The Curious Case of Session Identification

Florian Dietz[(✉)]

Berlin School of Library and Information Science, Humboldt-Universität zu Berlin,
Dorotheenstr. 26, 10117 Berlin, Germany
`florian.dietz@alumni.hu-berlin.de`

Abstract. Dividing interaction logs into meaningful segments has been a core problem in supporting users in search tasks for over 20 years. Research has brought up many different definitions: from simplistic mechanical sessions to complex search missions spanning multiple days. Having meaningful segments is essential for many tasks depending on context, yet many research projects over the last years still rely on early proposals. This position paper gives a quick overview of session identification development and questions the widespread use of the industry standard.

Keywords: Session identification · Task extraction · Interactive information retrieval

1 Introduction

Web usage mining has been around for quite some time now. Since the late 1990s and early 2000s, researchers have contributed dozens of studies about handling interaction logs and how to utilize them in their field of research. These early studies focus on search behaviour, interpreting how users interact with search systems and what is actually searched for [5,34]. Initial findings gave insight about average query length, amount of queries and reformulations or the number of visited result pages.

However, the actual identification of sessions in the interaction logs received a growing interest. Identifying patterns and segmenting logs into user sessions has grown to be a focal point, being the foundation for any further analysis or research [13]. Various methods were tested for finding reasonable session boundaries, often applying mechanical cuts like time outs. The most common inactivity time out of 30 min, most likely evolved from the 25.5 min proposed by [5], is still used today. Later, research interest went from mechanical sessions to a more intent-oriented approach, acknowledging that finding suitable user context is easier when sessions are logically segmented rather than mechanically. Therefore, definitions vary from mechanical [5] to logical [17].

Today, most related publications still apply the 30 min inactivity cut as a foundation. From user modelling to recommendation to personalisation - the 30 min rule seems to be omnipresent. This position paper is part of a dissertation project researching the impact of different session modelling concepts. A quick

timeline on the development of session concepts is presented and the solitary use of a temporal constraint discussed.

2   Literature Review

Session Identification. Early studies identifying sessions as the basic unit of measurement in interaction logs mostly relied on time gaps to decide if two consecutive queries belong to the same session, resulting in mechanically segmented sessions. [5] were among the first to introduce a temporal constraint. They report an average time of 9.3 min between interactions, adding 1.5 standard deviations to propose a temporal inactivity limit of 25.5 min. Other temporal cuts are also reported: 5 min [33], 15 min [14,15] or even 60 min and longer [3].

Over the years, these time constraints have evolved into a 30 min inactivity time out. Many works rely exclusively on this arbitrarily set time limit [4,8,21, 24,37], others recognized a need for more evidence, using stopping patterns [39] or dynamic time thresholds based on visited pages [7,41] and users [27]. After [35] reported multitasking during search sessions, even identifying interleaving intents, growing interest was directed to the identification of tasks rather than mechanical sessions.

Task Identification. Tasks may be similar to sessions, but they move away from purely mechanical thresholds to logical boundaries. Simple approaches use lexical similarity between adjacent queries [11] to identify topically related segments, assuming that queries that do not share any terms with previous ones indicate a new session [17] (although the sessions are identified with a temporal constraint in the first place). A prime example of the combination of lexical similarity and temporal relationship is [9], who use a geometric approach to calculate similarity between query pairs based on a 24 h temporal limit. Most approaches still use (mechanical) session-based features to calculate similarity between queries. Some use sequential patterns [28,30], others employ external sources to create a richer semantic context like thesauri [16] or pre-trained embeddings [10].

Even more advanced is the identification of cross-session tasks, recognizing the importance of interleaving and multiple tasks throughout the boundaries of mechanical sessions. [19] identified tasks as just another level of measurement. They define search sessions as user activity within a fixed time window, search goals as the atomic information need producing one or more queries and search missions as the overarching concept, connecting various search goals and therefore possibly spanning multiple sessions. This hierarchical point of view works well for describing user behaviour: visiting an information system in a session, searching for several goals belonging to one search mission. In [22], this concept is exploited via hierarchical clustering algorithms based on multiple query features. [12] and [13] propose a cascading method for connecting related adjacent queries by consecutively using lexical and semantic similarity, temporal proximity, search results and context comparison to find logically coherent search

missions. Other studies compare adjacent queries with binary classifiers [1,20], use latent structural Support Vector Machines [38] or utilize term and context embeddings [25,32].

3   Discussion

[40] qualitatively analysed real web sessions, identifying multiple factors as potential indicators for session boundaries: changing topics or tasks related to the topic, switching to a different phase of a mission, different environmental context (i.e. being among people) and the time gap as the traditional measure. Acknowledging the potential co-existence of these measures strongly supports a development from mechanical sessions to logically connected segments, possibly connecting multiple mechanical sessions and tasks. These concepts build upon each other and should be applied accordingly.

However, sessions identified with temporal boundaries are still widely used. 30 min of inactivity is the industry standard [2], despite clear indicators that solitary use of time gaps is not reliable [6,10,26]. Many applications using interaction logs still exclusively apply the 30 min inactivity time out rule as a foundation for algorithms or analysis. Receiving much attention lately is sequential user or topic modeling with recurrent neural networks. From predictions about sequences or session outcomes [36] to session-based or session-aware recommendation [23,29,31], either the 30 min or a slightly changed temporal constraint is used to detect sessions.

[12] criticized that published studies often do not state how sessions are built. But what is actually worse is that often mechanical sessions are used even when the aim of the study strongly suggests logical sessions [12]. Little thought is put into segmentation. Depending on the application, there are multiple possible definitions on how to structure a user's history [18] and the potential impact of different session models should be more present in research.

4   Conclusion

Algorithms need input data. In Information Retrieval, this input data comes excessively often in the form of interaction logs. Besides laboratory studies, interaction logs represent the main source of information regarding the understanding of users, their information needs and how they interact with search engines or information systems.

Although much effort has been put into segmenting logs in a meaningful way, and although task- and mission-based approaches have received much attention, many recent studies still apply only temporal constraints. They use mechanical sessions to model user context in many different ways (i.e. compare the recent wave of studies using recurrent neural networks). The actual basis for these algorithms are still sessions identified with a 30 min inactivity time out.

This position paper questions the lack of effort put into the pre-processing of interaction logs. A significant amount of thought should be put into the input for

any algorithm. The 30 min inactivity time out might be perfectly fine for most applications - but arbitrarily and unquestioningly applying it as the basis for any and all algorithms may lead to wrong conclusions, no matter the algorithm quality.

References

1. Agichtein, E., White, R.W., Dumais, S.T., Bennet, P.N.: Search, interrupted: understanding and predicting search task continuation. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2012, pp. 315–324 (2012). https://doi.org/10.1145/2348283.2348328

2. Bigon, L., et al.: Prediction is very hard, especially about conversion. Predicting user purchases from clickstream data in fashion e-commerce. CoRR abs/1907.00400 (2019). http://arxiv.org/abs/1907.00400

3. Buzikashvili, N., Jansen, B.J.: Limits of the web log analysis artifacts. In: WWW 2006 Logging Traces of Web Activity Workshop (2006)

4. Cao, H., et al.: Context-aware query suggestion by mining click-through and session data. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008, pp. 875–883 (2008). https://doi.org/10.1145/1401890.1401995

5. Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the worldwide web. Comput. Netw. ISDN Syst. **27**(6), 1065–1073 (1995). https://doi.org/10.1016/0169-7552(95)00043-7

6. Chitraa, V., Thanamani, D.A.S.: A novel technique for sessions identification in web usage mining preprocessing. Int. J. Comput. Appl. **34**(9), 23–27 (2011)

7. Dinuca, C., Ciobanu, D.: Improving the session identification using the mean time. Int. J. Math. Models Methods Appl. Sci. **6**, 265–272 (2012)

8. Downey, D., Dumais, S., Horvitz, E.: Models of searching and browsing: languages, studies, and applications. In: Proceedings of IJCAI 2007, IJCAI 2007, pp. 2740–2747 (2007)

9. Gayo-Avello, D.: A survey on session detection methods in query logs and a proposal for future evaluation. Inf. Sci. **179**(12), 1822–1843 (2009). https://doi.org/10.1016/j.ins.2009.01.026

10. Gomes, P., Martins, B., Cruz, L.: Segmenting user sessions in search engine query logs leveraging word embeddings. In: Doucet, A., Isaac, A., Golub, K., Aalberg, T., Jatowt, A. (eds.) TPDL 2019. LNCS, vol. 11799, pp. 185–199. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30760-8_17

11. Guan, D., Zhang, S., Yang, H.: Utilizing query change for session search. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2013, pp. 453–462 (2013). https://doi.org/10.1145/2484028.2484055

12. Hagen, M., Gomoll, J., Beyer, A., Stein, B.: From search session detection to search mission detection. In: Proceedings of the 10th Conference on Open Research Areas in Information Retrieval, OAIR 2013, pp. 85–92 (2013)

13. Hagen, M., Stein, B., Rüb, T.: Query session detection as a cascade. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM 2011, pp. 147–152 (2011). https://doi.org/10.1145/2063576.2063602

14. He, D., Göker, A.: Detecting session boundaries from Web user logs. In: Proceedings of of the BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research, pp. 57–66 (2000)

15. He, D., Göker, A., Harper, D.J.: Combining evidence for automatic Web session identification. Inf. Process. Manag. **38**(5), 727–742 (2002). https://doi.org/10.1016/S0306-4573(01)00060-7

16. Hienert, D., Kern, D.: Recognizing topic change in search sessions of digital libraries based on thesaurus and classification system. In: Proceedings of the 18th Joint Conference on Digital Libraries, JCDL 2019, pp. 297–300 (2019). https://doi.org/10.1109/JCDL.2019.00049

17. Jansen, B.J., Spink, A., Blakely, C., Koshman, S.: Defining a session on web search engines: research articles. J. Am. Soc. Inf. Sci. Technol. **58**(6), 862–871 (2007)

18. Jiang, D., Pei, J., Li, H.: Mining search and browse logs for web search: a survey. ACM Trans. Intell. Syst. Technol. **4**(4) (2013). https://doi.org/10.1145/2508037.2508038

19. Jones, R., Klinkner, K.L.: Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, pp. 699–708 (2008). https://doi.org/10.1145/1458082.1458176

20. Kotov, A., Bennett, P.N., White, R.W., Dumais, S.T., Teevan, J.: Modeling and analysis of cross-session search tasks. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, pp. 5–14 (2011). https://doi.org/10.1145/2009916.2009922

21. Liao, Z., et al.: A vlHMM approach to context-aware search. ACM Trans. Web **7**(4) (2013). https://doi.org/10.1145/2490255

22. Lucchese, C., Orlando, S., Perego, R., Silvestri, F., Tolomei, G.: Identifying task-based sessions in search engine query logs. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM 2011, pp. 277–286 (2011). https://doi.org/10.1145/1935826.1935875

23. Lv, Y., Zhuang, L., Luo, P.: Neighborhood-enhanced and time-aware model for session-based recommendation. arXiv abs/1909.11252 (2019)

24. Mehrotra, R.: Inferring User Needs & Tasks from User Interactions. Dissertation, University College London, London (2018)

25. Mehrotra, R., Yilmaz, E.: Task embeddings: learning query embeddings using task context. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, pp. 2199–2202 (2017). https://doi.org/10.1145/3132847.3133098

26. Montgomery, A., Faloutsos, C.: Identifying Web browsing trends and patterns. Computer **34**(7), 94–95 (2001). https://doi.org/10.1109/2.933515

27. Murray, G.C., Lin, J., Chowdhury, A.: Identification of user sessions with hierarchical agglomerative clustering. Proc. Am. Soc. Inf. Sci. Technol. **43**, 1–9 (2007). https://doi.org/10.1002/meet.14504301312

28. Piwowarski, B., Dupret, G., Jones, R.: Mining user web search activity with layered Bayesian networks or how to capture a click in its context. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, WSDM 2009, pp. 162–171 (2009). https://doi.org/10.1145/1498759.1498823

29. Quadrana, M., Karatzoglou, A., Hidasi, B., Cremonesi, P.: Personalizing session-based recommendations with hierarchical recurrent neural networks. In: Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, pp. 130–137 (2017). https://doi.org/10.1145/3109859.3109896

30. Radlinski, F., Joachims, T.: Query chains: learning to rank from implicit feedback. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD 2005, pp. 239–248 (2005). https://doi.org/10.1145/1081870.1081899

31. Ruocco, M., Skrede, O.S.L., Langseth, H.: Inter-session modeling for session-based recommendation. In: Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems, DLRS 2017, pp. 24–31 (2017). https://doi.org/10.1145/3125486.3125491

32. Sen, P., Ganguly, D., Jones, G.J.: Tempo-lexical context driven word embedding for cross-session search task extraction. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (Long Papers), vol. 1, pp. 283–292 (2018). https://doi.org/10.18653/v1/N18-1026

33. Silverstein, C., Marais, H., Henzinger, M., Moricz, M.: Analysis of a very large web search engine query log. SIGIR Forum **33**(1), 6–12 (1999). https://doi.org/10.1145/331403.331405

34. Spink, A., Jansen, B.J., Wolfram, D., Saracevic, T.: From e-sex to e-commerce: Web search changes. Computer **35**(3), 107–109 (2002). https://doi.org/10.1109/2.989940

35. Spink, A., Park, M., Jansen, B.J., Pedersen, J.: Multitasking during web search sessions. Inf. Process. Manag. **42**, 264–275 (2006). https://doi.org/10.1016/j.ipm.2004.10.004

36. Twardowski, B.: Modelling contextual information in session-aware recommender systems with neural networks. In: Proceedings of the 10th ACM Conference on Recommender Systems, RecSys 2016, pp. 273–276 (2016). https://doi.org/10.1145/2959100.2959162

37. Völske, M.: Retrieval enhancements for task-based web search. Dissertation, Bauhaus-Universität Weimar, Weimar, Germany (2019)

38. Wang, H., Song, Y., Chang, M.W., He, X., White, R.W., Chu, W.: Learning to extract cross-session search tasks. In: Proceedings of the 22nd International Conference on World Wide Web, WWW 2013, pp. 1353–1364 (2013). https://doi.org/10.1145/2488388.2488507

39. White, R.W., Drucker, S.M.: Investigating behavioral variability in web search. In: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, pp. 21–30 (2007). https://doi.org/10.1145/1242572.1242576

40. Ye, C., Wilson, M.L.: A user defined taxonomy of factors that divide online information retrieval sessions. In: Proceedings of the 5th Information Interaction in Context Symposium, IIiX 2014, pp. 48–57 (2014). https://doi.org/10.1145/2637002.2637010

41. Yuankang, F., Zhiqiu, H.: A session identification algorithm based on frame page and pagethreshold. In: 2010 3rd International Conference on Computer Science and Information Technology, vol. 6, pp. 645–647 (2010)