



INFORMATION VISUALIZATION DH2321

REPORT TO PROJECT 1

YIMING FAN
yimingf@kth.se

1 Introduction

This project aims at selecting best project groups based on anonymous survey data from students.

2 Selection Mechanism

1. Every group should have at least 2 (and no more than 3) ‘programmer’s in order to keep equal ability for working.
2. Group members should be chosen based on similar hobbies.

3 Workflow

3.1 Data Cleaning

3.1.1 Combine Similar Hobbies

Similar hobbies are combined into one word manually, for example ‘yoga, running, climbing...’ are combined into category ‘sports’. Below shows the original (left) and modified (right) attributes.

Alias	Interests	Alias	Interests
Bombadil	travel, sports, running, yoga, cycling, photography, coffee, gender equality, diversity, education, human rights	Bombadil	travel, sports, photography, coffee, politics
Frodo	graphics, friends, climb	Frodo	graphics, friends, sports
Bilbo	JavaScript, open source, design process, typography	Bilbo	programming, design
Sam		Sam	
Sauron	Sports, dancing, hiking, travelling	Sauron	sports, dancing
Boromir	Web Development, Competitive Programming, Data Science (Machine learning)	Boromir	programming
Galahadriel	programming, art, technology, basketball, equestrian sports, startups	Galahadriel	programming, art, sports
Legolas	developing front-end, web applications, new technologies, football, hunting & woodcrafting	Legolas	programming, sports
Pippin	Art, design, computer games, programming	Pippin	Art, design, programming
Lurtz	rock climbing, reading, drinking coffee	Lurtz	programming, reading, coffee
Celeborn	digital democracy, digital literacy, accessibility, usability, tv-shows, travelling, cooking	Celeborn	politics, tv-shows, travelling, cooking
Gollum	science, arts, technology, society, code, JavaScript, HTML, music, guitar, violin, cooking	Gollum	science, arts, politics, programming, music, cooking

3.1.2 Deletion of Insignificant Attributes

‘User’ and ‘graphics’ are deleted since only few of the results are different among others.

3.2 Recognition of Programmers

As shown above. First we divide quantitative attributes into two clusters. Then by comparing the sum value of those clusters we are done.

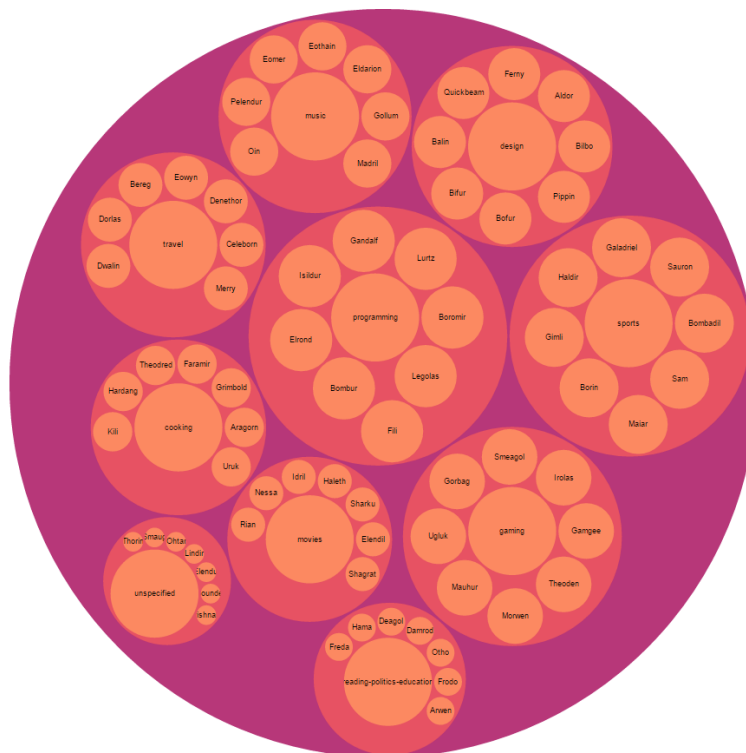
M	N	O	P	Q	R	S	T	U	V
is_prog	sum_desig	sum_prog	Statistics	Mathemati	Programm	IVIS	Art	HCI	UX
1	1	2	1	1	0	0	0	0	1
0	2	2	0	1	1	0	1	1	0
0	1	1	0	1	0	0	0	0	1
0	2	0	0	0	0	0	1	0	1
1	0	2	1	1	0	0	0	0	0
0	1	1	0	0	1	0	0	1	0
0	1	1	0	0	1	0	0	1	0
0	1	1	0	0	1	0	1	0	0
0	2	1	0	0	1	1	1	0	0
1	0	2	0	1	1	0	0	0	0
0	2	0	0	0	0	1	0	0	1
0	3	0	0	0	0	1	1	0	1
1	0	2	0	1	1	0	0	0	0

3.3 Group Selection

9 out of 10 groups are selected based on same hobbies while the other includes students harder to specify.

4 Result

As shown on the link: yimingf.github.io/dh2321-project-1/



5 Evidence

According to Shneiderman's best principle 'Overview first, zoom and filter, then detail-on-demand'[1], first we have an overview of 10 circles representing groups. By hovering on each small circle we can have detailed information including their common hobbies. Moreover bigger circles mean more common interests.

(Total: 191 words)

Reference

[1] Shneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualizations." In Visual Languages, 1996. Proceedings., IEEE Symposium on, pp. 336-343. IEEE, 1996.