

# Assessment of New Wikipedia Portal's Search Box Deployment

*Mikhail Popov*

*March 18, 2016*

## Executive Summary

On 10 March 2016 the Discovery/Portal team deployed a change to the Wikipedia Portal (wikipedia.org) as a result of a successful A/B test. However, when looking at the Discovery Dashboards, we noticed that the clickthrough rates we were seeing did not align with our expectations of the positive impact of the change. In this report we performed an intervention analysis of the clickthrough rate time series data. We found that the deployment of the new search box with suggestions has had a statistically significant positive impact ( $p = 0.011$ ) on the search clickthrough rate, raising it by approximately 1.88% on average.

## Introduction

On 10 March 2016 the Discovery/Portal team deployed a change to the Wikipedia Portal (wikipedia.org) as a result of a **successful A/B test**.

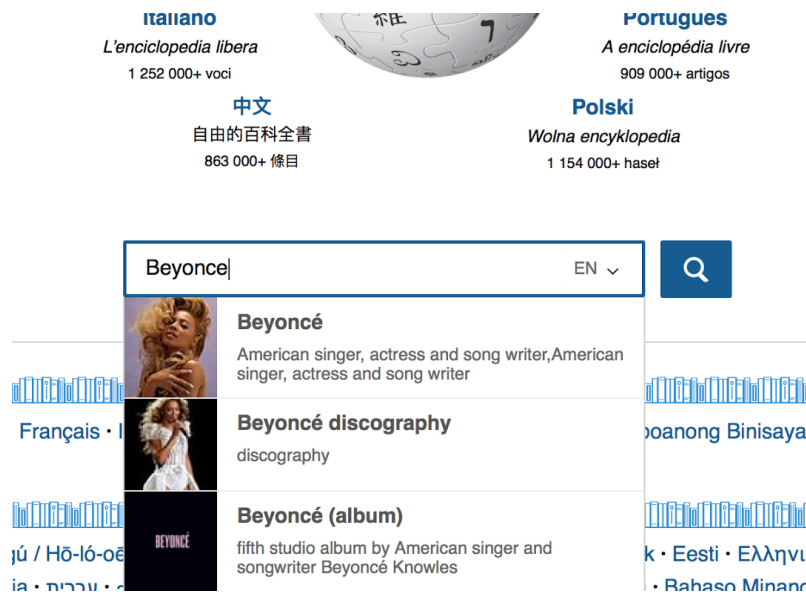


Figure 1: New search box and type-ahead update pushed to <http://www.wikipedia.org> on 10 March 2016.

In the **A/B test report**, we reported that 1.7-5.5% more sessions were likely to end in a clickthrough compared to the control, and that the probability of the user making a search-related clickthrough went up by 1.88-5.88%, both substantial improvements. However, when looking at the **Discovery Dashboards**, we noticed that the clickthrough rates we were seeing did not align with our expectations of the positive impact of the change.

The daily fluctuations, possibly small effect, and only a week of data post-deployment make it difficult to visually tell if there has been an impact (positive OR negative). In this report, we perform an intervention analysis of the clickthrough rates.

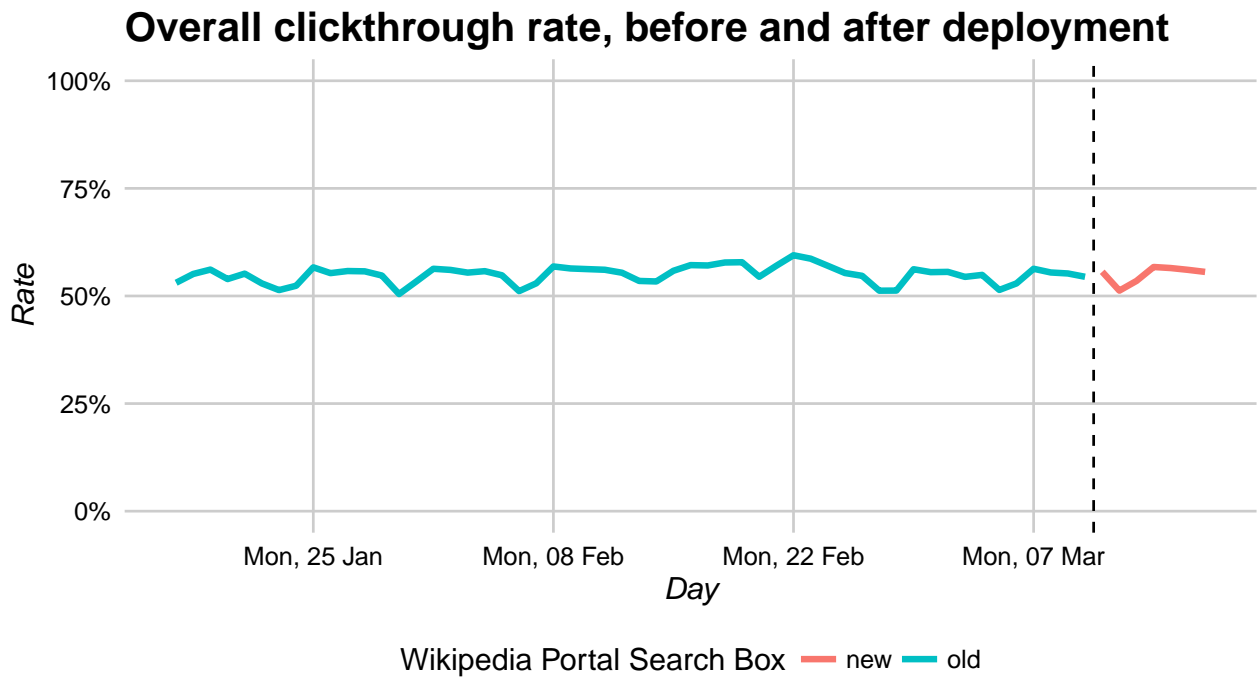


Figure 2: No visible impact on overall clickthrough rate from deploying the new search box.

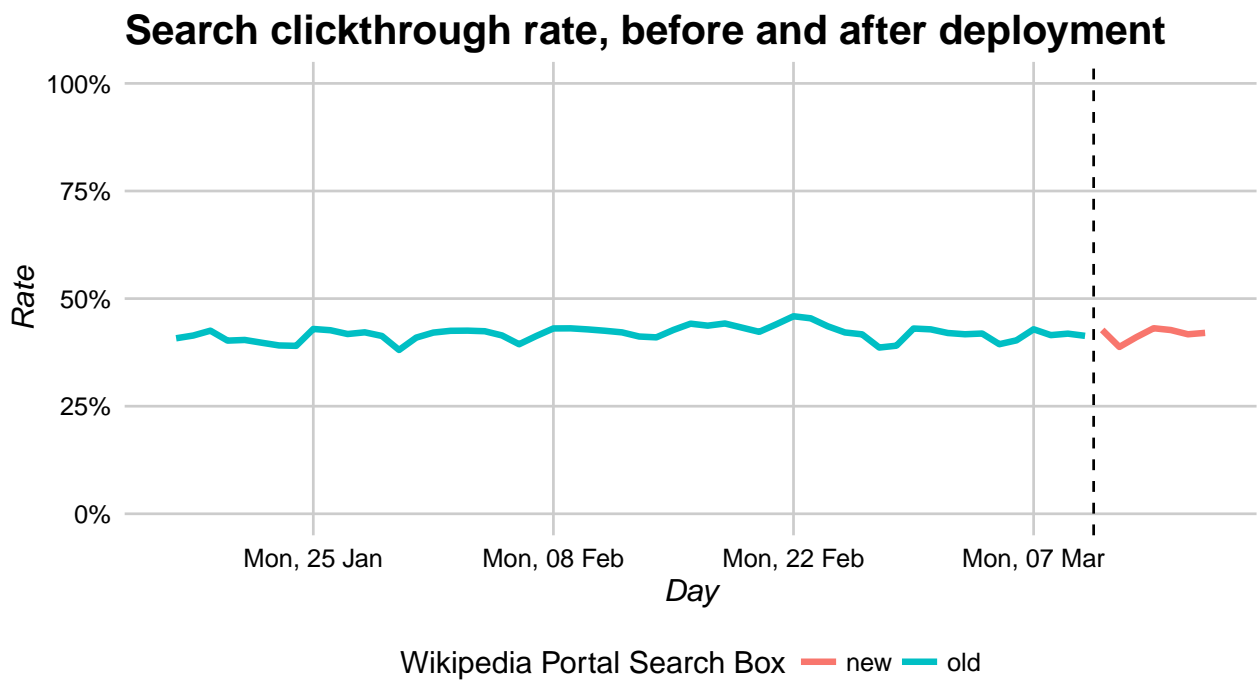


Figure 3: No visible impact on search clickthrough rate from deploying the new search box.

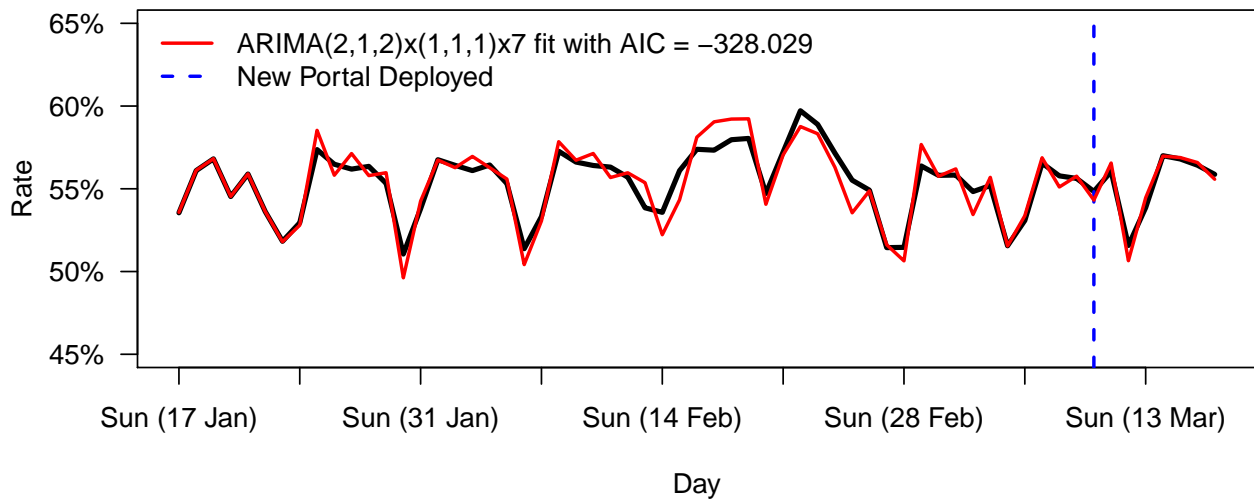
## Methods

We explored a variety of models and settled on seasonal ARIMA with a period of 7 days. This model of click-through rate  $\mathbf{y} = y_1, \dots, y_n$  with accompanying intervention indicator  $x_t \in \{0 \text{ if old portal, } 1 \text{ if new portal}\}$  is specified as:

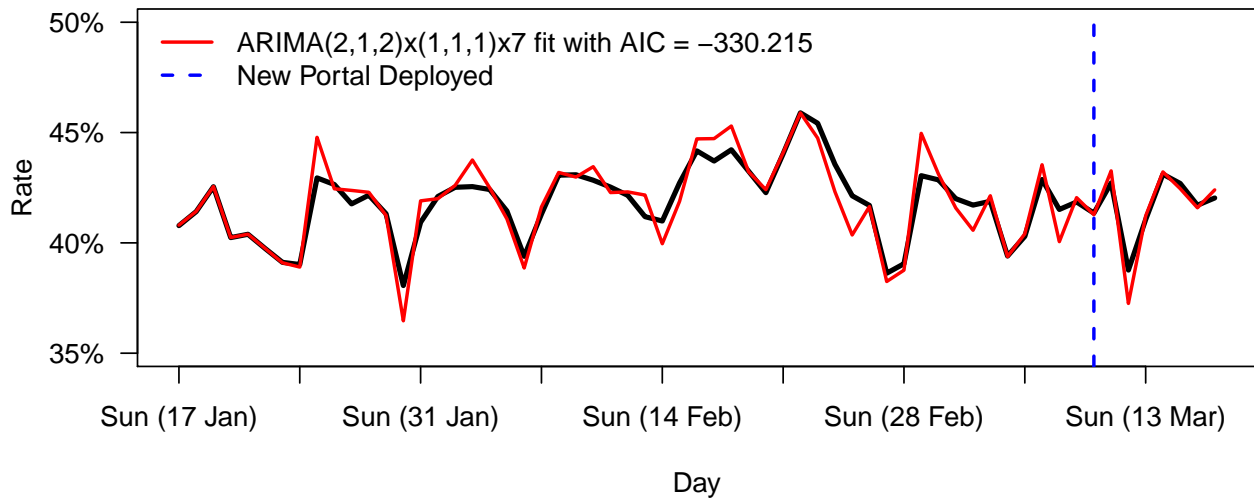
$$y_t = \epsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \beta x_t$$

We are interested in inference on  $\beta$ , the unknown permanent change in the mean due to the intervention – the deployment of the new search box.

### Overall Clickthrough Rate on Wikipedia Portal



### Overall Clickthrough Rate on Wikipedia Portal



Visual inspection and model evaluation metrics showed the models we used accurately captured the time series pattern, its seasonality, and without overfitting.

## Results and Discussion

Table 1: Summary table of inference on the effect of deploying the update.

Effect of deployment on...	% Change	Std. Error	95% C.I.	p-value
Overall (All Sections) CTR	1.400	0.725	(-0.02%, 2.82%)	0.054
Search CTR	1.881	0.739	(0.43%, 3.33%)	0.011 *

The deployment of the new search box with suggestions has had a statistically significant positive impact on the search clickthrough rate, raising it by approximately 1.88% (0.43%-3.33%) on average. As we expected, the deployment did not have statistically significant impact on the overall clickthrough rate, which includes users clicking on language links and Wikimedia project links.

It should be noted (again) that these results are based on a week of post-deployment event logging data. It will be interesting, if not necessary, to revisit this particular question again in several weeks' time and perform a re-evaluation of the deployment's impact. However, between the results of the initial A/B test and this report, the evidence does point toward a successful launch.

## References

Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis*. New York, NY: Springer Science & Business Media. <URL: <http://doi.org/10.1007/978-0-387-75959-3>>.

Hyndman RJ (2015). *forecast: Forecasting functions for time series and linear models*. R package version 6.2, <URL: <http://github.com/robjhyndman/forecast>>.

Hyndman RJ and Khandakar Y (2008). "Automatic time series forecasting: the forecast package for R." *Journal of Statistical Software*, 26(3), pp. 1-22. <URL: <http://ideas.repec.org/a/jss/jstsof/27i03.html>>.