

User Information Needs Through Query Reformulation

Asad Ullah
 Institute for Research in Applicable Computing
 University of Bedfordshire
 LU1 3JU
 UK
 Asadullah.asadullah1@study.beds.ac.uk

The main contribution of this paper is to identify user query types. Observing the query types, we can understand user behaviour in the context of information foraging. Using the essence of information foraging theory with our model we can also identify user information needs and their satisfaction through their query reformulation.

Keywords: Information foraging, User Behavior, Query reformulation

1. INTRODUCTION

A query is an entry point to a search engine so by examining the user queries we can understand a lot about the user behaviour. In this paper, we split the user queries into ten types. Four types of these queries were already defined by (Liu 2010), and six of them are our contribution. These queries types are validated from the empirical analysis of the user search log. In this paper, we will examine query reformulations and user behaviour in response to these reformulations. To fit the whole scenario in perspective, we need to use a model to justify our study. Previously we build a model EISE (Extended information goals, Search strategy and evaluation threshold) model (Asad 2016). Based on the ISE model (information goals, Search strategy and evaluation threshold) (Liu 2010) adapted from the information foraging theory (IFT) (Pirulli and card 1995)(Pirulli and card 1999). In this study, we will only focus on the information goals from our model to understand query reformulations. Search strategy and evaluation threshold will not be the subject of this paper.

2. BACKGROUND

Our study is not only depended on the empirical analysis of the data, but we also provide a theoretical background to the analysed data to fully understand the user behaviour in the context of human behaviour. Our EISE model (Asad 2016) is derived from established theories in information retrieval and psychology. We have a strong background to support our hypothesis. In this section, we will explain the immediate theories and models relevant to our study. Starting from Information foraging theory (IFT) which is the parent theory of our model. IFT is consists of three sections, Information scent, Information diet and Information

patch (Pirulli 1999). According to Information scent in the IFT users follows cues in the current information environment if they find positive cues in the environment user will stay for a long time otherwise it moves to another patch. Information diet in the IFT explains the user behaviour based on the user generalised and specific needs. Information patch is depended on the user approach whether the user wants to spend more time in one patch or want to move around between the patches to find relevant information. From the IFT theory, the ISE model (Liu 2010) is derived. The information goal of the ISE model derived information scent model and split into two types fixed information goals and evolving information goals. The Search strategy derived from information patch model and divided into two types cautious and risky. Evaluation threshold derived from information diet model and divided into two types weak and precise (Liu 2010).

Further, the ISE model is enhanced with psychology theories to produce the EISE model (Asad 2016). This is our contribution to the model to enhance it with relevant psychology theories. The information goal of the ISE model is improvised with the help of two mind set theory (Dweck 2006) to deeper understand the user behaviour. The information goals of the user will be the focus of this paper. The search strategy is upgraded with DiSC (dominance, influence, Steadiness, Conscientiousness) (DiSC 2015). The evaluation threshold involved decision making, the theory to understand decision-making behaviours is maximizers and satisfiers theory (Schwartz et al. 2002) to elevate the model. In this paper, we postulate that query reformulations are performed because of the user information goals to achieve. In this paper, our focus is on the information goals of the user. As we stated before that there are two types of information goals, fixed information

goals and evolving information goals. Now from the search log of the Bing users, we observed a pattern of query reformulations sync with these two types of information goals. According to a study query reformulations builds a session of user activity on the search engine (white and Drucker 2007) up to 32% of these sessions consist three or more queries (Jansen et al. 2005). Refer to another study (Huang and Efthimiadis) 28% of these sessions are the refinement to the previous query, and 52% of the users have the same behaviour. So this reformulation attract out intrest that the()

3. ANALYTICAL SETUP

Bing search log is used for this study to identify query types. Initially, ten users are selected for the preliminary data analysis. The search log for ten users contains a collection of 4231 queries average of 423 per user.

3.1 Information goals and Query Reformulations

The user performs query reformulations to satisfy his information needs. At the beginning of the session the user is not clear about his information needs, so his search will be exploratory learning from the retrieved information to reach the most relevant result (sloan 2015). In this case, the user will refine his query from an ambiguous query to a specific query to satisfy his information needs (song et al. 2009). In our model, this kind of behaviour is called user with evolving information goals. On the other hand, search process of the user with fixed information goal will be short and to the point. From our model (Asad 2016) we have operational definitions to differentiate between these two types of user's. The operational definitions are a predefined set of rules that how user's will perform in both cases. The operational definitions for fixed information goals explained with the help of table 1 and the operation definition for evolving information goals explained in table 2. Along with these operational definitions, we extracted ten types of query reformulations from the search log to help us in identifying the user with their information goals. The operational definitions are developed with the help of psychology theories. The existence of these operational definitions and query reformulations validates our hypothesis in the model shown in Table 1 and Table 2.

Table 1: Operational definitions for fixed information goals

Operational definitions	Descriptions
Less number of query iterations.	Query iterations are alterations to the query by user. (Asad and Liu 2016). e.g. Subset, Super-set,

Use small number of jumps.	Fixed jumps are the types of query jumps used when there are no changes to the information goals user. (Asad and Liu 2016).
Use small number of history.	History is an example of queries that are used in sessions and between the sessions. (Asad and Liu 2016)

Table 2: Operational definitions for evolving information goals

Operational definitions	Descriptions
large number of query iterations.	Query iterations are alterations to the query by user. (Asad and Liu 2016). e.g. Subset, Super-set,
Use large number of jumps.	Fixed jumps are the types of query jumps used when there are no changes to the information goals user. (Asad and Liu 2016).
Use large number of history.	History is an example of queries that are used in sessions and between the sessions. (Asad and Liu 2016)

During the session, the information goals of the user's changes, so the user refines his query according to his situation and understanding. In table 3 we have ten types of queries that users perform during their search process. These queries also help us to build up our operational definitions to distinguish between users.

Table 3: List of ten query types

Query type	Description
Repeat	Consecutive use of the same query. (Liu et al. 2010)
Subset	Subset of the previous query. (Liu et al. 2010)
Super-set:	The entire previous query with additional words. (Liu et al. 2010)
Overlap	Mix query with some words from previous query. (Liu et al. 2010)
Back	Same query used in a session but not consecutively. (Asad and Liu 2016)
Back Repeat	Repeat of same queries in between sessions. (Asad and Liu 2016)

Jump Query	New query within session and new information goal during the session. (Asad and Liu 2016)
Fixed Jump Query	New query with fixed information goals during the session (Asad and Liu 2016)

New jump	New query with new information goals between sessions. (Asad and Liu 2016)
Fixed New jump	New query with fixed information goals between sessions. (Asad and Liu 2016)

Table 4: List of quires types and user performance during the search process

	Repeat	subset	Super-set	overlap	Back	Back Repeat	Jump Query	Fixed Jump Query	New jump	Fixed New jump	New
User 1	19%	1%	3%	13%	1%	6%	18%	18%	5%	2%	
User 2	24%	1%	8%	9%	0%	4%	20%	12%	6%	1%	
User 3	23%	2%	3%	7%	1%	5%	24%	9%	7%	2%	
User 4	29%	2%	5%	17%	1%	5%	9%	11%	2%	1%	
User 5	24%	2%	6%	9%	1%	2%	24%	10%	1%	2%	
User 6	19%	1%	2%	7%	14%	8%	23%	10%	2%	1%	
User 7	13%	0%	0%	0%	38%	15%	4%	1%	0%	0%	
User 8	24%	1%	6%	13%	3%	10%	12%	11%	4%	2%	
User 9	23%	1%	6%	14%	2%	9%	17%	7%	3%	2%	
User 10	22%	1%	6%	19%	0	4%	19%	7%	3%	2%	

4. ANALYSIS

The information goals of the user be identified through these operational definitions and query reformulations performed by the users. In Table 5 we further clarify these operational definitions with the help of query reformulations that what are iterations, history and jumps. In Table 4 most of the users have repeat queries on average of 20%. Repeat is a use of history functionality, but we cannot only rely on only one reformulation to access a user behaviour. In combination with other operational definition, we conclude in Table 6 that which one the user fully satisfies the operational definitions with fixed information goals and evolving information goals.

Table 5: Categorizations of Queries

Iterations	History	Jumps
Subset	Repeat	Jump Query
Super-set	Back	New jump
overlap	Back Repeat	
	Fixed Jump Query	

	Fixed jump	New	
--	------------	-----	--

In Table 6 only user number seven falls in the category of Fixed information goals. If we look at Table 4, we will find out that only user seven consistently used back and repeat query during and between his sessions. It means that the user used the same query in his overall search process and the user is very fixed in his information goals. Although back and repeat are a history functionality to Table 5 one of our operational definitions also states that user with this kind of behaviour is evolving. But we should combine all our operational definitions together to achieve maximal results. In Table 4 we can see a complete summary of user's query reformulations and their percentage of the queries during their search. From the Table 4 above we can also conclude that most of the users are doing exploratory search and they have evolving information.

Table 6: User with Fixed IG and Evolving IG

Users	Fixed IG	Evolving IG
User 1		✓

User 2		✓
User 3		✓
User 4		✓
User 5		✓
User 6		✓
User 7	✓	
User 8		✓
User 9		✓
User 10		✓

5. CONCLUSION

The central theme of this paper was to identify deferent types of user with the help of query reformations. According to EISE model users are divided into two groups based on their information goals, fixed and evolving information goals. The ESIE model derived from a very stable theory in the field of information retrieval called information foraging theory. The model strengthened with psychology theories to understand the user behaviours. In this model, we have operational definitions to distinguish between the user information goals. The model as three stages but we only used the information goals section only to understand query reformulations and the initial interaction the user with the search engines.

6. REFERENCES

- A. Schwartz, B. Ward, J. Monterosso, K. Lyubomirsky, S. and White, and D. R. Lehman. Maximizing versus satisficing: Happiness is a matter of choice. (2002) *Journal of Personality and Social Psychology*, 83(5):1178–1197.
- Azure. DISC based personality assessment. <http://www.azureconsulting.com/files/1/74900324/DISCPastAndPresentAndWilliamMarston.pdf>, 2011. Accessed: 23 Dec. 2015.
- C. S. Dweck. (2006) *Mindset: The new psychology of success*. Random House, New York.
- H. Liu, P. Mulholland, D. Song, V. Uren, and S. Ruger. Applying information foraging theory to

understand user interaction with content-based image retrieval. In *Proceedings of the Third Symposium on Information Interaction in Context (IliX)*, New York, 2010, pages 135–144, ACM, USA.

Huang, J., & Efthimiadis, E. N. (2009). Analyzing and evaluating query reformulation strategies in web search logs. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09)* (pp. 77–86). ACM.

Jansen, B. J., Spink, A., & Pedersen, J. (2005). A temporal comparison of altavista web searching: Research articles. *Journal of the American Society for Information Science and Technology*, 56(6), 559–570.

Marc Sloan, Hui Yang, and Jun Wang. (2015). ewhfiuh A term-based methodology for query reformulation understanding. *Journal of Information Retrieval*. 18, 2, 145-165.

P. Pirolli and S. Card. Information foraging in information access environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages (CHI), Denver, Colorado, 1995, 51–58, ACM, USA.

P. Pirolli and S. K. Card. (1999) Information foraging. *Psychological Review*, 106(6):643–675.

Song, R., Luo, Z., Nie, J. Y., Yu, Y., & Hon, H. W. (2009). Identification of ambiguous queries in web search. *Information Processing and Management*, 45(2), 216–229.

ullah, Asad and Haiming Liu. "Theory-Based User Modeling For Personalized Interactive Information Retrieval". *Human Aspects In Adaptive And Personalised Interactive Environments In Conjunction With UMAP 2016*. Halifax: HAAPIE, Vol-1618, Canada.

White, R. W., & Drucker, S. M. (2007). Investigating behavioral variability in web search. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)* (pp. 21–30). ACM.