

LASER: A Living AnalyticS Experimentation System for Large-scale Online Controlled Experiments

Kwan Hui Lim, Ee-Peng Lim, Palakorn Achananuparp, Adrian Vu,
Agus Trisnajaya Kwee and Feida Zhu

Living Analytics Research Centre, Singapore Management University
{khlim, eplim, palakorna, adrianvu, aguskwee, fdzhu}@smu.edu.sg

ABSTRACT

Tracking user browsing data and measuring the effectiveness of website design and web services are important to businesses that want to attract the consumers today who spend much more time online than before. Instead of using randomized controlled experiments, the existing approach simply tracks user browsing behaviors before and after a change is made to website design or web services, and evaluate the differences. To address the effects caused by hidden factors (e.g. promotion activities on the website) and to give fair comparison of different website designs, we propose the LASER system, a unified experimentation platform that enables randomized online controlled experiments to be easily conducted with minimal human effort and modifications to the experimented websites. More importantly, the LASER system manages the various aspects of online controlled experiments, namely the selection of participants into groups, exposure of different user interface features or recommendation algorithms to these groups, measuring their responses, and summarizing the results in the visual manner.

Categories and Subject Descriptors: H.3.4 [Information Storage and Retrieval]: Systems and Software - Performance evaluation

General Terms: Measurement, Design, Experimentation

Keywords: Online Controlled Experiments, A/B Testing, User Selection, Recommendations, User Interface

1. INTRODUCTION

Large-scale online controlled experiments allows the effective evaluation of website design and web services using real users in real-life usage scenarios, moving away from the traditional trend of using recruited participants as test subjects. However, considerable effort is needed to conduct such experiments, which includes the selection of participants into groups, exposure of different user interface features or recommendation algorithms to these groups, measuring their responses, summarizing and analyzing results in the visual manner. Moreover, the implementation of such experiments normally requires extensive modification to the underlying

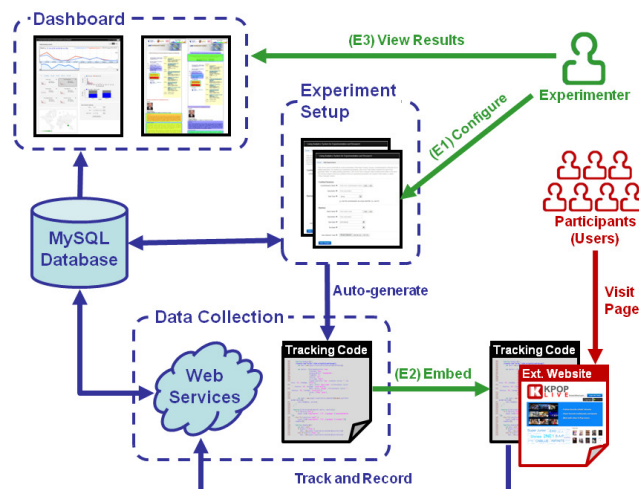


Figure 1: LASER System Overview

code of the experimented website. In this paper, we proposed a system that simplifies and automates the design and running of large-scale online controlled experiments.

While large companies such as Google, Yahoo! and Microsoft have used online controlled experiments to evaluate new features before their full-scale deployment on the web [3, 4, 6], many smaller companies and university researchers have not yet shifted to this experimentation approach. In fact, most university researchers still stick to small scale user experiments recruiting human subjects largely from university student population which may not be representative of real users. The experiments usually involve scenarios that are artificially created rather than the natural ones. Lately, with the help of Mechanical Turks, one can scale up their experiments to involve many more human subjects. This however has not addressed the realism of experiment scenarios.

There are commercial applications (e.g. Google Analytics¹, MixPanel², Flurry Analytics³, Optimizely⁴, etc) that support basic tracking and simple A/B testing. However, these applications are not designed for complex customized experiments such as multi-variate tests or experiments using a customized user selection algorithm. In addition, these applications are not designed to evaluate recommendation

¹<http://www.google.com/analytics>

²<https://mixpanel.com>

³<http://www.flurry.com>

⁴<http://www.optimizely.com>

Figure 2: Experiment Setup Form

algorithms. As such, we designed the LASER system to provide experimenters with the flexibility to define their own user selection algorithm, evaluate recommendation algorithms and track fine-grained user activities, in a real-life experimentation setting.

1.1 Contributions

Our main contribution includes proposing the LASER system that allows experimenters to conduct customized online experiments with minimal human effort and modifications to their existing websites. The key features of the LASER system are as follows:

- A unified experimentation platform for evaluating user selection algorithms, recommendation algorithms, UI features, webpage variations, or to simply perform basic tracking.
- Ease of customizing experiments without the need for any programming. Experiment setup is performed via a GUI form.
- Minimal effort to implement the LASER tracking on the experimented website, by including a single LASER tracking code that is auto-generated.
- The ability to track both registered users (i.e. websites with log-in functions) and anonymous users. These users can be further separated into control and treatment groups, using either the built-in randomized grouping or a customized user selection algorithm.
- The automatic tracking of fine-grained user statistics (e.g. clicks and mouse-overs on various HTML elements) and monitoring of overall performance metrics (e.g. conversion rate based on user registration, check-out, etc)
- A dashboard that provides a visual summarization of the experiment results in real-time.

2. SYSTEM ARCHITECTURE

The LASER system is developed as a web-based system on the experimenter platform (i.e. viewing and setup of experiment parameters) while JavaScript/jQuery/AJAX is used for the tracking code that will be included on the website being experimented on. LASER also includes a dashboard component which gives an overview of key experiment statistics that are displayed using Google Charts.

The LASER system can be divided into three components according to the main functionalities provided. Fig. 1 gives an illustration of how an experimenter typically interacts

Table 1: Experiment Parameters

Parameter Name	Description
Web Page Tracking	A list of URLs indicating the webpages on which the tracking will be performed.
Groups	The user groups defined for an experiment. Each group can be assigned a particular recommendation algorithm, landing website or UI to be evaluated on.
Cookies/Sessions	A list of cookies that is used by the website. In particular, if a cookie contains the unique userID, this userID would be used as input for the user selection algorithm (for user grouping).
Batches	Each batch of an experiment would have an associated start/end date where an experiment is active only when there is a batch with a currently active start/end date.

with these LASER components. In Step E1, the experimenter defines the type of experiment to be conducted and its associated details using the experiment setup component. Thereafter, a tracking code is automatically generated by LASER and the experimenter embeds this tracking code on the experimented website (Step E2). This tracking code forms part of the data collection component which manages the primary task of tracking the user activities and storing it on the LASER system. Upon a user arriving, the tracking code also assigns this user into treatment or control groups. Finally in Step E3, the experimenter is able to use the data visualization component to view key statistics relating to the collected experiment data, either in real-time or post-experiment. We now further elaborate on each LASER component.

2.1 Experiment Setup

This component allows the experimenter to configure various parameters of the experiment via a GUI form as shown in Fig. 2. For example, the Groups parameter defines the number of experimental groups and type of treatment to administer. Table 1 shows a list of these parameters and their main purposes in the experiment setup. These experiment parameters provide the flexibility to define many variations of experiments. To better illustrate how various customized experiments can be conducted, we give some examples of the possible usage scenarios, as follows:

1. **Perform simple website usage analysis.** In its simplest form, we can perform basic tracking by defining only a single group.
2. **Conduct a simple randomized controlled test for UI elements or recommendation algorithms.** We can simply define two Groups: one as a control and the other being a treatment that is exposed to a particular UI element or recommendation algorithm. Being a simple controlled test, we could use the in-built LASER function for the randomized grouping of users (new visitors). For the exposure of UI (for a particular Group), this can be either via dynamic HTML modification (using jQuery) or website redirection. The dynamic HTML modification is preferred for testing a single UI element and the website redirection for testing multiple, extensive UI elements.

3. **Conduct a split test using a specific user grouping algorithm.** Similar to usage scenario 2, except that we now group users into control and treatment using a particular user selection algorithm. For example, to measure any cultural difference in user behavior, we could use a customized user selection algorithm that groups all users from U.S. into control and users from Asia into treatment.
4. **Conduct a split test using a pre-defined user grouping.** Similar to usage scenario 3, except that the evaluated website is one where users are able to log-in with a unique user ID. In this case, we might want to compare between two groups of pre-selected users (e.g. one group could comprise mutual friends while the other comprises only strangers).
5. **Conduct a multi-variate test.** In some cases, we might be interested in conducting a multi-variate test to evaluate multiple UI features and/or recommendation algorithms. We achieve this by defining a control group and multiple treatment groups, each of which is exposed to a particular UI element or recommendation algorithm.

Upon completing the experiment setup, the tracking code will be auto-generated based on the experiment parameters defined. The next step for the experimenter is to simply embed this tracking code on the experimented website (e.g. `<script type=text/javascript src=http://laser.com/track.js>`).

2.2 Data Collection

The data collection component comprises two main sub-components, namely the tracking code that collects user data relating to the experiment, and supporting web services that this tracking code interfaces with. At this stage, the only intervention required of the experimenter is to embed the tracking script on the experimented website.

2.2.1 Tracking Code

This tracking code is stored on the LASER server with the main advantage being that experiment parameters can be kept current within the tracking code, should the experimenter make any changes. The inclusion of the tracking code should be placed at the end of the HTML page where the tracking will be conducted. Placing the tracking code at the end of the page ensures that the webpage is loaded first and shown to the user while the tracking code executes subsequently (in the background).

The tracking code first checks if it is on a valid website and there is a currently active batch. If so, it calls the LASER web services to check the user group this current user belongs to and the session ID it is assigned to. The user group and session ID are then stored as two separate LASER cookies which are continuously refreshed based on a predefined time interval (to keep track of the user session). While these cookies are active (meaning it is a valid user session), tracking and logging of the user events are performed. As an effort to speed-up the execution time of the tracking code, calls to web-services are made asynchronously.

Due to the implementation of the tracking code in JavaScript, one limitation of the LASER system is the requirement for users to have JavaScript enabled in their web browsers. However, this limitation has minimal impact on the effectiveness

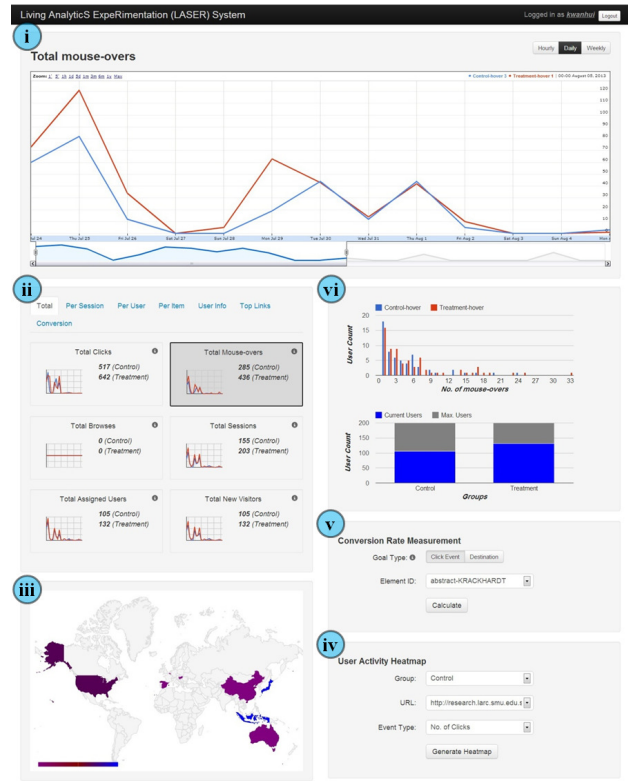


Figure 3: LASER Dashboard

of the LASER system as studies have shown that only 2% of people have JavaScript disabled [7].

2.2.2 Web Services

These web services provide an interface for the tracking code to interact with the LASER system and its back-end database. Broadly, these web-services allow the tracking code to perform the following tasks: (i) Update the user group that a new user is assigned to; (ii) Track user sessions (log-ins or anonymous visits to the experimented website); and (iii) Log user activities (e.g. mouse clicks, mouse-overs, item browses, etc).

In addition, there are web-services such as the user selection and recommendation algorithms that could potentially be de-coupled from the LASER system. This de-coupling allows for LASER to access the user selection and recommendation algorithms as black-box functions (via the external web-services). The main advantage to this approach is when the vendor holds proprietary data about its users and items which is not released to the LASER system (for the purposes of user selection and item recommendation). Many similar systems have also utilized designs where selected functionalities are de-coupled or loosely-coupled and provided via external components [1, 2].

2.3 Data Visualization

The dashboard provides an overview of the key statistics on experiments as these results are being tracked and computed in real-time. The dashboard comprises a series of pre-defined charts that are generated using Google Charts⁵ and

⁵<https://developers.google.com/chart/>

can be divided into six main component as shown in Fig. 3, namely (anti-clockwise from the top): (i) a temporal line chart of the selected statistic (mouse clicks, mouse-overs, session duration, new visitors, etc); (ii) aggregated statistic of user activities, sessions, demographics, etc; (iii) world map of visitor location; (iv) heatmap of user activity (examples shown in Fig. 4); (v) dynamic calculation of conversion rate; (vi) histogram of user grouping statistic.

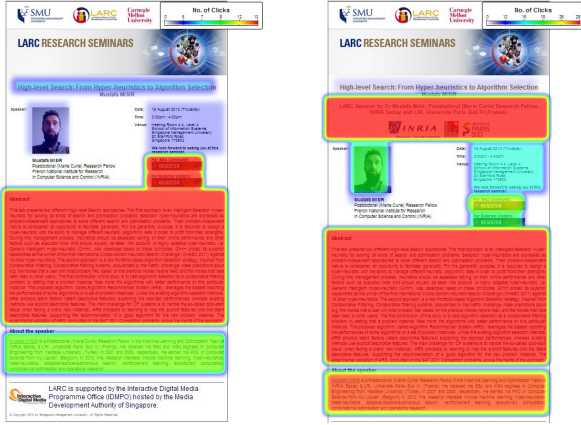


Figure 4: Mouse-clicks Heatmap of Control (Left) and Treatment (Right) groups. In this example, we are experimenting with the effect of an additional authority message (Treatment) highlighting the credentials of a seminar speaker.

3. SYSTEM DEMONSTRATION

Our demonstration will revolve around the three main components (experiment setup, data collection, and data visualization) of the LASER system. We aim to provide the audience with a comprehensive experience of using the LASER system by showcasing the following:

- A brief overview of the experiment setup process for conducting the different types of experiments (as per the various usage scenarios in Section 2.1).
- Hands-on interaction with a website embedded with the LASER tracking code to show the real-time collection and viewing of user activity data.
- Usage of the dashboard to show the results from several concluded experiments based on the live deployment of the LASER system.

We will first demonstrate how the various types of experiments (recommendation algorithm, UI testing) can be conducted via the experiment setup. The audience will then be able to interact with a sample website embedded with the LASER tracking code. In addition, we will also allow the audience to access the LASER dashboard to view a summary of the data that is being collected in real-time.

The LASER system has since been used to conduct various experiments and tracking on more than 30 different websites between May 2013 to Feb 2014. Till now, LASER has successfully tracked more than 70,000 users (unique visitors) who generated over 1.5 million user events (mouse clicks, mouse-overs, user sessions). In particular, we have conducted a series of controlled experiments to examine the

effect of authority messages on user behavior on events websites. Using results from these actual experiments, we will demonstrate how the dashboard can be used to view and study the collected data from a top-level view of aggregated statistics to a fine-grained analysis of user activity. One of our main finding is how an authority message effectively draws the visitor's attention to page elements related to the authority speaker. As this paper focuses on the technical aspects of the LASER system, we only present an overview of our experimental findings and refer readers to [5] for a more detailed analysis.

4. CONCLUSION AND FUTURE WORK

In this paper, we presented the LASER system that provides a non-intrusive approach to conducting online controlled experimentations to evaluate UI features and recommendation algorithms. The experimenter has the flexibility to configure various aspects of the experiment via a GUI form, and only needs to embed the auto-generated tracking code on the experimented website. Thereafter, the LASER system manages the running of this experiment including the grouping of users, showing of different treatment and tracking of user responses. All the collected result is then made available in real-time or post-experiment as a dashboard.

We have also performed a preliminary deployment of the LASER system by conducting various experiments and tracking on more than 30 different websites using the LASER system. As part of our future work, we will extend the LASER system with the functionality to also conduct network experiments where users have to be carefully selected to avoid interference of treatments.

Acknowledgments. This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office, Media Development Authority (MDA). The authors would also like to thank Michael Finegold and Juan Du for their comments.

5. REFERENCES

- [1] R. Bierig, J. Gwizdka, and M. J. Cole. A user centered experiment and logging framework for interactive information retrieval. In *Proceedings of SIGIR'09 Workshop on Understanding the User*, pages 8–11, Jul 2009.
- [2] M. M. Hall and E. Toms. Building a common framework for IIR evaluation. In *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, volume 8138 of *LNCS*, pages 17–28, Sep 2013.
- [3] R. Kohavi, A. Deng, B. Frasca, T. Walker, Y. Xu, and N. Pohlmann. Online controlled experiments at large scale. In *Proceedings of KDD'13*, pages 1168–1176, Aug 2013.
- [4] R. A. Lewis, J. M. Rao, and D. H. Reiley. Here, there, and everywhere: Correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of WWW'11*, pages 157–166, Mar 2011.
- [5] K. H. Lim, B. Jiang, E.-P. Lim, and P. Achananuparp. Do you know the speaker? An online experiment with authority messages on event websites. In *Proceedings of WWW'14 Companion*, Apr 2014.
- [6] D. Tang, A. Agarwal, D. O'Brien, and M. Meyer. Overlapping experiment infrastructure: more, better, faster experimentation. In *Proc. of KDD'10*, pages 17–26, Jul 2010.
- [7] N. C. Zakas. How many users have javascript disabled? Internet, Oct 2010. Available from: <http://developer.yahoo.com/blogs/ymdn/many-users-javascript-disabled-14121.html>.