Client-Side Indexes for Fast Full-Text Searching

Amy X. Zhang
axz@mit.edu

Lea Verou
leaverou@mit.edu

Manali Naik
manalinaik@mit.edu

ABSTRACT

Many applications on the web use a combination of client-side and server-side data stores to facilitate fast interactive and data-intensive experiences. However, standard client side databases within browsers do not currently support full-text searching. In this paper, we describe a client-side search engine built on top of IndexedDB that makes use of several types of indexes common to many well-known server-side search engines. We compare the performance of different indexes on different types of full-text content and queries and find that.... We also compare the performance of our system with that of fully server side systems and examine scenarios where a hybrid approach may be fastest. We find that...

1. INTRODUCTION

Today

2. BACKGROUND AND RELATED WORK

There are a lot of cases where apps store a heavy amount of data in the browser. This study gives users ubiquitous access to data by allowing browser session migration [5]

client-side profiles for personalized advertising, privacy concerns [3]

Client-side database storage can improve the performance of data intensive websites by executing portions of web applications client-side and synchronizing with a web server. [2]

Search engines can be useful in the absence of connectivity on mobile phones. This system builds on a user study showing that revisitation is common. [1]

Previous research has demonstrated that it is feasible to store a reverse index within IndexedDB [4] though it is much slower than using a server-side application such as Lucene.

3. LUCY.JS

4. INDEX IMPLEMENTATIONS

5. EVALUATION

6. DISCUSSION

7. FUTURE WORK

8. CONCLUSION

9. REFERENCES

[1] A. Balasubramanian, N. Balasubramanian, S. J. Huston, D. Metzler, and D. J. Wetherall. Findall: A local search engine for mobile phones. In *Proceedings of the 8th international conference on Emerging networking experiments and technologies*, pages 277–288. ACM, 2012.

[2] E. Benson, A. Marcus, D. Karger, and S. Madden. Sync kit: a persistent client-side database caching toolkit for data intensive websites. In *Proceedings of the 19th International Conference on World Wide Web*, pages 121–130. ACM, 2010.

[3] M. Bilenko and M. Richardson. Predictive client-side profiles for personalized advertising. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 413–421. ACM, 2011.

[4] J. Lin. *On the Feasibility and Implications of Self-Contained Search Engines in the Browser*. arXIV, 2014.

[5] J. T. K. Lo, E. Wohlstadter, and A. Mesbah. Imagen: Runtime migration of browser sessions for javascript web applications. In *Proceedings of the 22nd international conference on World Wide Web*, pages 815–826. International World Wide Web Conferences Steering Committee, 2013.