

The Broad Side Of Video Tagging – A Classification Of Tags From YouTube And Viddler

Overview

This research presents a classification study of tags collected from both Viddler and YouTube videos. The aim of the research was to discover if the language of the tags or tag type is different depending on whether the tags were entered into a broad (Viddler) or a narrow (YouTube) tagging system. In addition the research aims to analyse whether the tag type is affected by the category of the video.

Research Questions

1. Is there evidence of collaborative tagging on viddler?
2. Does collaborative tagging have an impact on tag type?
3. Does the category of video have an impact on tag type? Do people tag Entertainment videos differently to Informative videos?

Category Type	YouTube	Viddler
Entertainment	Comedy	Comedy
	Entertainment	Entertainment
	Gaming	Game/Games (10 from each)
	Music	Music
Informative	Technology	Technology
	Sport	Sports
	Travel	Travel
	News	News

Table 1 Video Categories

Methods

Eight YouTube categories were chosen and their respective tag found in Viddler. Viddler does not have pre-defined categories to choose from so the corresponding tag had to be used (see Table 1).

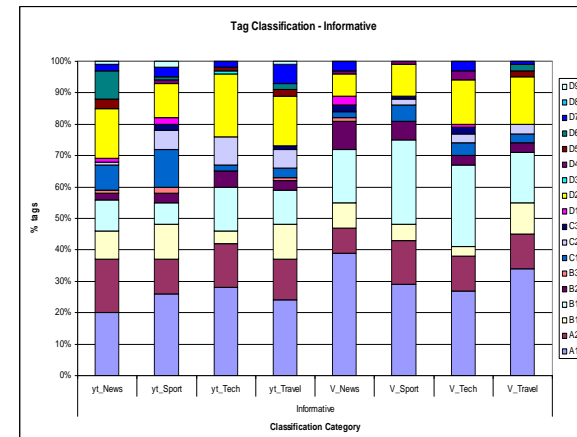
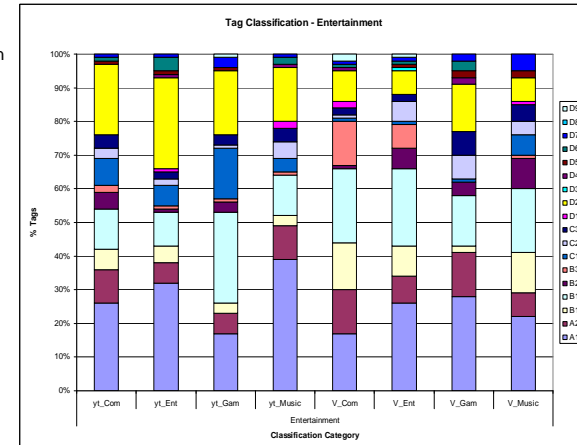
The YouTube API and Viddler API were used to capture a dataset of unique videos and associated textual data (title, user, views and tags). For YouTube, 300 of the 'most recent' videos in each category were retrieved each day over a 5 day period. For Viddler the feed was ordered by 'most recent', all videos were retrieved in each of the 8 categories on one day. The dataset was cleaned, removing videos that had been deleted from the system, or contained 100% non-English word/character tags. This left a YouTube dataset of 10,870 unique videos and a Viddler dataset of 46,573 unique videos.

Findings

- Entertainment has 68% more Opinion Expression (B3) tags than Informative. Viddler has 42% more B3 tags than YouTube.
- Informative has 28% more Conjunctions and Prepositions (D7) than Entertainment.
- Entertainment has 56% more tags that denote ownership (C3) than Informative.
- Entertainment has 18% more subjective (B2+B3) tags than Informative.
- Viddler has 34% more subjective (B2+B3) tags than YouTube.
- YouTube has 28% more Multiword tags (D2) than Viddler. Marginally (4%) more in Entertainment than Informative.
- Viddler has 14% more tags that denote ownership (C3) than YouTube.
- YouTube has 46% more Foreign Word tags (D6) than Viddler.
- Viddler has 24% more B1b tags than YouTube. Marginally more (4%) in Entertainment than Informative.
- Viddler has 30% more Interpretive B2 tags than YouTube and there are 7% more in Informative than Entertainment.
- YouTube has 44% more Refining (C1) tags than Viddler.

Conclusion

- There is no clear evidence that collaborative tagging is used on Viddler as it is not possible to find out who assigned the tag.
- The classification findings suggest that Viddler users tag with more specific vocabulary at a more subjective, rather than objective level. This may imply that some videos are collaboratively tagged.
- The study found that people are more likely to tag Entertainment videos subjectively, with Informative videos being tagged at a more basic level.
- Further work will analyse the data further to find any correlation between no. views and no. of tags, to find evidence of collaborative tagging. A statistical analysis of the classification results will be conducted to try and empirically answer the research questions.



Pre-Iconographic	A	Generic relationship between tag and video content	Objective
1	1	Tag generically identifies what video is 'of'	Basic
2	2	Tag identifies video Category/Genre	Basic
Iconographic	B	Specific relationship between tag and video content (semographic)	Objective
1(a)	1(a)	Tag specifically identifies what video is 'of' (place names/events)	Specific
1(b)	1(b)	Tag specifically identifies what video is 'of' (people/animals/objects)	Specific
2	2	Tag identifies what video is 'about'	Subjective
3	3	Tag identifies Opinion Expression	Subjective
Interpretive	C	Tag only useful to a minority of users, specific individual or group	Subjective
1	1	Refining tag	
2	2	Self-reference tag	
3	3	Tag which explicitly denotes ownership of video	
D	D	Irrelevant/Non Useful Tags	Subjective
1	1	Compound tag	
2	2	Multi Word Tags	
3	3	Attention Attracting Tags	
4	4	Misspelling	
5	5	Unable to determine relationship	
6	6	Foreign word/character	
7	7	Conjunctions and prepositions	
8	8	Repeated tags	
9	9	URL	

Table 2 - Classification Scheme

A random selection of 1 tag each from 100 videos were extracted from each category, from both datasets. A custom classification scheme was used to classify the tags for tag type and language parameters. Analysis of the classification will discuss the specificity of the vocabulary of the tags and whether this is affected by first the tagging system and second, the video's category.