# Using Petri Nets to Capture Search Behavior Patterns in the Context of Query Reformulation

Vesna Gega[1] and Pece Mitrevski[2]

*Abstract* – **In recent years, discovering and understanding users' search behavior attract attention in the research community. Different approaches have been proposed for (i) learning and modeling how users search, and (ii) predicting future users' search behavior patterns, most of them based on statistical analysis and application of data mining techniques on a query log data. In this paper we focus on the application of Petri Nets on already discovered users' search behavior patterns, where consecutive actions for query reformulation in a single user session are considered as transitions.**

*Keywords* – **Petri Nets, modeling, search behavior, query log data, patterns.**

## I. Introduction

For several decades, the main goal of IR researchers is to refine or to develop novel techniques and methods for effective or efficient information retrieval [1]. But recently, the course has changed – analyzing and understanding users' behavior, learning how to interpret users' actions and making predictions of users' behavior patterns during the searching phase, attract IR researchers' attention [2], [3]. This means that a successful retrieval could be realized by integrating the knowledge of how users search, considering users' interaction context [4] and building a users' behavior model.

There are plenty of scientific papers that are focused on this problem. Despite this fact, we came across difficulties to identify a research work which considers users' search behavior (i.e. actions) as transitions. This is the main idea of our work, inspired by [5]: to show how Petri Nets, as transition based models, can be applied in modeling users' search behavior.

The rest of the paper is organized as follows. In Section 2 we present related work in this field. Section 3 represents an overview of our contribution. In Section 4 we give some conclusions and steps for future work.

## II. Related work

To build an effective users' search behavior model means to develop an accurate predictive mathematical model of the users' behavior. Usually, the models are built on analyzing query log data [6] formulated in a period of time [7]. Query log data are long and complex, but also an important source of information about users' behavior. There are two types of query log data: a) data obtained on a client side and b) data obtained on a server side.

In order to study server side log data and to gain knowledge of how users search, statistical analysis and application of data mining techniques need to be performed. Usually, the emphasis is placed on developing models for: discovering navigational patterns and explaining typical users' behavior on one hand [8], and on the other hand, predicting next user action [9] – both based on analyzing the click-through part (history) of the query log file [10].

The most exploited mathematical models so far are: first order Markov models, high-order Markov models and their extensions. For example, Baeza-Yates et al. [11] proposed three models. First, they built model for the number of clicks expected in a session where the number of queries formulated is known. Second, they calculated a Markov model of transitions in a query session. Third, they calculated a time distribution transition model considering times between query formulation and document selection. Their work is based on a query log data generated by a Chilean search engine called TodoCL (www.todocl.cl). Their results show that users formulate short queries; select few pages and an important proportion of them refine their originally proposed query in order to retrieve more relevant documents. Also, they illustrated that the query space is very sparse, which means that around 80% of the queries are formulated only once. Mixture of Hidden Markov Models for modeling users' behavior was postulated by Ypma et al. [12]. Their experiments were realized on a one-day log file from a commercial web site in the Netherlands in order to cluster the users based on their surfing patterns.

Probabilistic data mining approaches, such as Naïve Bayes, Bayesian Belief Network [13] and others [10], are also used in order to model and infer user's intention. The former developed a tool that automatically collect user's log data in IE environment, stored as an XML file. In their case, five types of user's actions were recorded: browse, click, query, save and close. Their results show that in the case of prediction, their proposed model is effective and it is really close to the human prediction.

Users' search behavior was also examined in the field of question answering tasks [2]. Here, the effect of topic familiarity and length of answers on users' search behavior was investigated. They showed that users give more accurate answers when they are more familiar with a topic. Contrary to this, with no topic familiarity, the ability for correct answer detection was low. Snippets as a retrievable unit were more preferable by the users than full context. But, the length of the answer was difficult to establish, because users prefer short as

[1]Vesna Gega is with the University for Information Science and Technology, Partizanska bb, Ohrid 6000, Republic of Macedonia, E-mail: vesna.gega@uist.edu.mk.

[2]Pece Mitrevski is with the Department of Computer Science and Engineering, Faculty of Technical Sciences, St. Clement Ohridski University, Ivo Lola Ribar bb, 7000 Bitola, Republic of Macedonia, E-mail: pece.mitrevski@uklo.edu.mk.

well as long answers. Also, different numbers of results presented were preferred by different users, depending on their knowledge level.

Modeling users' behavior can be useful and can increase search efficiency in the field of personalized search, for every user with respect to its type. This is mentioned in the work of Pitkow et al. [14].

One difficult but interesting issue is modeling users' changing behavior over time. Radinsky et al. [15] developed modeling framework adapted from physics and signal processing that can be used to predict future time varying users' behavior based on historical data and illustrate dynamic nature of search behavior. They are focused on modeling behaviors such as changes in query frequencies, clicked URLs and query-URL pairs over time, based on Bing query log data. The results of their experiments indicate that they achieved significant improvements in prediction compared to baseline models that weigh historical evidence the same for all queries.

The Petri Nets [16] as well developed graphical and mathematical models were also proposed to describe the behavior of users. Kantor et al. [5] have presented fairly poor theoretical or conceptual framework, which can describe different kinds of actions that may occur during searching considered as transitions. Starting from this point, finding it interesting and challenging to explore, inspired by this work we conduct a research and investigate whether Petri Nets as transition based models can be applied in modeling users' search behavior, in order to improve search quality.

## III. Using Petri Nets to Capture Search Behavior Patterns

Our study is based on the previous work [6], where a survey of many measures used to describe and evaluate the efficiency and effectiveness of large-scale search services is given. They covered research in six fields and presented rich visualizations on the: query space, users' query sessions, user behavior, operational requirements, the content space, and user demographics.

Their analyses are based on large and complex AOL[1] query log file. This file consists of ~20M web queries collected from ~650k users over three months. It includes the following columns:

1) AnonID - an anonymous user ID number.
2) Query - the query formulated by the user, case shifted with most punctuation removed.
3) Query Time - the time at which the query was submitted for search.
4) Item Rank - if the user clicked on a result, the rank of the item on which they clicked is listed.
5) ClickURL - if the user clicked on a search result item, the domain portion of the URL in the clicked result is listed.

Pass et al. [6] discovered and described the patterns of query reformulation as a part of the searching process within a single user session. In these patterns, several actions exist:

- A user formulates a new query;
- A user modifies (refines) a query:
  - add,
  - delete,
  - change word/s in a query;
- User returns to a previous query;
- User sees more results for the same query;
- User ends the session.

The dynamic nature of Petri Nets indicates that they can be accommodated for modeling such real users' search behavior. In this direction, we employ two models in order to capture search behavior patterns in the context of query reformulation. Each action is presented as a transition, as shown in Table 1.

TABLE 1
ALL POSSIBLE TRANSITIONS

| | |
|---|---|
| Formulate a new query | tFNQ |
| See results (first page) | tSR |
| Return to a previous query | tRPQ |
| See more results (next page) | tSMR |
| Modify a query | tMQ |
| Delete word/s in a query | tDWQ |
| Add word/s in a query | tAWQ |
| Change word/s in a query | tCWQ |
| End the session | tES |

The first model shown in Fig. 2 is related to simple users' behavior. According to the activity diagram in Fig. 1, first of all, the user formulates a new query. After seeing the results (usually the first page) the user can: end the search process, see more results (usually next page/s), return to a previous query, modify (refine) his/her query or ask a new query.

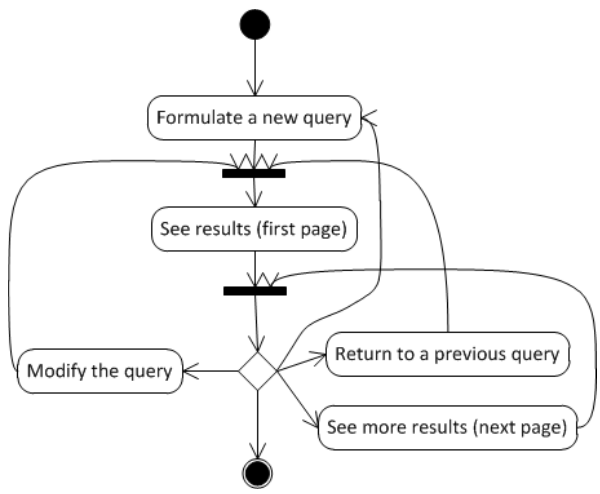


Fig. 1. Activity diagram for simple user behavior

In the Petri Net notation, initially, a token appears in the input place S and the user places a new search. When the transition tFNQ fires it means the user asks a new query, the token is removed from the place S and it is placed in the output place NQ. Next, the transition tSR fires, so the user sees the results for the query asked and the token now appears

[1] http://www.aol.com/

in the output place R. At this moment, five transitions are ready to fire: tES, tFNQ, tRPQ, tMQ, tSMR. Only one transition can fire at a time, so depending on the user's choice, the token is removed from place R and appears in output place: E if the user wants to end the session, R if the user wants to see more results for the same query, PQ if the user wants to see results for some previous query, MQ if the user wants to modify the query and finally NQ if the user wants to ask a new query. This is a recurring process, until the user ends the session.

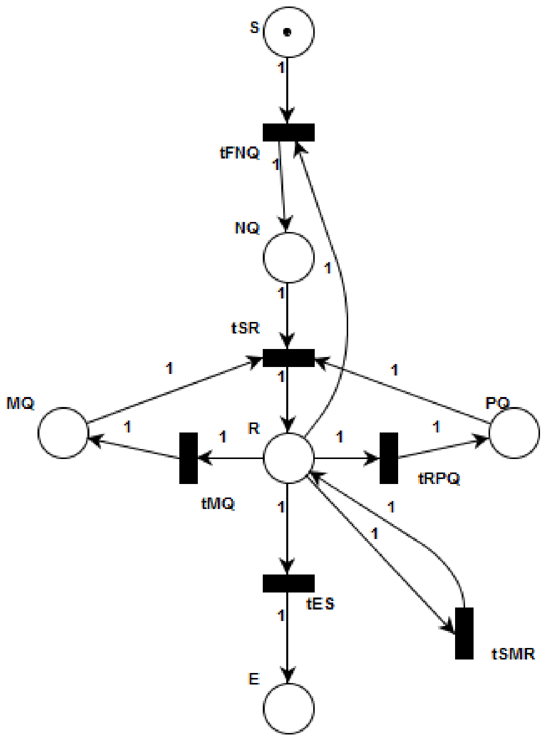


Fig. 2. Petri Net for simple users' behavior

The second model shown in Fig. 4 is related to extended users' behavior. According to the activity diagram in Fig. 3, the states are same as those in the simple model, except the modification state is replaced with three detected behavioral patterns: add word/s, delete word/s and change word/s to an existing query.

Once more, a token appears in the input place S and the user places a new search. When the transition tFNQ fires, the token is removed from the place S and it is placed in the output place NQ. Next, the transition tSR fires, so the user sees the results for the query asked and the token now appears in the output place R. At this moment, seven transitions are ready to fire: tES, tFNQ, tRPQ, tAWQ, tDWQ, tCWQ, tSMR. Only one transition can fire at a time, so depending on the user's choice, the token is removed from place R and appears in output place: E if the user wants to end the session, R if the user wants to see more results for the same query, PQ if the user wants to see results for some previous query, +Q if the user wants to add word/s to the query, -Q if the user wants to delete word/s from the query, CQ if the user wants to change word/s to the query and finally NQ if the user wants to ask a new query. Yet again, this is a recurring process.

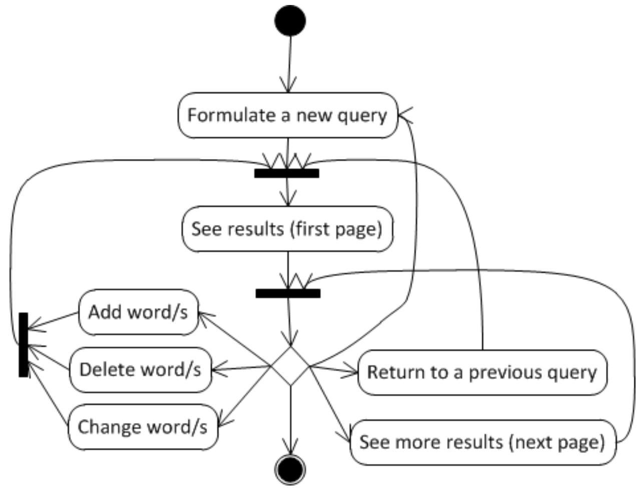


Fig. 3. Activity diagram for extended user behavior

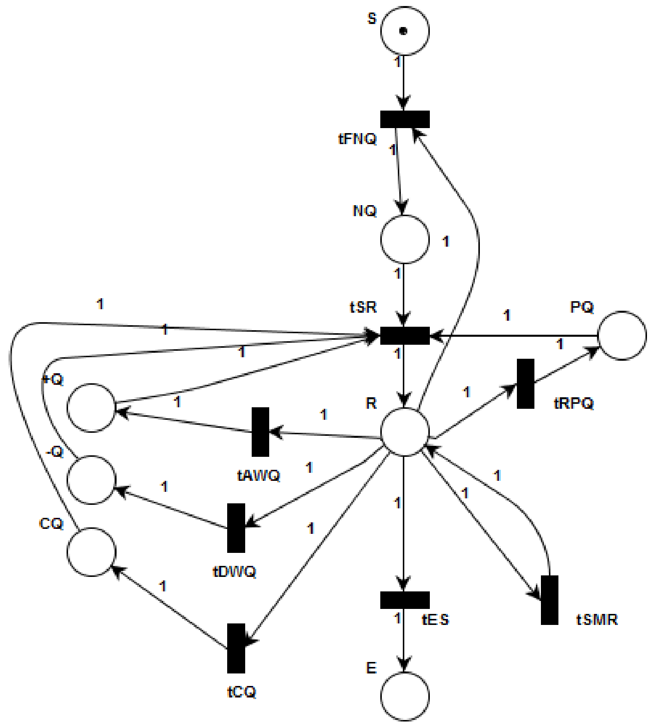


Fig. 4. Petri Net for extended users' behavior

According to the probability matrix built by Pass et al. [6] for average series of query formulations within a single user session, 42 transitions from a state to a state are possible, but the following most probable users' behavior patterns can be revealed:

1. When a token appears in the place NQ, it means that the transition tFNQ has already been fired and a new query is asked. Next, a token appears in the place R, which means the transition tSR is fired and user sees results for the new query. After this, the user starts a new search, the transition tFNQ is fired and a token is placed in the place NQ again.

$$\{NQ\} \xrightarrow{tSR} \{R\} \xrightarrow{tFNQ} \{NQ\}$$

2. When a token appears in the place -Q, it means that the transition tDWQ has already been fired, so some word/s from the query was/were deleted. Next, a token appears in the place R, which means the transition tSR is fired and user sees results for the modified query. After this, the user starts a new search, the transition tFNQ is fired and a token is placed in the place NQ.

$$\{-Q\} \xrightarrow{tSR} \{R\} \xrightarrow{tFNQ} \{NQ\}$$

3. When a token appears in the place CQ, it means that the transition tCWQ has already been fired, so some word/s from the query was/were changed. Next, a token appears in the place R, which means the transition tSR is fired and user sees results for the modified query. After this, the user changes the query, the transition tCQ is fired and a token is placed in the place CQ again.

$$\{CQ\} \xrightarrow{tSR} \{R\} \xrightarrow{tCQ} \{CQ\}$$

4. When a token appears in the place +Q, it means that the transition tAWQ has already been fired, so some word/s was/were added to the query. Next, a token appears in the place R, which means the transition tSR is fired and user sees results for the modified query. After this, the user changes the query, the transition tCQ is fired and a token is placed in the place CQ.

$$\{+Q\} \xrightarrow{tSR} \{R\} \xrightarrow{tCQ} \{CQ\}$$

5. When a token appears in the place R, it means that the transition tSR has already been fired, so the user sees results for a query. Next, a token appears in the place R again, which means the transition tSMR has already been fired and the user wants to see more results for the same query.

$$\{R\} \xrightarrow{tSMR} \{R\}$$

6. When a token appears in the place PQ, it means that the transition tPQ has already been fired, so the user returns to a previous query. Next, a token appears in the place R, which means the transition tSR is fired and user sees results for a previous query. After this, the transition tSMR is fired and a token appears in the place R again. It means that the user wants to see more results for the previous query.

$$\{PQ\} \xrightarrow{tSR} \{R\} \xrightarrow{tSMR} \{\mathrm{R}\}$$

## IV. Conclusions and Future Work

We have proposed a Petri Net modeling approach for describing existing users' search behavior in the context of query reformulation. In this work, we have presented all the identified actions in a query reformulation stage (as a part of the searching process within a single user session) as transitions.

Detailed experimentation and further investigation is scheduled in order to provide effective characterization and modeling of real users' search behavior with the help of the Petri Nets formalism. In order to evaluate several performance metrics, e.g. the number of clicks during the course of a query session, the distribution of time between query formulation and document selection, etc., we will employ the class of Generalized Stochastic Petri Nets (GSPN) [17], where immediate transitions have firing weights, timed transitions have exponentially distributed firing times, and the underlying stochastic process is a Continuous Time Markov Chain (CTMC). We hope that this work will motivate further research in this area.

## References

[1] V. Gega, P. Mitrevski, "On the General Principles of Human-Computer Information Retrieval", ICT Innovations 2012, Web Proceedings, pp. 371-380, Ohrid, Macedonia, 2012.

[2] A. Al-Maskari, M. Sanderson, "The Effects of Topic Familiarity on User Search Behavior in Question Answering Systems", LWA, vol. 1/2006 of Hildesheimer Informatik-Berichte, pp. 132-137, 2006.

[3] K. Järvelin, P. Ingwersen, "Information seeking research needs extension towards tasks and technology", Information Research, vol. 10, no. 1, 2004.

[4] I. Ruthven, "Information Retrieval in Context", Advanced Topics in Information Retrieval, The Information Retrieval Series, vol. 33, pp. 187-207, 2011.

[5] P. B. Kantor, R. Nordlie, "Models of the Behavior of People Searching the Internet A Petri Net Approach", Larry Woods (Ed.), Proceedings of the 62nd ASIS Annual Meeting, 1999.

[6] G. Pass, A. Chowdhury, C. Torgeson, "A picture of search", Proceeding InfoScale '06 Proceedings of the 1st international conference on Scalable information systems, Article no.1, 2006.

[7] C. Silverstein, H. Marais, M. Henzinger, M. Moricz, "Analysis of a very large web search engine query log", ACM New York, NY, USA, vol. 33, pp.6-12, 1999.

[8] R. W. White, S. M. Drucker, "Investigating Behavioral Variability in Web Search", Proceedings of the 16th international conference on World Wide Web, pp. 21-30, ACM New York, NY, USA, 2007.

[9] D. Downey, "Models of Searching and Browsing Languages Studies and Applications", Proceeding IJCAI (2007), pp. 2740-2747, 2007.

[10] B. Piwowarski, H. Zaragoza, "Predictive User Click Models Based on click-through history", Proceedings of the sixteenth ACM Conference on information and knowledge management, pp. 175-182, ACM New York, NY, USA, 2007.

[11] R. Baeza-Yates, C. Hurtado, M. Mendoza, G. Dupret, "Modelling User Search Behavior", LA-WEB '05: Proceedings of the Third Latin American Web Congress, 2005.

[12] A. Ypma, T. Heskes, "Automatic Categorization of Web Pages and User Clustering with Mixtures of Hidden Markov Models", WEBKDD, pp. 35-49, 2002.

[13] Z. Chen , F. Lin , H. Liu , Y. Liu , L. Wenyin , W. Ma, "User Intention Modeling in Web Applications Using Data Mining", Journal World Wide Web, vol. 5, pp. 181–191, Kluwer Academic Publishers Hingham, MA, USA, 2002.

[14] J. E. Pitkow, H. Schütze, T. A. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, T. M. Breuel, "Personalized search", Commun. ACM vol. 45, no. 9, pp. 50-55, 2002.

[15] K. Radinsky, K. M. Svore, S. T. Dumais, J. Teevan, A. Bocharov, E. Horvit, "Modeling and predicting behavioral dynamics on the web", WWW, pp. 599-608, 2012.

[16] T. Murata, "Petri Nets: Properties, Analysis and Applications", Proceedings of the IEEE, Vol.77, No.4, pp. 541-580, 1989.

[17] M. Ajmone Marsan, G. Conte, G. Balbo, "A Class of Generalized Stochastic Petri Nets for the Performance Evaluation of Multiprocessor Systems", ACM Transactions on Computer Systems, Vol. 2, No. 2, pp. 93-122, 1984.