



Faculty Publications

2009-03-12

Patrons Cataloging? The Role and Quality of Patron Tagging in Item Description

William Lund
billlund@icloud.com

Allyson Washburn
allyson_washburn@byu.edu

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>

 Part of the [Library and Information Science Commons](#), and the [Social Media Commons](#)

Original Publication Citation

Lund, Bill, Washburn, Allyson. Patrons Cataloging? The Role and Quality of Patron Tagging in Item Description, ACRL Fourteenth National Conference 29 Mar 12 - Mar 15 Seattle, WA 263-271.

Recommended Citation

Lund, W. B., and Washburn, A. (2009). Patrons Cataloging? The Role and Quality of Patron Tagging in Item Description. In *Proceedings of the Fourteenth National Conference of the Association of College and Research Libraries* (pp. 263-271). Seattle, Washington, USA.

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

Patrons Cataloging? The Role and Quality of Patron Tagging in Item Description

William Lund and Allyson Washburn

Problem

Consider the case of a searcher wanting to find a recently published mystery regarding a young man with an autism disorder. He searches on the terms “mystery” and “asperger” for Asperger Syndrome, an autism spectrum disorder. Such a book would be *The curious incident of the dog in the night-time*. The search does not retrieve the book in the traditional library online catalog because the Library of Congress Subject Heading (LCSH) is “autism”. (See Figure 1.) However, when the searcher refers to the social cataloging site LibraryThing (LT), he finds the book because a contributor tagged the book with the term “asperger.” This scenario is typical of many searchers who use current online public access catalogs (OPAC) of integrated library systems (ILS) that do not support patron tagging. Many searchers whether searching an online library catalog, article databases, Google or the internet, tend to use natural language keywords as opposed to Library of Congress subject headings (LCSH) or other controlled vocabularies.¹ Consequently, a large number of searches return no records or records that are not what the user wants²⁻⁶ (Norgard et al. 1992).

Additionally, users construct queries based on their expectation of the terms to be found describing the document, rather than an unfamiliar controlled vocabulary⁷⁻¹⁰ (Yu and Young, 2004). As early as 1987, Frost¹¹ reported that “researchers found that, for a majority of users, the library’s source of controlled subject headings remains intellectually inaccessible.” From the perspective of one of the authors who provides both reference and instruction at the library, this is still largely true. The recent advent of tagging and folksonomies presents an opportunity to supplement library catalogs and improve user search results.

This paper will compare the user-created tags from the LibraryThing folksonomy with the assigned LC subject headings of the collection at the Harold B. Lee Library at Brigham Young University, a major academic research library with 3.7 million volumes. LibraryThing, found at <http://www.librarything.com>, is a social networking site for cataloging books and currently contains records of close to 4 million books tagged by over 574,000 contributors. The controversy surrounding the practice of community contributions to catalog records comes from the traditional view that

William Lund is Assistant University Librarian for Information Technology at Harold B. Lee Library, Brigham Young University, e-mail: bill_lund@byu.edu; Allyson Washburn is eLearning and User Assessment Librarian at Harold B. Lee Library, Brigham Young University, e-mail: allyson_washburn@byu.edu

Figure 1. Catalog Record Showing Library of Congress Subject Headings

Lee Library OPAC Record	LibraryThing Record										
<p>The curious incident of the dog in the night-time Haddon, Mark.</p> <p>Personal Author: Haddon, Mark.</p> <p>Title: The curious incident of the dog in the night-time / Mark Haddon.</p> <p>Edition statement: 1st ed.</p> <p>Publication info: New York : Doubleday, 2003.</p> <p>Physical description: 226 p. : ill. ; 22 cm.</p> <p>General Note: Despite his overwhelming fear of interacting with people, Christopher, a mathematically-gifted, autistic fifteen-year-old boy, decides to investigate the murder of a neighbor's dog and uncovers secret information about his mother.</p> <p>Subject terms: Autism--Fiction.</p> <p>Subject terms: Savants (Savant syndrome)--Fiction.</p> <p>Geographic terms: England--Fiction.</p> <p>LCCN: 2002031355</p> <p>ISBN: 0365509456</p>	<p>The Curious Incident of the Dog in the Night-Time (Vintage Contemporaries) by Mark Haddon</p> <table border="1"> <thead> <tr> <th>Members</th> <th>Reviews</th> <th>Popularity</th> <th>Average rating</th> <th>Conversations</th> </tr> </thead> <tbody> <tr> <td>16,223</td> <td>411</td> <td>18</td> <td>3.9 (3.91)</td> <td>318</td> </tr> </tbody> </table> <p>Your library</p> <p><input checked="" type="checkbox"/> The Curious Incident of the Dog in the Night-Time (Vintage Contemporaries) by Mark Haddon, Vintage (2004), Paperback</p> <p>Members: 16,223</p> <p>Recently added by: han-hahling, moekang, blizulets, chend0, etharw, HesterPrynne, soppet, Josephberg, trendee37, emfarley</p> <p>Private watch list: odnuckey, JFB</p> <p>Your top 50 similar libraries: wdwidowis, Karys0133, bitter_suho, JFB</p> <p>Member tags: 1000 2004 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818 2819 2820 2821 2822 2823 2824 2825 2826 2827 2828 2829 2830 2831 2832 2833 2834 2835 2836 2837 2838 2839 2840 2841 2842 2843 2844 2845 2846 2847 2848 2849 2850 2851 2852 2853 2854 2855 2856 2857 2858 2859 2860 2861 2862 2863 2864 2865 2866 2867 2868 2869 2870 2871 2872 2873 2874 2875 2876 2877 2878 2879 2880 2881 2882 2883 2884 2885 2886 2887 2888 2889 2890 2891 2892 2893 2894 2895 2896 2897 2898 2899 2900 2901 2902 2903 2904 2905 2906 2907 2908 2909 2910 2911 2912 2913 2914 2915 2916 2917 2918 2919 2920 2921 2922 2923 2924 2925 2926 2927 2928 2929 2930 2931 2932 2933 2934 2935 2936 2937 2938 2939 2940 2941 2942 2943 2944 2945 2946 2947 2948 2949 2950 2951 2952 2953 2954 2955 2956 2957 2958 2959 2960 2961 2962 2963 2964 2965 2966 2967 2968 2969 2970 2971 2972 2973 2974 2975 2976 2977 2978 2979 2980 2981 2982 2983 2984 2985 2986 2987 2988 2989 2990 2991 2992 2993 2994 2995 2996 2997 2998 2999 3000 3001 3002 3003 3004 3005 3006 3007 3008 3009 3010 3011 3012 3013 3014 3015 3016 3017 3018 3019 3020 3021 3022 3023 3024 3025 3026 3027 3028 3029 3030 3031 3032 3033 3034 3035 3036 3037 3038 3039 3040 3041 3042 3043 3044 3045 3046 3047 3048 3049 3050 3051 3052 3053 3054 3055 3056 3057 3058 3059 3060 3061 3062 3063 3064 3065 3066 3067 3068 3069 3070 3071 3072 3073 3074 3075 3076 3077 3078 3079 3080 3081 3082 3083 3084 3085 3086 3087 3088 3089 3090 3091 3092 3093 3094 3095 3096 3097 3098 3099 3100 3101 3102 3103 3104 3105 3106 3107 3108 3109 3110 3111 3112 3113 3114 3115 3116 3117 3118 3119 3120 3121 3122 3123 3124 3125 3126 3127 3128 3129 3130 3131 3132 3133 3134 3135 3136 3137 3138 3139 3140 3141 3142 3143 3144 3145 3146 3147 3148 3149 3150 3151 3152 3153 3154 3155 3156 3157 3158 3159 3160 3161 3162 3163 3164 3165 3166 3167 3168 3169 3170 3171 3172 3173 3174 3175 3176 3177 3178 3179 3180 3181 3182 3183 3184 3185 3186 3187 3188 3189 3190 3191 3192 3193 3194 3195 3196 3197 3198 3199 3200 3201 3202 3203 3204 3205 3206 3207 3208 3209 3210 3211 3212 3213 3214 3215 3216 3217 3218 3219 3220 3221 3222 3223 3224 3225 3226 3227 3228 3229 3230 3231 3232 3233 3234 3235 3236 3237 3238 3239 3240 3241 3242 3243 3244 3245 3246 3247 3248 3249 3250 3251 3252 3253 3254 3255 3256 3257 3258 3259 3260 3261 3262 3263 3264 3265 3266 3267 3268 3269 3270 3271 3272 3273 3274 3275 3276 3277 3278 3279 3280 3281 3282 3283 3284 3285 3286 3287 3288 3289 3290 3291 3292 3293 3294 3295 3296 3297 3298 3299 3300 3301 3302 3303 3304 3305 3306 3307 3308 3309 3310 3311 3312 3313 3314 3315 3316 3317 3318 3319 3320 3321 3322 3323 3324 3325 3326 3327 3328 3329 3330 3331 3332 3333 3334 3335 3336 3337 3338 3339 3340 3341 3342 3343 3344 3345 3346 3347 3348 3349 3350 3351 3352 3353 3354 3355 3356 3357 3358 3359 3360 3361 3362 3363 3364 3365 3366 3367 3368 3369 3370 3371 3372 3373 3374 3375 3376 3377 3378 3379 3380 3381 3382 3383 3384 3385 3386 3387 3388 3389 3390 3391 3392 3393 3394 3395 3396 3397 3398 3399 3400 3401 3402 3403 3404 3405 3406 3407 3408 3409 3410 3411 3412 3413 3414 3415 3416 3417 3418 3419 3420 3421 3422 3423 3424 3425 3426 3427 3428 3429 3430 3431 3432 3433 3434 3435 3436 3437 3438 3439 3440 3441 3442 3443 3444 3445 3446 3447 3448 3449 3450 3451 3452 3453 3454 3455 3456 3457 3458 3459 3460 3461 3462 3463 3464 3465 3466 3467 3468 3469 3470 3471 3472 3473 3474 3475 3476 3477 3478 3479 3480 3481 3482 3483 3484 3485 3486 3487 3488 3489 3490 3491 3492 3493 3494 3495 3496 3497 3498 3499 3500 3501 3502 3503 3504 3505 3506 3507 3508 3509 3510 3511 3512 3513 3514 3515 3516 3517 3518 3519 3520 3521 3522 3523 3524 3525 3526 3527 3528 3529 3530 3531 3532 3533 3534 3535 3536 3537 3538 3539 3540 3541 3542 3543 3544 3545 3546 3547 3548 3549 3550 3551 3552 3553 3554 3555 3556 3557 3558 3559 3560 3561 3562 3563 3564 3565 3566 3567 3568 3569 3570 3571 3572 3573 3574 3575 3576 3577 3578 3579 3580 3581 3582 3583 3584 3585 3586 3587 3588 3589 3590 3591 3592 3593 3594 3595 3596 3597 3598 3599 3600 3601 3602 3603 3604 3605 3606 3607 3608 3609 3610 3611 3612 3613 3614 3615 3616 3617 3618 3619 3620 3621 3622 3623 3624 3625 3626 3627 3628 3629 3630 3631 3632 3633 3634 3635 3636 3637 3638 3639 3640 3641 3642 3643 3644 3645 3646 3647 3648 3649 3650 3651 3652 3653 3654 3655 3656 3657 3658 3659 3660 3661 3662 3663 3664 3665 3666 3667 3668 3669 3670 3671 3672 3673 3674 3675 3676 3677 3678 3679 3680 3681 3682 3683 3684 3685 3686 3687 3688 3689 3690 3691 3692 3693 3694 3695 3696 3697 3698 3699 3700 3701 3702 3703 3704 3705 3706 3707 3708 3709 3710 3711 3712 3713 3714 3715 3716 3717 3718 3719 3720 3721 3722 3723 3724 3725 3726 3727 3728 3729 3730 3731 3732 3733 3734 3735 3736 3737 3738 3739 3740 3741 3742 3743 3744 3745 3746 3747 3748 3749 3750 3751 3752 3753 3754 3755 3756 3757 3758 3759 3760 3761 3762 3763 3764 3765 3766 3767 3768 3769 3770 3771 3772 3773 3774 3775 3776 3777 3778 3779 3780 3781 3782 3783 3784 3785 3786 3787 3788 3789 3790 3791 3792 3793 3794 3795 3796 3797 3798 3799 3800 3801 3802 3803 3804 3805 3806 3807 3808 3809 3810 3811 3812 3813 3814 3815 3816 3817 3818 3819 3820 3821 3822 3823 3824 3825 3826 3827 3828 3829 3830 3831 3832 3833 3834 3835 3836 3837 3838 3839 3840 3841 3842 3843 3844 3845 3846 3847 3848 3849 3850 3851 3852 3853 3854 3855 3856 3857 3858 3859 3860 3861 3862 3863 3864 3865 3866 3867 3868 3869 3870 3871 3872 3873 3874 3875 3876 3877 3878 3879 3880 3881 3882 3883 3884 3885 3886 3887 3888 3889 3890 3891 3892 3893 3894 3895 3896 3897 3898 3899 3900 3901 3902 3903 3904 3905 3906 3907 3908 3909 3910 3911 3912 3913 3914 3915 3916 3917 3918 3919 3920 3921 3922 3923 3924 3925 3926 3927 3928 3929 3930 3931 3932 3933 3934 3935 3936 3937 3938 3939 3940 3941 3942 3943 3944 3945 3946 3947 3948 3949 3950 3951 3952 3953 3954 3955 3956 3957 3958 3959 3960 3961 3962 3963 3964 3965 3966 3967 3968 3969 3970 3971 3972 3973 3974 3975 3976 3977 3978 3979 3980 3981 3982 3983 3984 3985 3986 3987 3988 3989 3990 3991 3992 3993 3994 3995 3996 3997 3998 3999 4000 4001 4002 4003 4004 4005 4006 4007 4008 4009 4010 4011 4012 4013 4014 4015 4016 4017 4018 4019 4020 4021 4022 4023 4024 4025 4026 4027 4028 4029 4030 4031 4032 4033 4034 4035 4036 4037 4038 4039 4040 4041 4042 4043 4044 4045 4046 4047 4048 4049 4050 4051 4052 4053 4054 4055 4056 4057 4058 4059 4060 4061 4062 4063 4064 4065 4066 4067 4068 4069 4070 4071 4072 4073 4074 4075 4076 4077 4078 4079 4080 4081 4082 4083 4084 4085 4086 4087 4088 4089 4090 4091 4092 4093 4094 4095 4096 4097 4098 4099 4100 4101 4102 4103 4104 4105 4106 4107 4108 4109 4110 4111 4112 4113 4114 4115 4116 4117 4118 4119 4120 4121 4122 4123 4124 4125 4126 4127 4128 4129 4130 4131 4132 4133 4134 4135 4136 4137 4138 4139 4140 4141 4142 4143 4144 4145 4146 4147 4148 4149 4150 4151 4152 4153 4154 4155 4156 4157 4158 4159 4160 4161 4162 4163 4164 4165 4166 4167 4168 4169 4170 4171 4172 4173 4174 4175 4176 4177 4178 4179 4180 4181 4182 4183 4184 4185 4186 4187 4188 4189 4190 4191 4192 4193 4194 4195 4196 4197 4198 4199 4200 4201 4202 4203 4204 4205 4206 4207 4208 4209 4210 4211 4212 4213 4214 4215 4216 4217 4218 4219 4220 4221 4222 4223 4224 4225 4226 4227 4228 4229 4230 4231 4232 4233 4234 4235 4236 4237 4238 4239 4240 4241 4242 4243 4244 4245 4246 4247 4248 4249 4250 4251 4252 4253 4254 4255 4256 4257 4258 4259 4260 4261 4262 4263 4264 4265 4266 4267 4268 4269 4270 4271 4272 4273 4274 4275 4276 4277 4278 4279 4280 4281 4282 4283 4284 4285 4286 4287 4288 4289 4290 4291 4292 4293 4294 4295 4296 4297 4298 4299 4300 4301 4302 4303 4304 4305 4306 4307 4308 4</p>	Members	Reviews	Popularity	Average rating	Conversations	16,223	411	18	3.9 (3.91)	318
Members	Reviews	Popularity	Average rating	Conversations							
16,223	411	18	3.9 (3.91)	318							

For example, both systems use the tags “autism,” “fiction,” and “England,” however, only the library catalog record uses the term “Savants.” Likewise, the LibraryThing tags include additional terms such as “divorce,” “asperger,” and “mathematics,” not found in the Lee Library catalog entry, but which relate to the work and could be argued provide a broader description. Lastly, note that LibraryThing includes tags apparently used by the individual user to indicate information not relevant to the work itself, for instance “read” and “unread.” This research does not compare non-descriptive tags such as those to LCSH.

Methodology

Comparing the descriptive metadata, such as LCSH or LibraryThing tags, between a collection using LCSH and a folksonomy in which there is no authority control required finding two collections where common materials could be compared and evaluated. The library OPAC provides a broad collection of 3.7 million volumes in which items are cataloged with LCSH assigned by professionally trained catalogers. On the folksonomy side, the developers of LibraryThing, a user-driven personal library site, provided a folksonomy of 3.9 million records. A folksonomy differs from the controlled vocabulary of the LCSH in that users provide descriptive terms based on their own understanding and vocabulary.

Linking the two collections required a common unambiguous point to match records. This turned out to be the ISBN, which is provided by both systems. Although there were over 3 million records in each collection, ultimately only about 433,000 matches were discovered based on the ISBN. The authors had hoped to be able to compare at least a million records. The smaller number of matches between the two systems was not ideal, but unavoidable. With some reflection this is not terribly surprising given the different nature of each collection. For example, the Lee Library’s collection extends back over 100 years, adding approximately 50,000 new titles each year. LibraryThing, based on personal collections, is likely to favor more recent titles. Looking at the books with the most references in LibraryThing shows that they tend toward trade books, while the Lee Library collection tends toward materials appropriate to undergraduate education and graduate research.

The records from each system were linked through the ISBN and compared using the LC subject head-

ings from the library catalog and the tags from the folksonomy. From the library catalog the comparisons used the MARC fields 650 (Topical Term), 651 (Geographic Name), and 655 (Index Term--Genre/Form). LibraryThing provided individual tags linked to the work ID. Table 2 displays the number of records, LCSH entries and LibraryThing tags used in this study. The LCSH terms were evaluated in two ways:

Records used in the study:	433,416
LCSH Entries (650:)	830,658
LCSH Geographic Entries (651):	134,571
LCSH Genre Entries (655):	75,519
Total LC Subject Headings:	1,040,748
Total LibraryThing tags:	18,783,751

as specified by the Library of Congress and split into individual keywords. For example, the LCSH term “Education--Political aspects--United States” was evaluated for matches in LibraryThing both as indicated above and as individual keywords “Education,” “Political aspects,” and “United States.” Also, to counteract some variance in the folksonomy, all terms from both systems were converted to lower case for comparisons. One of our original thoughts was that there would be very few LibraryThing tags which matched the form of LCSH using the double dash “--”. This turned out to be incorrect. We suspect there are librarians using LibraryThing for their own collections.

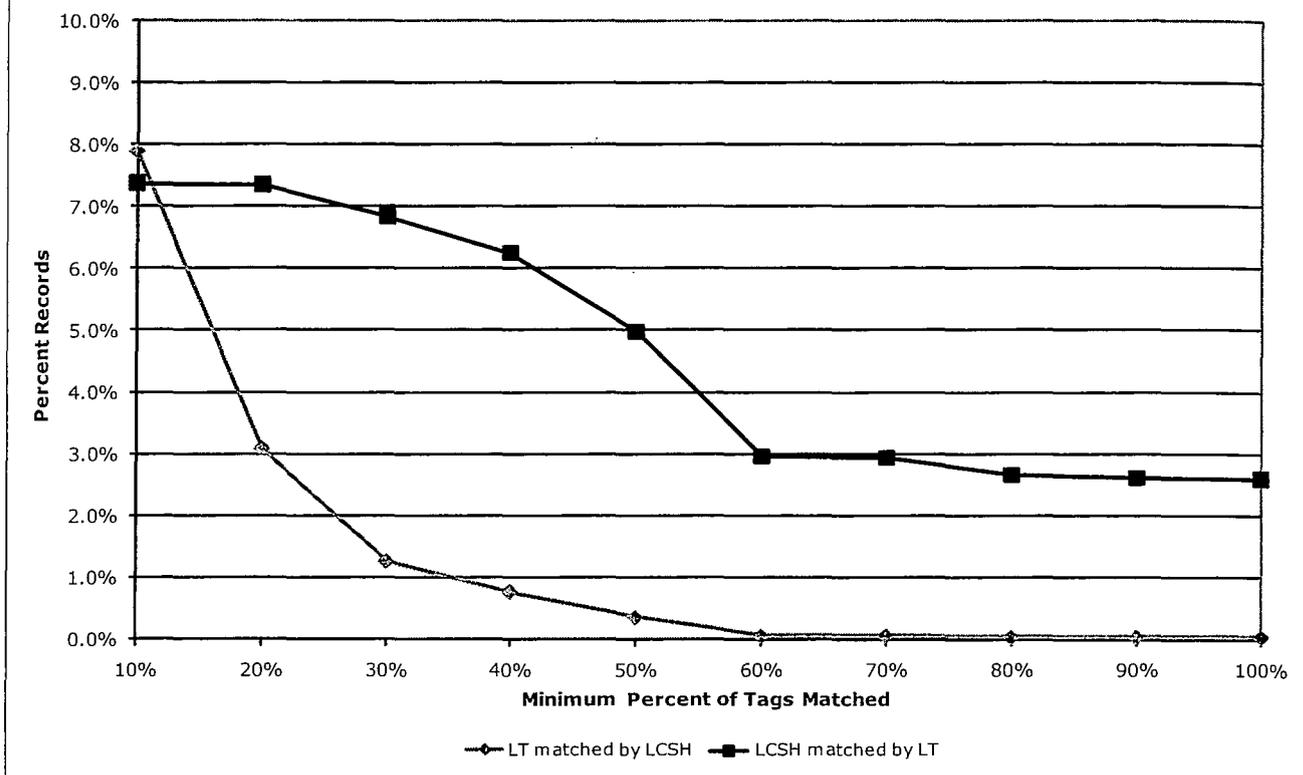
In order to facilitate query processing, the extracted records were housed in a MySQL database, from which SQL queries and Perl code could evaluate and compare the records.

Findings and Discussion

Exact Matches

The first and most obvious question concerns the matches of LCSH and the tags found in LibraryThing (LT). One would think that the number would be quite small, and that is true, but surprisingly there are 149 records, which exactly match between the Lee Library and LibraryThing. Figure 2 shows the minimum exact match of tags between the LCSH found in the MARC record and LT tags. Reading the graph, 7.9 percent of the LT records have at least 10 percent of their tags ex-

**Figure 2. Cumulative Matching
Exact Match Between LCSH and LibraryThing Tags**



actly matching those from the LCSH of the associated Lee Library MARC records, while 7.4 percent of the LCSH from the MARC records exactly match the tags from LT. Similarly, 3.0 percent of the LT records have at least 60% of their tags exactly matching subject headings from LCSH, while 0.1 percent of the Lee Library records have 60 percent of the subject headings exactly matched by LT tags. Finally, looking at 100 percent matching, 2.6 percent of the LCSH from the library catalog MARC records were an exact match with tags from LibraryThing. Inspecting those exact matches, it appears that they occur exclusively where the LC subject headings are simple (e.g. “acting,” “child development,” or “democracy”) without any subheadings.

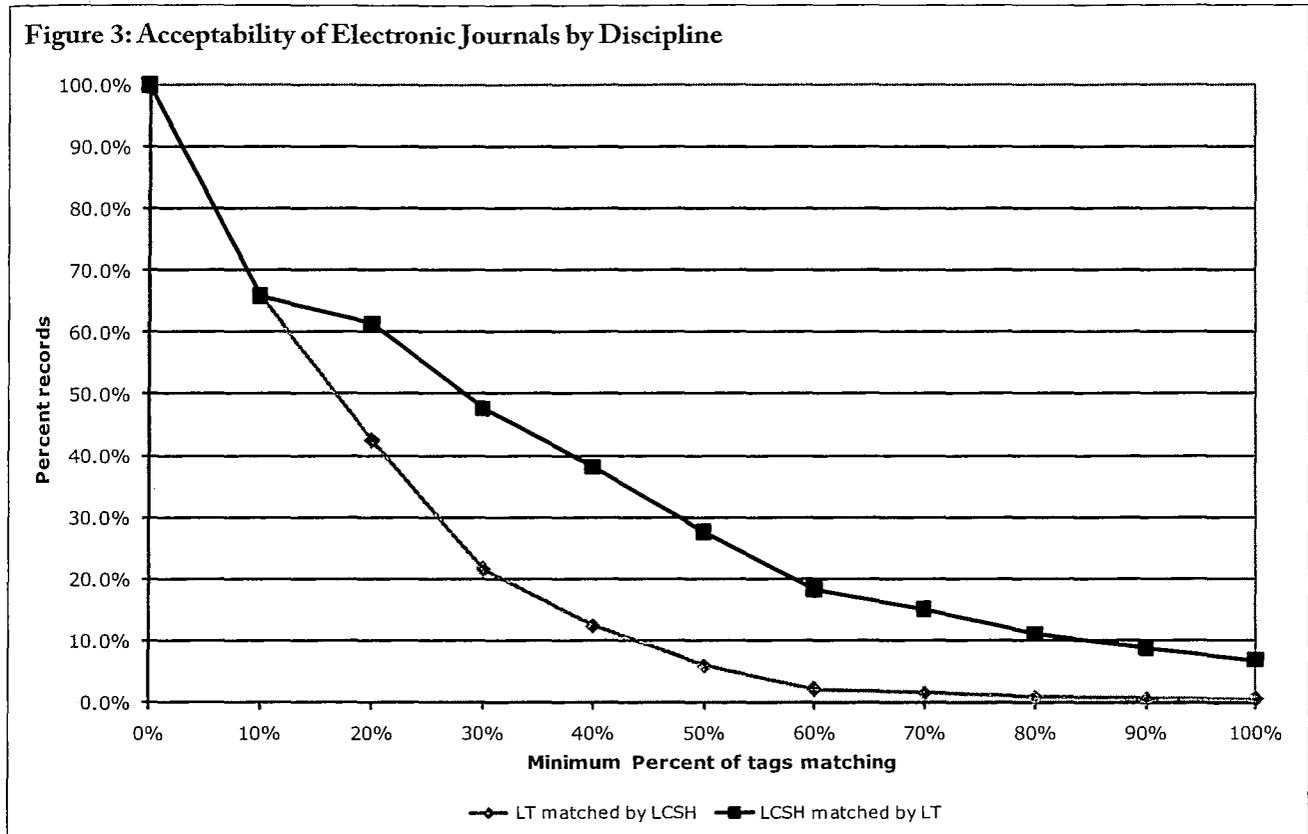
With the exception of the first data point at 10 percent, there were always more instances of LibraryThing tags matching the LCSH entries in the library catalog than the reverse. In any case, given that only exact matches between full Library of Congress subject headings and LibraryThing tags were considered, the total number of matches is still quite small. The vast majority of the records in both the library catalog and LibraryThing had no matches when con-

sidering only exact matches of the formal Library of Congress subject headings.

Keyword Matches

As indicated in Villen-Rueda¹⁵ most library searchers do not use LCSH, but select keywords instead. Like most integrated library systems, the one at the Lee Library will in fact take a keyword search and return results based on matching a portion of a subject heading. For instance a search on “autism” will return the work *The curious incident of the dog in the night-time* in which the LCSH “Autism—Fiction” occurs. An exact match with the keyword in the search was not necessary. In LibraryThing, that same work has a number of tags, one of which is “autism.”

Based on this, the next step in the study separated all subject headings into their component parts, i.e., splitting “Autism—Fiction” into two tags “autism” and “fiction” for evaluation purposes. This would appear to be closer to how the user and the system would interact. Figure 3 shows these results. Immediately, it is apparent that there is a better match between LCSH and the LT tags. For instance 65 percent of the re-



records of both the library catalog and LT have at least a 10 percent match and 27.7 percent of the catalog records have 50 percent match with the LibraryThing tags. Consider the following examples.

For the work *The imperfect panacea: American faith in education, 1865-1990* shown in Table 3 there are only two LC subject headings, all parts of which are matched by tags from LibraryThing. LT

TABLE 3
Comparing Lcsh and Librarything Metadata. The Imperfect Panacea : American Faith in Education, 1865-1990

Library Catalog Record		LibraryThing Record	
Library of Congress Subject Headings:	Evaluated as	LibraryThing User Supplied Tags	Evaluated as
Education--United States--History. Education--Philosophy.	philosophy history united states education	#edu 370.97320 american education American History education education in america education in the U.S. education in the united states non-fiction History history of education philosophy of education united states education	education nonfiction fiction in 370.97320 history america #edu states of american united the philosophy us

TABLE 4
Comparing LCSH and LibraryThing Metadata.
Integrative Health Promotion : Conceptual Bases for Nursing Practice

Library Catalog Record		LibraryThing Record	
Library of Congress Subject Headings	Evaluated as	LibraryThing User Supplied Tags	Evaluated as
Holistic nursing. Health promotion. Alternative medicine. Health Promotion--methods. Health Behavior. Holistic Health. Nursing Theory.	holistic methods theory medicine nursing alternative behavior health promotion	health promotion holistic	holistic promotion health

users have created tags which fully matched all of the individual components of the subject headings. Full matches occur in only 6.8 percent of the library catalog records.

Table 4 illustrates the converse, for the work *Integrative health promotion: Conceptual bases for nursing practice* where the LCSH completely match all of the LT tags. However, referring back to Figure 3 it can be seen that the subject headings from the catalog match the LT tags much less frequently, in only 0.6 percent of the records. One possible explanation for this in the case of the work in Table 4 is that there are only four LT users who have included this work in their collection.

This is the source of much of the observed difference between the library catalog and the LT records. Whereas the library catalog records, are for the most part static once they have been created, the LT records continue to grow as users add tags, which for them are meaningful descriptions of a work. The cataloging standard for creating a subject heading is that when a subtopic comprises 20 percent of a work¹⁶ it should be added as a heading. However, the LT tags are created based on user utility rather than a specific degree of topic coverage. An extreme example of this in LibraryThing is the work *Narrative of the life of Frederick Douglass* for which there are 504 distinct tags in LT and only two subject headings in the library catalog. This work is held by 1,873 users of LT, which is one explanation of the proliferation of tags.

In general, there are far more metadata entries and keywords in LibraryThing than LCSH in the library records. Figure 4 shows the number of works that have the given number of LCSH tags. For example, 21,932 works have five LC

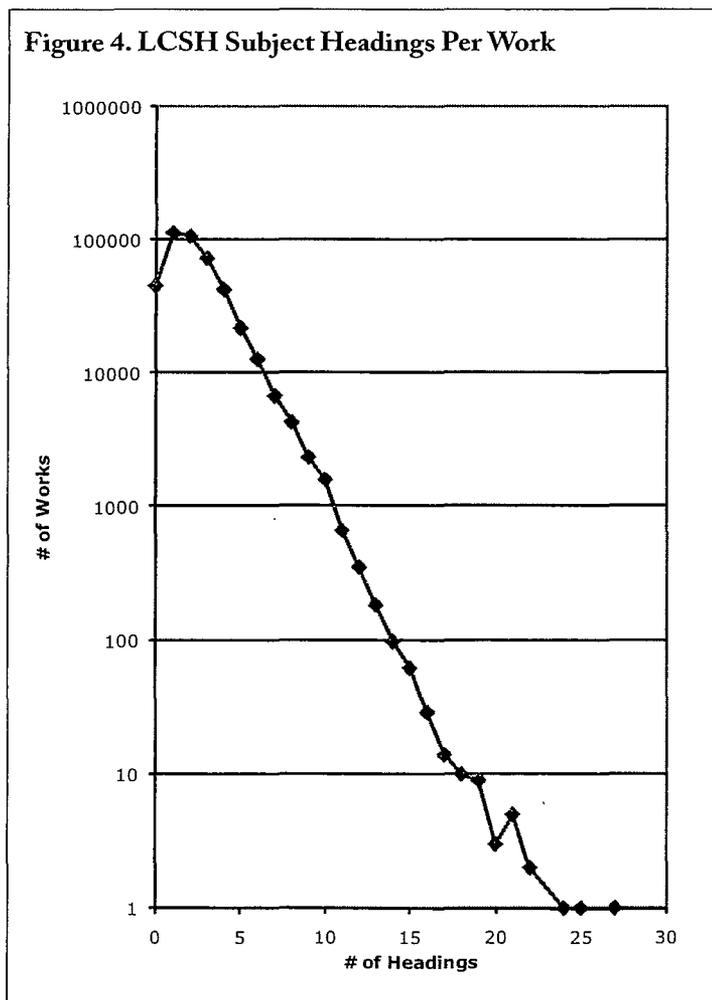
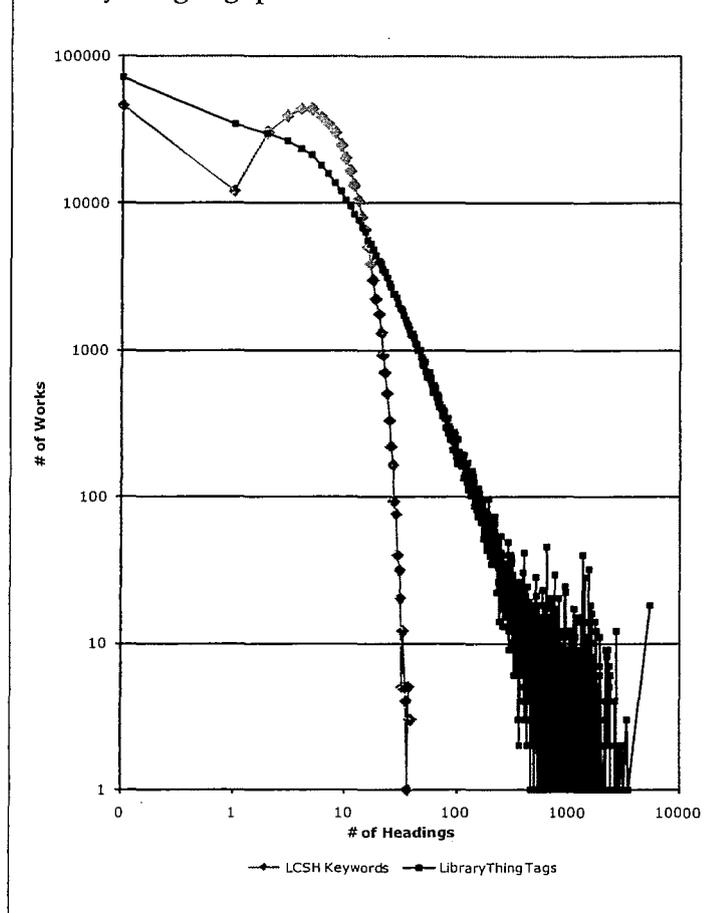


Figure 5: LCSH Keywords (Split Subject Headings) and LibraryThing Tags per Work



subject headings in the OPAC. In fact 93% of the works in this study had five or fewer LC subject headings.

Dividing the LCSH into keywords and comparing this to the number of tags found in LT, we see in Figure 5 that LT tends to have more tags than the same work has keywords derived from splitting apart the subject headings. Specifically, 10 percent of the works had more than 12 keywords (as derived from the LCSH of the MARC record) while 32 percent of the LibraryThing records had more than 12 keywords (or tags). As can be seen from the graph, the maximum number of LCSH derived keywords is 40, and the maximum number of LibraryThing keywords is 5,561.

Quality of Patron Contributions

The findings of this study with regard to the number of tags contributed by users were similar to those reported by Trant¹⁷ in a “proof of concept” study for the

Metropolitan Museum of Art in 2005. Thirty-nine non-professional participants, i.e. museum administrative staff and volunteers, viewed thirty works and assigned terms to the works. “88 percent of the terms supplied by participants . . . were not found in the basic museum descriptions” (p.97). Further, the Museum Subject Cataloging committee reviewed these terms and judged more than three-fourths of the terms as valid. Combined with the findings of this study, it appears that user supplied terms for the identical work, match or provide more, descriptive terms, thus increasing successful searching and access to collections.

A recent masters thesis in Computer Science at Brigham Young University titled *Improving Library Searches Using Word-Correlation Factors and Folksonomies*¹⁸ further explores this concept by creating an enhanced library catalog using both Library of Congress Subject Headings as well as tags to create a relevancy rated retrieval of results. The author states that “experimental results show that [the system] (i) significantly reduces the amount of queries that retrieve no results, (ii) obtains high precision in retrieving and accuracy in ranking relevant results, and (iii) achieves a processing time comparable to existing library catalog search engines.”¹⁹ Experimentally, the author of this thesis found that 16.2 percent of the queries to the OPAC returned no results

compared to 1.0 percent of the queries using the LibraryThing folksonomy. Further, 61 percent of the time the first returned result from the OPAC was deemed relevant by human reviewers compared to 84 percent of the time for the results using the enhanced library such with the LibraryThing folksonomy. Clearly these results are an improvement over the OPAC alone.

Conclusions

Graham²⁰ states “When we suspect that a significant proportion of users’ needs may not be adequately met, it is legitimate to consider alternatives to current cataloging practices and policies in order to serve catalog users better” Just as Graham had some success by adding user search terms as cross-references to catalog records, tagging can serve the same function of reducing the number of no-hits searches in OPACs in general. The results of these studies indicate that in general the LibraryThing folksonomy does a better

job of representing what is in the MARC record than the converse. There is value in considering the contributions of informal folksonomies to describe library materials, providing additional access points beyond the formal LC subject headings.

Further Research

Areas of interest for future research may include stemming the LT folksonomy to include variations of a term (e.g. plurals, alternate spellings, etc.) and the exclusion of idiosyncratic tags. To illustrate, many tags in LT are meaningless outside of the individual user's context. The tag "Box 1" obviously refers to the location of a work in a user's collection, but does nothing to describe a work. Likewise "read," "unread," provide only personal user information. Future research could explore a way to exclude tags, that are not descriptive in nature, perhaps based on frequency across works to indicate that they are meaningless as descriptive information.

Acknowledgments

The authors are grateful to the catalogers of the Lee Library and contributing libraries for the LC subject headings found in the MARC records, which were extracted for this work. We are also very grateful to Tim Spalding and Abby Blachy of LibraryThing and to the contributors to LibraryThing for the folksonomy provided for use in this study. Lastly, we acknowledge the contributions of the Library Information Technology division of the Lee Library for their assistance.

References

1. Patricia M. Wallace, "How Do Patrons Search the Online Catalog When No One's Looking? Transaction Log Analysis and Implications for Bibliographic Instruction and System Design". *RQ* 33, no.2 (1993): 239(14 p).
2. Rhonda N. Hunter, "Successes and Failures of Patrons Searching the Online Catalog at a Large Academic Library". *Reference and User Services Quarterly* 30 (1991): 395-402.
3. Ray R. Larson, "Classification, Clustering, Probabilistic Information Retrieval and the Online Catalog". *The Library Quarterly* 61, no.2 (1991): 133-173.
4. Barbara Norgard, Michael G. Berger, Michael K. Buckland and Christian Plaunt. 1993. The Online Catalog: From Technical Services to Access Service. In *Advances in Librarianship*, 111-148. New York: Academic Press.
5. Deborah D. Blecic, Nirmala S. Bangalore, Josephine L. Dorsch, Cynthia L. Henderson, Melissa H. Keonig, and Ann C. Weller, "Using Transaction Log Analysis to Improve OPAC Retrieval Results". *College & Research Libraries* 59 (January 1998): 39-50.
6. William Lund, "Unicorn Search Results from August 2006 to December 2006". Unpublished internal report of Harold B. Lee Library, Brigham Young University, Provo, UT.
7. Allyson Carlyle, "Matching LCSH and User Vocabulary in the Library Catalog". *Cataloging and Classification Quarterly* 10, no. 1/2 (1989): 37-63.
8. Holly Yu and Margo Young, "The Impact of Web Search Engines on Subject Searching in OPAC". *Information Technology and Libraries*, 23, no.4(2004): 168-180.
9. Karen Markey, *Subject Searching in Library Catalogs Before and After the Introduction of Online Catalogs*. Dublin, OH: OCLC Online Computer Library Center, Inc., 1984.
10. Karen M. Drabenstoff and Diane Vizine-Goetz, *Using Subject Headings for Online Retrieval: Theory, Practice, and Potential*. San Diego, CA: Academic Press, 1994.
11. Carolyn O. Frost, "Subject Searching in an Online Catalog". *Information Technology and Libraries* 6 (March 1987): 60-63.
12. Adam Mathes, "Folksonomies - Cooperative Classification and Communication Through Shared Metadata," 2004. 1-13. Retrieved from: <http://adammathes.com/academic/computer-mediated-communication/folksonomies.html>
13. Elaine Peterson, "Parallel Systems: The Co-Existence of Subject Cataloging and Folksonomy". *Library Philosophy and Practice*, April 2008: 5 pages.
14. Ellen Kroski, posting to Infotangle blog, December 11, 2005, <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/>.
15. Luis Villen-Rueda, Jose A. Senso, and Felix de Moya-Anegon, "The Use of OPAC in Large Academic Library: A Transactional Log Analysis Study of Subject Searching". *The Journal of Academic Librarianship* 33 no. 3(2007): 327-337.
16. This standard is explained by the following reference: *General topic and subtopic; principle vs. specific case*. If a work discusses a general topic with emphasis on a particular subtopic, or presents a principle and illustrates the principle with a specific case or example, assign headings for both the general topic or principle and for the subtopic or specific case or example, provided that the treatment of the latter forms at least 20% of the work. *Example: Title: Revolutions yesterday and today*. [A survey of revo-

lutions with emphasis on the Cuban Revolution of 1959] 650 #0 \$a Revolutions \$x History.#0 \$a Cuba \$x History \$y Revolution, 1959. Retrieved from: <http://desktop.loc.gov/nxt/gateway.dll/Subject%20Cataloging%20Manuals/scmshm/3/10?f=templates&fn=document-frame-htm.htm&an=&c3.0&q=>

17. J. Trant, "Exploring the Potential for Social Tagging and Folksonomy in Art Museums: Proof of Concept". *New Review of Hypermedia and Multimedia* 12, no. 1 (2006): 83-105.

18. Sole Pera, "Improving Library Searches Using Word-Correlation Factors and Folksonomies." Master's thesis, Brigham Young University, 2008.

19. Ibid, 5

20. Rumi Y. Graham, "Subject No-Hits Searches in an Academic Library Online Catalog: An Exploration of Two Potential Ameliorations." *College & Research Libraries*, 65, no.1(2004): 36-54.