

Lala Hajibayova
Doctoral Candidate
Indiana University Bloomington
lhajibay@indiana.edu
812 327 5075

Blog, news, ecommerce: Does genre matter for taggers?

Abstract

This study presents an analysis of over 7000 tags assigned to 36 online resources. Analysis of the genre tags revealed that users' preferences in assigning tags emerged from the context of the resource. Study revealed significant variance in tagging vocabulary across four content categories and three resource genres. Based on Bakhtin's genre theory (1986), this study suggests that utilization of genre should be grounded on language, in particular, on its meaning it carries to a particular group of individuals.

Keywords

Tagging, user-generated tagging vocabulary, genre.

Introduction: Genre Studies

Genre has been investigated in several different domains, including rhetoric (e.g., Jamieson & Campbell, 1982), linguistics (e.g., Askehave & Swales, 2001; Bhatia, 1993; Schryer, 2002), anthropology (e.g., Hanks, 1987), media (e.g., Corner, 1991; Neale, 1995), psychology (e.g., Mandler, 1984), human-computer interaction (e.g., Vaughan & Dillon, 2006) and library and information science (LIS) (e.g., Crowston & Kwasnik, 2003; Montesi & Navarrete, 2008; Montesi, 2010; Rosso, 2005). Researchers in these domains have studied genre from uniquely disciplinary perspectives and have offered various definitions of genre. In rhetoric, for example, Jamieson and Campbell (1982) define a genre as a "dynamic fusion of substantive, stylistic, and situational elements and . . . constellations that are strategic responses to the demands of the situation and purposes of the rhetor" (p. 146). Based on Jamieson and Campbell's definition, Miller (1984) argues that genre refers to a "conventional category of discourse based in large-scale typification of rhetorical action; as action, it acquires meaning from situation and from the social context in which that situation arose" (p. 163). Thus, from Miller's perspective, genre is a "rhetorical means for mediating private intentions and social exigence; it motivates by connecting the private with the public, the singular with the recurrent" (p. 163); and learning about genre helps us to "understand better the situations in which we find ourselves and the potentials for failure and success in acting together" (p. 165). And Bauman, a linguistic anthropologist, suggests that each genre is defined by its "thematic or referential capacities, as a routinized vehicle for encoding and expressing particular orders of knowledge and experience" (2002, p. 85). He specifically points to the referential orientation of genres as part of the "indexical field implicated in relationships of generic intertextuality" (p. 85). For example, curing chants within healing rituals can also be performed in other contexts for entertainment, for the chanter's pleasure, or for instructional purposes to teach a novice healing practitioner (p. 85). In

this way, Bauman's notion of genres is closely aligned to Bakhtin and Medvedev's (1928/1985) conceptualization of genre as an "aggregate of the means for seeing and conceptualizing reality" (p. 137). For Bakhtin (1986) genres are kinds of practice that characterized by their "addressivity" (p. 95): Different genres correspond to different "conceptions of the addressee" and "determined by that area of human activity and everyday life to which the given utterance is related" (p. 95).

Each discipline has reflected its particular purpose of genre categorization, such as linguists employed genre to interpret specific features of language with particular types of writing; rhetoricians used genre to explain the social and cultural factors of language in use; psychologists examined the influence of text structure on reading comprehension and constructed text categorization according to the mode of cognitive processing required by different types of text (Rosso, 2005).

LIS genre studies draw heavily on Yates and Orlikowski's (1992) conceptualization of the notion of genre (e.g., Crowston & Kwasnik, 2003; Roussinov et al., 2001; Montesi & Navarrete, 2008), which posits that the "genre of organizational communication (e.g., recommendation letters or proposals) is a typified communicative action invoked in response to a recurrent situation" (1992, p. 301).

Researchers in LIS have suggested that genre may have implications for information seeking and retrieval (e.g., Crowston & Kwasnik, 2003; Freund, 2012; Montesi & Navarrete, 2008; Montesi, 2010; Rosso, 2005; Roussinov et al., 2001). Beghtol (2001) argued that genre analysis may be used to create a "framework of analysis for a domain" that can help "structure and interpret texts, events, ideas, decisions, explanations and every other human activity in that domain" (p. 19). Crowston and Kwasnik (2003) suggested that "identification and use of document genre as a facet of document and query representation" (p. 346) may improve performance and enhance resource representation in information retrieval systems. In contrast, Rosso's (2005) study of the implications of genre for web search and retrieval found that genre-annotated search results produced no significant improvement in a user's ability to make relevant judgments.

Studies of user-generated vocabularies have also suggested that users tend to employ genre related tags (e.g., Montesi, 2010; Munk & Mork, 2007). Munk and Mork's (2007b) quantitative analysis of 178,460 Delicious tags found that genre tags comprised 3.8% of the analyzed tags (p. 122), whereas Lamere's (2008) study of the 500 most frequently applied tags at Last.fm indicated that genre accounted for 68% of all tags assigned. Unlike traditional indexing systems, which generally assign a single genre, a resource in Last.fm can be "tagged with all potentially applicable" (p. 103) genre tags, thereby providing for better representation of the "fuzziness at the boundaries of [musical] genre categories" (p. 106) (see also Inskip, 2009).

This study aims to analyze the effect of the genre on tagging vocabulary generated by a potentially more diverse population of students and faculty at Indiana University in Bloomington (IUB). In particular, study examines whether tagging vocabulary varies across the three resource genres (news article, blog and ecommerce) and four content categories (TOOL, CLOTHING, FRUIT, VEHICLE).

Method

A quasi-experiment study was conducted to investigate whether the genre of a resource has an effect on tagging vocabulary. A group of 40 subjects (native speakers of English) were recruited from members of the faculty and the student body at Indiana University Bloomington (IUB). Each subject was asked to assign as many tags as he/she felt was appropriate to represent the 36 resources in order to retrieve them at some point in the future. Each resource represented one of three resource genres identified by Rosso (2005) (i.e., news article, blog, ecommerce) and one of four content categories used by Rosch and her associates (1976) (i.e., TOOL, CLOTHING, FRUIT, VEHICLE). This study used three of the genres identified by Rosso because these genres provide a distinctive representation of the “genre ecology of the Internet” (Herring, Scheidt, Bonus, & Wright, 2004, p. 7). The following definitions of news article, blog, and ecommerce were used to identify resources:

- News article – textual material, issued or distributed by someone other than the author, that reports on an event and includes supporting facts;
- Blog – a weblog entry that presents the personal opinion of the author;
- Ecommerce – an offering by an online retailer that describes or advertises a product that is for sale online.

The four content categories (TOOL, CLOTHING, FRUIT, VEHICLE) were adapted from Rosch et al. (1976) as representative of the “most common categories of concrete objects” in the English language (p. 387).

Selection of the set of 36 resources used in the experiment was based on an evaluation of how well each of the 90 preliminary resources represented both the category and the genre with which it had been identified. Evaluations were based on the ratings of a group of five judges who were recruited from the population of IUB faculty and doctoral students.

Genre tags were coded by two coders, inter-rater agreement between two raters was outstanding ($k=0.80, p<0.05$) (Landis & Koch, 1977) (see Appendix).

Results

Overall, 7617 tags were collected and analyzed according to the genre and the content category of the resources. As shown in Table 1, the percentage of tags assigned to the resources in the blog genre (36.3%) was slightly higher than of the news genre (33.9%), which was followed closely by ecommerce (29.8%). Analysis of variance, used to test variances in tagging vocabulary across three genres of resources, found statistically significant differences across tags assigned to the resources representing news, blog and ecommerce genres, $F(1, 39) = 193.94, p < .00$. There were also statistically significant differences in assignment of tags across four content categories and three genres: TOOL category, $F(1, 39) = 99.94, p < .00$, CLOTHING category, $F(1, 39) = 68.03, p < .00$, VEHICLE category, $F(1, 39) = 50.36, p < .00$, and FRUIT category, $F(1, 39) = 48.64, p < .00$.

Table 1
Tagging Vocabulary: Genres and Content Categories

Category/Genre	News Article	Blog	Ecommerce	Total
TOOL	647 (8.5%)	635 (8.3%)	612 (8%)	1894 (24.9%)
CLOTHING	658 (8.6%)	721 (9.5%)	610 (8%)	1989 (26.1%)
FRUIT	647 (8.5%)	735 (9.6%)	514 (6.7%)	1896 (24.9%)
VEHICLE	629 (8.3%)	672 (8.8%)	537 (7.1%)	1838 (24.1%)
Total	2581 (33.9%)	2763 (36.3%)	2273 (29.8%)	7617

Following, categories emerged from analysis of the tagging vocabulary: *content* related tags (78.3%), tags specifying the *genre* of a resource (12.1%), tags identifying the *source* of a resource (5.3%), tags specifying the *geographical location* (1.8%), *date* (1.4%) and *author* or individual (1.1%) responsible for the intellectual content of a resource (see Table 2).

Table 2
Tagging Vocabulary by Categories and Genres

Genre/Categories	News Article	Blog	Ecommerce	Total
Content	25.8%	28.4%	24.1%	78.3%
Genre	3.9%	4.6%	3.6%	12.1%
Source	2.3%	1.4%	1.7%	5.3%
Geographical Location	0.8%	0.9%	0.2%	1.8%
Date	0.6%	0.5%	0.3%	1.4%
Author	0.6%	0.5%	0%	1.1%
Total	33.9%	36.3%	29.8%	

A breakdown of the emergent categories by the genres of resources, revealed a slight prevalence of content tags assigned to the resources representing the blog (28.4%) genre, followed by news (25.8%) and ecommerce (24.1%) (see Figure 1). Specifically, content tags prevailed in the resources representing the blog genre of the FRUIT (8.14%) and CLOTHING categories (7.69%), followed by news (6.88%) and ecommerce (6.63%) genres of the TOOL category. A minimum number of the content tags were assigned to the resources representing the ecommerce genre of the FRUIT and VEHICLE categories. Genre tags were prevalent in the resources representing the blog genre (4.6%), followed by news (3.9%) and ecommerce (3.6%). Most of the genre tags, however, were assigned to the news resources representing CLOTHING category (1.5%), followed by blog resources representing VEHICLE category (1.4%), and ecommerce resources representing CLOTHING (1.1%) and FRUIT (1.1%) categories. Tags representing the source, author, and date of the resources were prevalent in the news resources – 2.3%, 0.6%, 0.6%, respectively.

Figure 1. Distribution of Categories of Tags

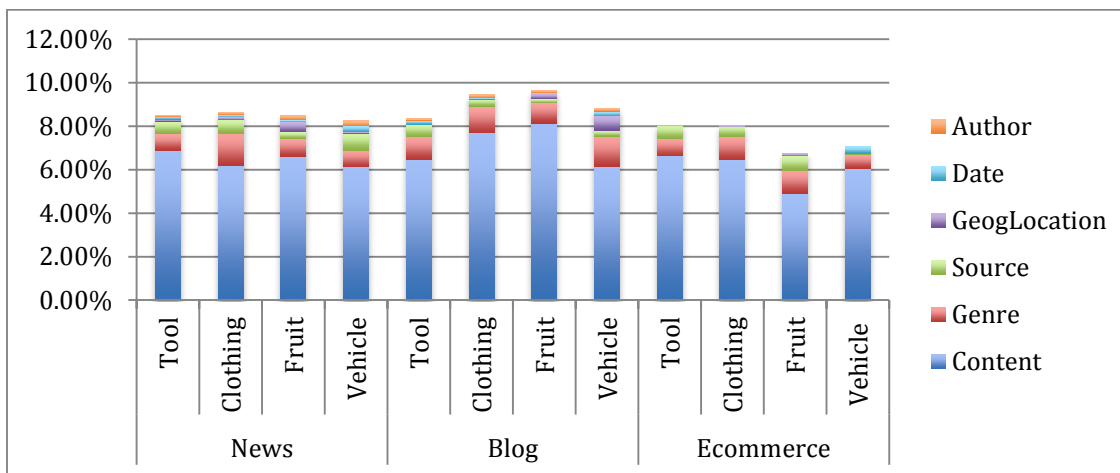


Figure 1. Representation of emergent categories across four content categories and three genres.

Analysis of a set of genres or genre repertoire (Yates & Orlikowski, 1992) across four content categories and three genres of the resources (see Table 3), showed that subjects not only recognized the predefined genre of the resources, but also assigned genre tags that emerged from the content of the resources tagged. For example, resources representing the news genre were also tagged as *review*, *press-release*, *guide*, *list*, *commentary*, *op/ed*, *critique* and *facts*. Genre tags assigned to the blog resources also specified the traditional resource format, such as *look book*, *photo album* and *catalogue*. Ecommerce resources were also tagged as *homepage*, *manual* and *synopsis*. Overall, assigned genre tags represented all three categories of the web genre identified by Crowston and Williams (2000): reproduced (such as *new article*, *press-release*, *op/ed*) adapted (such as *blog*), and emergent (such as *ecommerce*, *home page*).

Table 3.
Genre Repertoire

<i>Genre/Category</i>	<i>News Article</i>	<i>Blog</i>	<i>Ecommerce</i>
TOOL	Guide Ecommerce List News article Press-release Recommendations Review	Blog List Press-release Ranking Review	List Review Manual Ecommerce
CLOTHING	About	Blog	About

	Commentary Critic Editorial History List News article Op/Ed Opinion Review	Catalogue Ecommerce Guide Link List Look book Recommendations Review Website	Ecommerce List Reviews
VEHICLE	Column List News article Op/Ed Review	Blog Ecommerce Humor News article Photographs Review Story	Advertisement Ecommerce Homepage List Overview Review Synopsis
FRUIT	About Editorial Ecommerce Facts History News article Overview	About Blog Ecommerce History Guide News article Review	Ecommerce List

Assignment of tags representing genres of the resources, however, was statistically significant only for the resources representing FRUIT category and news genre, $F(1, 39) = 3.78, p < .005$, and for the resources representing VEHICLE category and blog genre, $F(1, 39) = 5.84, p < .02$.

Discussion

Observed differences in tagging vocabulary across four content categories and three genres of the resources, suggest that content and genre of the resources may have an effect on the user-generated tagging vocabularies. Although study revealed a small level of significance of tags representing genres of the resources, in line with previous studies (e.g., Lamere, 2008; Inskip, 2009) genre tags provided complex and rich representation of the news, blog, and ecommerce resources. For example, tags assigned to the online news article about the history of the citrus fruits ranged from *news article* to *column*, *fact*, *history*. Tags assigned to the resources representing blog also revealed users' comprehensive understanding of the overlapping genres. For example, analysis of the genre tags assigned to the fashion blog revealed that subjects identified not only the actual genre of the resource (*blog*), but also assigned genre tags that emerged from the context of the blog entry: *review*, *look book*, *tips*, *catalogue*, *photographs* (*photo albums*). Although there were relatively fewer genre-tags assigned to ecommerce resources, no agreement on a single genre assignment was observed. For example, genre tags assigned to the ecommerce resources representing the TOOL category ranged from *advertisement*, *synopsis* to *review*, *shopping*.

Recognition of the predefined genres of resources suggests that mental representation of the text structure and design serve as cues that allow users to quickly identify, select and make use of the resources (Dillon, 1991; Rumelhart, 1984; Vaughan & Dillon, 2006). More importantly, variations in tagging vocabulary across content categories and genres as well as multifaceted representation of the resource genre suggest that utilization of genre is grounded on language, in particular, on its meaning carried to a particular group of individuals or “addressivity” (Bakhtin, 1986, p. 95). Since each genre corresponds to a particular conception of the “addressee”, it determines an area of human activity and everyday life to which the given utterance is related. For Bakhtin (1986), certain features of language (such as lexicological, semantic and syntactic) are “knit together” with the genre providing a framework that facilitates communication.

Conclusions

Even though this study is limited to the analysis of tagging vocabulary across four content categories and three genres of the resources, it provides evidence that genres are perceived and considered useful for the resource representation and organization of resources. Along with variations in tagging vocabulary across predefined genres, study also revealed multifaceted representation of the resource genres – no single genre representation was observed across four categories and three genres. This study suggests that taxonomic representation of genre does not reflect users perspectives on and understanding of the resource genre.

Genre is intertwined with the language or context that ultimately represents, produces, reproduces and modifies particular genres (Bakhtin, 1986). Based on Bakhtin’s (1986) notion of genre, this study suggests that genre should be defined within the content category or context that is shared among members of a given language group. This approach may potentially provide greater opportunity for “making meaning” with resources, and, eventually, build systems that “understand” an individual’s information ecology.

References

- Askehave, I., & Swales, J. (2001). Genre identification and communicative purpose: A problem and a possible solution. *Applied Linguistics*, 22(2), 195-212.
- Bakhtin, M. (1981). Discourse in the novel. In M. Holquist, M. (Ed.), *The dialogic Imagination: Four essays by M. M. Bakhtin* (pp. 259- 422). Austin, TX: University of Texas Press.
- Bakhtin, M. (1986). Toward a methodology for the human sciences. In C. Emerson & M. Holquist (Eds. & Trans.), *Speech genres and other late essays* (pp.159- 172). Austin, TX: University of Texas Press.
- Bakhtin, M., & Medvedev, P. (1985). *The Formal Method in Literary Scholarship: A Critical Introduction to Sociological Poetics*. (A. J. Wehrle, Trans.). Cambridge: Harvard University Press (Original work published 1928).
- Bauman, R. (2000). Genre. *Journal of Linguistic Anthropology*, 9 (1-2), 84-87.

- Beghtol, C. (2001). The concept of genre and its characteristics. *Bulletin of the American Society for Information Science and Technology*, 27(2), 1-5.
- Bhatia, V. (1993). *Analysing genre: Language use in professional settings*. London: Longman.
- Corner, J. (1991). Meaning, genre and context: the problematics of 'public knowledge' in the new audience studies. In J. Curran & M. Gurevitch (Eds.), *Mass Media and Society* (pp. 267-284). London, England: Edward Arnold.
- Crowston, K., & Kwasnik, B. (2003). Can document-genre metadata improve information access to large digital collections? *Library Trends* 52(2), 345-361.
- Crowston, W. & Williams, M. (2000). Reproduced and emergent genres of communication on the World Wide Web. *The Information Society* 16, 201- 215.
- Dillon, A. (1991). Readers' models of text structures: The case of academic articles. *International Journal of Man-machine Studies*, 35(6), 913-925.
- Freund, L. (in press) A cross-domain analysis of task and genre effects on perceptions of usefulness. *Information processing and management* (2012). doi: <http://dx.doi.org/10.1016/j.ipm.2012.08.007>
- Hanks, W. (1987). Discourse genres in a theory of practice. *American Ethnologist*, 14(4), 668-692.
- Herring, S., Scheidt, L., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. *Proceedings of the 37th Hawaii International Conference on System Sciences, USA*, 1-11.
- Inskip, C. (2009). Tagging – A new frontier? *Library and Information Update*, 3, 27.
- Jamieson, K.H., & Campbell, K.K. (1982). Rhetorical hybrids: Fusions of generic elements. *Quarterly Journal of Speech*, 68, 146-157.
- Klahr, D., Chase, W. G., & Lovelace, E. A. (1983). Structure and process in alphabetic retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(3), 462-477. doi:10.1037/0278-7393.9.3.462
- Lamere, P. (2008). Social tagging and music information retrieval. *Journal of New Music Research*, 37(2), 101-114.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159-74.
- Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech* 70, 151-167.

- Montesi, M. (2010). Genre analysis of bookmarked webpages. *Online Information Review*, 34(6), 954-971.
- Montesi, M., & Navarrete, T. (2008). Classifying web genres in context: A case study documenting the web genres used by a software engineer. *Information Processing and Management*, 44(2008), 1410-1430.
- Munk, T., & Mork, K. (2007). Folksonomy, the power law & significance of the least effort. *Knowledge Organization*, 34(1), 16-33.
- Munk, T., & Mork, K. (2007b). Folksonomies, tagging communities, and tagging strategies- An empirical study. *Knowledge Organization*, 34(3), 115-127.
- Neale, S. (1995). Questions of genre'. In O. Boyd-Barrett & C. Newbold (Eds.), *Approaches to Media: A Reader* (pp. 460-472). London: Arnold.
- Orlikowski, W., & Yates, J. (1994). Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly* 39, 541-574.
- Rosch, E., Mervis, C, Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382-439.
- Rosso, M. (2005). *Using genre to improve web search* (Doctoral dissertation). Retrieved from http://ils.unc.edu/~rossm/Rosso_dissertation.pdf
- Roussinov, D., Crowston, K., Nilan, M, Kwasnik, B., Cai, J., & Liu, X. (2001). Genre based navigation on the Web. *Proceedings of the 34th Annual Hawaii Conference on System Sciences, USA*, 1-10.
- Rumelhart, D. (1975). Notes on a schema for stories. In D.G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 211-236). New York, NY: Academic Press.
- Schryer, C. (2002). Genre and power: a chronotopic analysis. In R. Coe, L. Lingard, & T. Teslenko (Eds.), *The Rhetoric and Ideology of Genre* (pp. 73-102). London, England: Taylor & Francis.
- Van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Vaughan, M., & Dillon, A. (2006). Why structure and genre matter for users of digital information: A longitudinal experiment with readers of a web-based newspaper. *International Journal of Human-Computer Studies*, 64(6), 502-526.
- Yates, J., & Orlikowski, W. (1992). Genres of organizational communication: A structurational approach to studying communication and media. *The Academy of Management Review*, 17(2), 299-326.

Appendix. Genre Coding.

Genre Coding	Genre Tags
About	about, about mandarin oranges, about clothing, about oranges
Advertisement	Acura advertisement, ad, clothes for toddlers ad
Blog	blog, CNY farm blog, Syracuse blog, Syracuse CNY Farms blog, LA Weekly blog, blog post, fruit blog, citrus blog, orange blog, news blog, fashion blog, style blog, women's fashion blog, fashion blogger, Biodwell blog, car blogs, blog review, AutoKeepUp blog, personal blog, blog spot, The Culpers blog, English Car blog, blogging, England blog, auto blog, blog spot, farm blog, blogging, Already Pretty blog. Women's fashion blog
Catalogue	
Column	column, Warren Brown column
Commentary	
Critique	fashion critic
Ecommerce	shopping, tool shopping, product page, purchasing information for drill, sales website, online shopping, clothes shopping, online store, store, retail, shop, pre-school sale, online ordering, order form, car shopping, vehicle pricing, pricing, pears for sale, order Asian pears online, food shopping, for sale, sales website, retail, buy fruit, store, order form, online gifts, to buy, purchase online, shipped to order, online organic clothing retailers, online retailer, clothing retailers, internet shopping, sale, order mangos online order, pears online
Editorial	fashion editorial, editorial about jeans
Facts	orange facts, fun facts

Guide	price guide, how-to guide, how-to, price guide for screwdrivers
History	history of jeans, history of denim, fashion history, history of objects, history of mandarin oranges, mandarin orange history, food history, history of oranges, food history, history of oranges
Homepage	company site, Acura homepage, Ford homepage, main page, zappos homepage
Humor	
Link	links to like products, link to online stores, organic clothing links
List	favorite tool list, gift list, wish list, company list, men's ecofriendly clothing list, top tool list, tool list
Lookbook	
Manual	manual, handi saw manual
News article	news, industry news, news website, news magazine, online news, article, online newspaper, fashion news, local news, Iowa newspaper, transportation news, weird news, MSNBC news, Trinidad Express newspaper, newspaper, ThomasNet news, product news, Mirror news, BBC News, San Francisco Chronicle
Op/Ed	
Opinion	opinion, opinion casual clothes, fashion opinion
Overview	overview, vehicle overview
Photographs	photo album, image, car photos
Press-release	
Ranking	
Recommendations	recommendations, purchasing advise,

	pozidrive tip, fashion tips, tools recommendations
Review	product review, clothing review, auto review, vehicle performance review, ACURA RDX review, car review, Warren Brown's car review, Volvo review, Chevy Malibu review, hammer and drill review, review of strawberry variety, power tool review, Wilton B.A.S.H sledgehammer review, fashion review, review of strawberry variety, milled-face hammer reviews, fruit review, handi saw reviews
Story	personal stories, the story of jeans
Synopsis	vehicle synopsis
Video	video, video pajama jeans
Website	good clothes website, zappos clothes website, company site, fresh pear website, online website