
FINAL YEAR PROJECT

CM0343 – 40 Credits

INITIAL PLAN

“Label It”

A study analysing the relation between user demographic information and how they label an image.

Produced by Lowri Williams

Supervised by Irena Spasic, Moderated by Stuart M Allen

Project description

We are constantly tagging and labelling objects, whether it being over social media or items that we can touch. Labels (also known as tags) are used to specify explicitly what or who the content of that object is about. Tagging and labelling images from a demographic point of view is something I believe to be overlooked and has yet to be fully exposed.

This project will concentrate on a study in which the relation between user demographic information and how they label an image will be analysed.

The data will be collected from a simple web-based image labelling game called "Label It," built with basic web language tools such as HTML, CSS and PHP. The user is expected to enter five words they believe best describe the image content; throughout this study these will be referred to as labels - a form of metadata (i.e. data about data) that describes the content of the image or some other information relevant to the content. The user will also enter their information such as their gender, age and nationality, which along with the image labels will be stored in a relational database to facilitate subsequent processing. The database will be mined using a variety of statistical measures such as correlation, mutual information, etc. Labels will be explored using methods such as clustering, latent semantic analysis etc. to identify synonymy, homonymy and hyponymy or - in general - related labels. The results will be curated manually and organised into a taxonomy - defining and naming groups of objects that share the same characteristics. A formal definition similar to a dictionary entry will be given to each node in the taxonomy. The associated vocabulary will include all varieties in the jargon and lingo across different user groups.

Visualisation of the results from this study can be displayed in graph form and the labels within the taxonomy can be visually shown within tag clouds.

Project aims and objectives

Interim report content

- Research into databases, languages, hosting, graphing software and tag clouds - I intend on using the basic web language skills that I already possess from the Web Development module in the first year of my degree, which are HTML, CSS and basic PHP, although there are other web skills I would like to research into such as AJAX to build a responsive and interactive website. I am considering using PHPMyAdmin to store the data as its high performance, high reliability and ease of use works well alongside PHP. The university's Project Site web server is a convenient method for running the database and hosting the front-end of the website. Hosting the website requires some research as I am unsure if it is possible to host without having username and password restrictions so an alternative method such as free hosting may be better suited. I have experience in using Microsoft Excel's graphing functionality and so this software might potentially be what I use to graph the results, although I would like to research into areas where I don't have to simply copy and paste the data from the database into the spread sheet. I also know of websites that can produce tag clouds, but researching into how to build tag clouds using PHP is something I would like to look into. All research will definitely be included in my interim report.
- Design website and database(s) – To secure good data management and the ability to produce timely and correct data without redundancy or inconsistency, I will design the database to ultimately decide what data will be stored. Alongside this, to ensure that the website's usability is of a high standard and a blueprint to follow when building it, I will also produce a design of its potential functionalities. I expect users from all levels of IT literacy to use the website and so having a simple and smooth design should make usability easy. This could potentially be included in my interim report as a conclusion to the research, however a good base to being the final report from. I have yet to decide this and it all depends on how well I have followed my work plan and the structure of the interim report.
- Build prototype of website – To ensure excellent usability and a plan for the website's design, I hope to build a prototype using Microsoft PowerPoint as the hyperlink functionality will help with testing the flow and usability of the design before building it, although I intend to research on other methods and software that may be more appropriate for building prototypes. This research will also be included in the interim report along with the research I intend to complete during week four and five of my work plan. The prototype can be included as a conclusion or a part of the design.

Final report content

- Build website – screen shots of website can be included in the final report.
- Test website – Browser support is very important to test as many users are comfortable with using different browsers. As well as this, it is important to test if the HTML is valid and that the links are correct to ensure the smooth usability flow. I will run a series of tests in which I will ask some potential users to use the site before it is live. From this I hope to find errors or usability issues that I can fix before distributing the URL. Testing practices run on the website along with their results will be included in the final report.
- Distribute website - I will be using the social media sites Facebook and Twitter to distribute the URL to "Label It." Most of the target audience for this study use this form of communication and so not only do most of my friends and the people I know use Facebook, but their friends also use it. As conversations act like magnets; when one starts, people prick up their ears and take notice. This can have a knock on effect where more and more people jump into the conversation. This could help spread the URL wide and far and have many users. I will also use email for distribution although only to a selected few people who I am not connected with over social media. Possible screen shots of the distribution of the website can be added to the final report.
- Statistically analyse the occurrence of image labels in relation to different user characteristics.
- Normalise labels– using the image labels as well as their co-occurrence with similar labels to group them with different tags with the same meaning or different spelling according to the user's demographic information.
- Define - formal definition similar to a dictionary entry will be given to each normalised label as it is expected that homonymy will arise as users from different nationalities and cultures will use different vocabulary, jargon and lingo. I will use either an online or a physical dictionary for definitions.
- Folksonomy and taxonomy – sorting the image labels into categories derived from the consensus of the user and processing the data to form taxonomies.
- Graph outcomes – Graphing the results of a taxonomy against another will visually show the relations between how users label images. This will definitely be included in my final report as this will show the main results of the study.
- Tag clouds – Tag clouds are used to visualise free form text or tags on websites. I have used tag clouds before and find them a different and quirky way of displaying results as important or most frequent tags are shown in different text sizes or colours.

Work plan

I intend on achieving these milestones by dedicating a couple of extra days than expected to fully complete the task in hand as by the end of the second and the beginning of the third week of the autumn semester, I will start receiving coursework and so my concentration will be on completing them. I also plan to write a weekly log of my work so that it'll make it easier to remember issues that arose and those I potentially overcame, when I write up my interim and final report.

As this is will be an exploratory study, the discoveries from the data may lead to slight deviations from the original plan.

October '12						
Su	M	Tu	W	Th	F	Sa
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

November '12						
Su	M	Tu	W	Th	F	Sa
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

December '12						
Su	M	Tu	W	Th	F	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

January '13						
Su	M	Tu	W	Th	F	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

February '13						
Su	M	Tu	W	Th	F	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28		

March '13						
Su	M	Tu	W	Th	F	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

April '13						
Su	M	Tu	W	Th	F	Sa
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

May '13						
Su	M	Tu	W	Th	F	Sa
			1	2	3	

Date	Description
01/10/2012- 13/10/2012	Produce initial plan, submit for feedback from supervisor
14/10/2012- 18/10/2012	Work on initial plan feedback after return and submit.
19/10/2012	Submit initial plan
20/10/2012 - 03/11/2012	Research
04/11/2012- 10/11/2012	Design website and prototype
11/11/2012- 01/12/2012	Build website
02/12/2012- 13/12/2012	Produce interim report
14/12/2012	Submit interim report
15/12/2012- 06/01/2013	Christmas recess - continue to build website, test and distribute URL
14/01/2013- 25/01/2013	Examination period
28/01/2013- 01/03/2013	Database work (frequency of occurrence, folksonomy, taxonomy, normalisation, definitions)
02/03/2013- 22/-03/2013	Graph, analysis and tag cloud
23/03/2013- 14/04/2013	Easter recess - continue to finish database work. If finished, start producing final report
15/04/2013- 02/05/2013	Produce final report
03/05/2013	Submit final report