characteristics of internet short video sharing: Youtube as a case study. Technical Report arXiv:0707.3670v1 [cs.NI], Cornell University, arXiv e-prints, July 2007.	Using a case study approach, this article looks at the characteristics of YouTube videos and the implications these have for the future of the site. During a 3 month period in 2007, 2,676, 388 videos were collected for the study. Characteristics of the videos examined included: video category, length, file size, access patterns, life span, views, ratings, and comments. The data set also allowed the authors to study growth trends and the social networking of the site. They found that the links to related videos that come from uploader's choices contribute to the social networking (suggests correlations for the videos as well as peer-to-peer distribution schemes. Social networking aspects included communities, groups, awards, user characteristics (number of friends and number of videos), and small-world network. From the data gathered about the videos and how the users connect through video sharing and linking, the authors take a critical eye to the future of the site and solutions to the problems of storage and access (caching), continued good user experience, and peer-to-peer challenges with short video sharing. (Source:http://scholar.google.com/scholar?hl=en&lr=&q=Understanding+the+characteristics+of+internet+short+video+sharing%3A+Youtube+as+a+case+study.+&btnG=Search)	Content Social networking Access Communities Sharing Groups Growth Networking Future

Reference	Synopsis/Findings	Themes
Gill, P., Arlitt, M., Li, Z., & Mahanti, A. (2007). YouTube traffic characterization: A view from the edge. Paper presented at the 7th ACM SIGCOMM conference on Internet measurement, San Diego, CA.	This article examines traffic characteristics of YouTube during a three month time period in 2007. Usage patterns, video properties, popularity, referencing characteristics, and transfer behaviors are compared to traditional web media. Suggestions for the future and improvements include decentralized resources for better user experience and metadata creation (exploiting Web 2.0 technologies). Looking at the top 100 most viewed videos for data, the authors draw insights into global and local popularity. Looking at the usage by students during this time period showed an increase in usage about 4 weeks into a new semester (higher education) where students are more settled with their coursework and usage decreases during school breaks. In looking at the social networking piece of YouTube, the authors looked at the rating system that shows how much users like or dislike a video. Suggestions for the future of YouTube include an examination of the workload patterns to plan and design for a better delivery infrastructure, strains on centralized resources, and utilizing Web 2.0 metadata availability. (Source: http://portal.acm.org/citation.cfm?id=1298306.1298310)	User characteristics User behavior Content Popularity Web 2.0 Rating system Future
Gueorguieva, V. (2007). Voters, myspace, and youtube: The impact of alternative communication channels on the 2006 election cycle and beyond. Social Science Computer Review.	This article discusses the implications of YouTube and MySpace on election campaigns and the control that is taken from the campaign and given to the users. Explains the tagging system that allows users to create keywords associated with a video, the comments area for user reflection, and users subscribing to videos of other users. Being user driven and popular sites, they have a large impact on election campaigns. The authors suggest that user demographic data show the implications of sites like these are impacting campaigns because they are gaining appeal across generations. Using these sites for campaigning in the past showed their usefulness for advertising and fund-raising. Specifically in YouTube, campaigns could use short videos that reached many people for great communication power. And all of this is free for users and campaigners. They are changing communication avenues and how candidates represent themselves, along with other campaigning challenges that force the importance of blended networking to include online and offline emphasis. (Source: http://ssc.sagepub.com/cgi/content/abstract/0894439307305636v1)	Political campaigning User control Tagging Commenting Subscriptions User characteristics Impacts Online communication

Reference	Synopsis/Findings	Themes
Halvey, M. J., & Keane, M. T. (2007). Exploring social dynamics in online media sharing. Paper presented at the International World Wide Web Conference, Baniff, Alberta, Canada.	During 2006, the online behavior of users was examined specifically for the methods of social interaction (linking to friends individual and in groups, and interactions through comments and subscriptions). The data shows that users visit the site to view videos more than upload them, widespread failure of users using the community options. This suggests poor data option for recommenders and personalization of the site. But, users who do use the options have high usage patterns. There is an indication of growth in using the social features as subscription numbers increase. (Source: http://portal.acm.org/citation.cfm?id=1242804)	User behaviors Comments Subscriptions Usage Recommenders
Madden, M. (2007). Online video. Washington, DC: PEW Internet & American Life Project.	This is a report from a study asking about online video usage. The study was conducted through phone interviews asking questions about a users experience and usage of online video. Included in it is information about users with YouTube and the characteristics that draw them to the site. These include free access, social applications for users, communication options, demographics of users, popularity, rating/feedback/comments, linking to popular/similar videos, and type (music videos). (Source: http://www.pewinternet.org/PPF/r/219/report_display.asp)	User characteristics Comments Linking User behavior Access Networking Popularity
Rainie, L. (2008). Pew internet project data memo (memo). Retrieved February 28, 2008 from http://www.pewinternet.org/pdfs/Pew_Videosharing_memo_Jan08.pdf.	Phone survey/interview conducted in late 2007 asking users about online video sharing sites (like YouTube). Shows an increase in usage of these sites, nearly twice as large at the end of 2007 as it was in 2006. (Source: http://www.pewinternet.org/pdfs/Pew_Videosharing_memo_Jan08.pdf)	User behavior

Reference	Synopsis/Findings	Themes
Silva, P. A., & Dix, A. (2007). Usability - not as we know it! Paper presented at the 21st BCS HCI Group Conference, Lancaster University, UK.	This article discusses and evaluates usability for YouTube. It looks at the site from a conventional usability perspective showing it's faults and failures indicating that users are not expected to return and use the site. It then evaluates YouTube from three perspectives: users, content, and design. From this perspective usability of the site succeeds because 1) users are content designers who are communicating and socializing on the web 2) the content of YouTube makes it successful with user personal broadcasting ability that is shared with a vast community 3) apparent bad design features of keeping the user in the system longer does not stop users. Crucial features include user control, engagement, and play. (Source: http://scholar.google.com/scholar?hl=en&lr=&q=Usability+-+not+as+we+know+it%21+silva&btnG=Search)	Usability Evaluation User characteristics Content quality Visual design User control

Search Methodology

Database/Journal	Search Terms	Records Retrieved
Chronicle of Higher Education (http://chronicle.com)	YouTube	0
	Video sharing + YouTube	12
Curtis Bonk	YouTube	1
Digital Dissertations	YouTube	4
	Video sharing + YouTube	1
Education Full Text	YouTube	23
	Video sharing + YouTube	0
ERIC via EBSCO Host	YouTube	905
	Video sharing + YouTube	0
Google Scholar	YouTube	5,980
	Video sharing + YouTube	1,980
Wikipedia (http://en.wikipedia.org/wiki/Youtube)	YouTube	1