

# Clustering of library users by similarity of visiting paths using location information

Noriko Sugie sugie@surugadai.ac.jp

Faculty of Media and Information Resources, Surugadai University, Japan

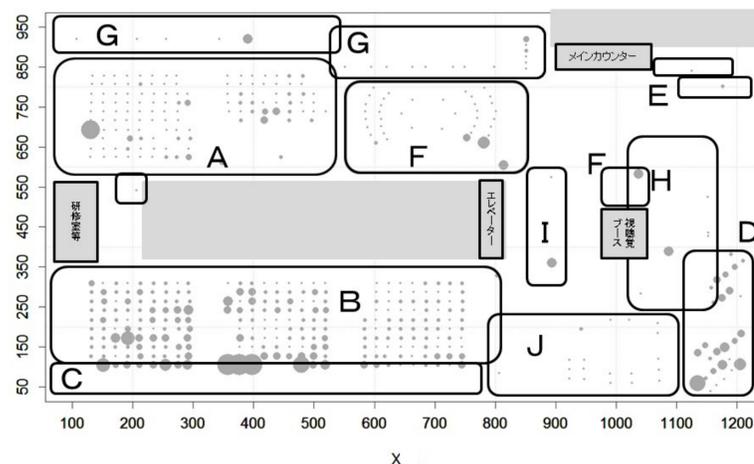
## 1. Statement of Problem

It is difficult to collect exact data on a user's actual searches using traditional methods such as observation and interview, and yet such data are central to understanding the user's information-seeking behavior and developing library services.

## 2. Purpose of Study

The purpose of this study is to classify users' behavior through analysis of location information acquired using observation methods adapting RFID technology.

## 3. Data collection



A: Research zone, B & C: General book zone, D: Paperback, E: Books returned, F: Information search zone, G: Reading chairs, H: Space for display, I: Library entrance, J: Magazine zone

Figure 1. Locations of the RFID tags and zones in the library

## 4. Data analysis

- The position coordinates of each user were converted into alphabetic characters referring to zones. The character string "FFFFGGA" indicates that the user sequentially visited the information search zone, reading chairs, and research zone.
- The edit distances were calculated from the character strings to express the degree of similarity among the visiting paths of users.
- Clustering the users' paths using Ward's method was conducted to identify user groups. The dendrogram was divided by a length of 12, yielding two clusters.
- The features of each group identified via clustering were analyzed with reference to the questionnaire responses.

- A behavioral investigation using radio frequency identification (RFID) and questionnaire survey were conducted at Chiyoda Public Library in Japan in 2012.
- There are 120,000 books and magazines on the floor, and each one has an RFID tag(953MHz).
- Participants were given an antenna and a personal digital assistant (PDA).
- The antenna received the radio waves from RFID tags, and the PDA recorded the data from the antenna while the participants were in the library, enabling their location to be identified.

## 5. Results

### (A) Frequency of the visited points

- The mean value for Cluster 1 was 8,453 with a median value of 7,237. The mean value for Cluster 2 was 5,853 with a median value of 3,840. The mean values differed significantly between the two groups, according to Welch's test ( $t(207) = 2.42, p < 0.05$ ).

### (B) Mean and percentage of visit frequency by zone

- Zones that exhibited significantly differing mean visit-frequency values between clusters were zone A ( $t(207) = -3.66, p < 0.01$ ), zone B ( $t(207) = 3.40, p < 0.01$ ), and zone C ( $t(207) = 5.12, p < 0.01$ ).

Table 1. Mean value of visit frequency by zone

Cluster	Zone	A	B	C	D	E	F	G	H	I	J
1	Mean	371	5,655	1,375	643	27	210	32	64	25	52
	SSD	1,022	4,571	2,987	1,463	166	372	127	113	52	163
2	Mean	2,433	1,564	102	900	13	563	130	76	29	44
	SSD	4,203	8,626	346	1,619	54	2,614	486	144	57	88

- The percentage of visiting frequency for A was highest in Cluster 2 (41.6%), and for B was highest in Cluster 1 (66.9%) ( $p < 0.01$ ).

### (C) Analysis of the questionnaire responses

- In Cluster 1, more users borrow and fewer sit compared with Cluster 2.
- In Cluster 2, more users don't borrow and more sit compared with Cluster 1.
- Testing for the difference in population proportions revealed significant differences between the two clusters for each option in both questionnaire items at a 1% level of significance.

Table 2. Cluster-wise borrowing behavior

Cluster	Borrowers		Non-borrowers		Sum	
	Num.	%	Num.	%	Num.	%
1	101	66.9	50	33.1	151	100.0
2	18	31.0	40	69.0	58	100.0

Table 3. Cluster-wise sitting behavior

Cluster	Sitters		Non-sitters		Sum	
	Num.	%	Num.	%	Num.	%
1	82	54.3	69	45.7	151	100.0
2	45	77.6	13	22.4	58	100.0

## 6. Conclusions

- Clustering based on the edit distance between users' character strings enabled identification of groups of users with different behaviors in the library.
- Users in Cluster 1 were likely to look for materials to borrow without sitting. They especially visited general book zone where materials for lending were located.
- Users in Cluster 2 were likely to look for materials and sit to read. They might visit the library to spent time reading materials or doing research.