First in Search – How to Optimize Search Results in E-Commerce Web Shops

Gerald Petz and Andreas Greiner

University of Applied Sciences Upper Austria, Campus Steyr, Austria
{gerald.petz,andreas.greiner}@fh-steyr.at

Abstract. Customers of e-commerce web sites frequently use the full text search to find the desired products. The ranking of the search result page depends on various criteria such as the matching of search terms or popularity of the product. E-commerce vendors usually use additional ranking criteria and may want to increase conversion rates by varying the rankings of the search hits. This paper proposes a method to measure the impact of changing the ranking of the search result page. The method is applied to a b2b e-commerce shop with office products.

Keywords: ranking, search result page, web shop, e-commerce, measurement, A/B-testing, conversion.

1 Introduction and Motivation for Research

Although the estimates about global e-commerce trends vary, it is certainly true that e-commerce is growing fast; developed economies dominate the market, but emerging economies are expected to catch up soon. [1–5] According to the "e-commerce-guideline"-study e-commerce revenues in Germany rose from 18.3 billion Euros in 2010 to an estimated 25 billion Euros in 2012.[6] About 86% of Germany's online retailers run their own web shop; of course other channels such as online auction platforms are used as well.[6]

Whenever a user wants to buy products from a specific web shop he can use various alternatives to find the desired product: rummage in product lists, browse products by category, use faceted search or use full text search. The full text search plays an essential role in an e-commerce system: up to 80% of the visitors use only the full text search to find the desired products – a phenomenon that seems to be learned from usage of the Google search engine. One third of the visitors leave a web shop because they cannot find the desired products – even if the products are offered. The search engine has to deliver search hits accurately and quickly and has to be tolerant of typing mistakes and synonyms, and the search has to understand industry jargon.[6] If the customer uses the full text search it leads to the question as to how the search hits in the search result page can be ordered. Several sort criteria can be identified and of course combined together such as matching of search text to product title and / or product description and / or product category, average customer review, popularity, price, and many others. However, the ranking that is desired by the customer will be different to the desired ranking of

F.F.-H. Nah (Ed.): HCIB/HCII 2014, LNCS 8527, pp. 566–574, 2014.

the company that runs the web shop. A company may want to rank products according to different criteria. It may, for example, rank those with the highest contribution margins highest or those which are discontinued items or fast moving consumer goods. Alternatively, it may be important for a company to rank goods which are on stock highest or which should be sold as quickly as possible for various other reasons. As a result of this, web shop operators try to combine and weigh several ranking criteria and assume that products that are displayed at the top of the page have better conversion rates and are thus ordered more often. A number of software products (like Factfinder, exorbyte, celebros, and many others) support these considerations. However, two questions remain: What is the impact on the conversion when a product is better ranked? And how can we measure this impact in an environment that does not support a simple parallel A/B test setting due to technical restrictions?

The *objective* of this paper is to develop a method to measure the impact of variations of the search result page of a web shop, to apply this method and to evaluate the factual impact. In order to attain this objective, we set up a methodology as follows:

1. Development of the measurement method.
   The quite technical complex infrastructures of larger web shops (web shop software, combined with an ERP-system, a system that enables rankings according to specific product attributes, load balancers) and the complex environment (b2c- as well as b2b-customers with different price structures, product portfolios, etc.) prohibit the application of a simple A/B-test setting. Hence a method to measure the impact of different rankings has to be developed. Due to these technical restrictions, the method is a trade-off between a scientifically sound measurement and a technically realizable measurement method.
2. Implementation of the method.
   The method will be implemented with several tools, e.g. Google Analytics.
3. Application of the method.
   The method will be used in one specific case (web shop with office materials).

2 Related Work, Background

A/B testing and multivariate testing are commonly used in web development; these methods allow website operators to run experiments on website users. A/B testing is an experiment that compares two versions (A and B) of a webpage; the versions are identical except for one variation. The versions are randomly displayed to the visitors; the version that contributes most to the goal conversions is the one which is preferred by the visitors. Multivariate testing is similar to A/B testing, but enables us to test more than two different versions at the same time. [7, 8]

Several studies and publications focus on "success-factors" for e-commerce websites. A number of researchers investigate the connection between usability and the success of e-commerce sites: [9] evaluated commercial websites in order to find usability problems; [10] emphasize the importance of user-friendly interface of electronic shops; they applied heuristic evaluation to examine the usability of several e-commerce sites. As a result the authors provided a set of usability guidelines. Some researchers broadened the evaluations and also took related attributes into consideration (e.g. design attributes [11], aesthetic design and complexity [12]). Some researchers discuss convenience as an important factor for online shopping; convenience in e-commerce is defined as the range to which customers feel that a website is simple, sensory and user-friendly. [13, 14] Additionally, sometimes the cultural context in multilingual websites is considered as well (e.g. [15]).

An important factor for the success of e-commerce sites is trust. Users often hesitate to place orders on web shops because of uncertainty about the vendor, vendor behavior or perceived risks. A variety of research work focusses on this topic, e.g. [16] developed a typology and trust measures for e-commerce, [17] investigate the impact of trust on purchase decisions in the context of e-products and e-services.

The impact of online reviews and electronic word-of-mouth offers a broad range of research activities: e.g. [18] investigate the impact of online reviews on revenues of consumer electronics and video games. The authors show that reviews have significant impacts on revenues, but that the effect decreases over time. [19] determine the impact of online travel reviews, [20] test the impact of hotel reviews.

Interestingly, to our best knowledge we could not find scientific papers that research the impact of the ranking of search results in e-commerce shops. One can find many blogs and more or less reliable "studies" about this topic (e.g. [21]); especially in the area of search engine optimization we can find many hints, blog posts and "studies". We can summarize the discussions simply as: the better the ranking of a search hit, the better the conversion rate is. Unfortunately, there are no publicly available reliable investigations about the impact of search rank on the conversion rate of a product in e-commerce systems.

3 Measurement Method

The setup of the measurement method is proposed as follows:

- Experimental setup.
  We define a control group and an experimental group of products in three different product categories. The control groups have the "usual" ranking factors; the experimental groups are based on other ranking factors. Since it is technically not possible to measure the effects (comparable to an A/B test setting) temporally parallel, the measurements are carried out alternately in time sequence (see figure below). In order to avoid biases, we use product categories that have no seasonal fluctuations and chose weeks that contain no bank holidays.

Table 1. Setup of timing

		Week							
		1	2	3	4	5	6	7	8
Category 1	control group 1	X		X		X		X	
	experimental group 1		X		X		X		X
Category 2	control group 2	X				X		X	
	experimental group 2		X		X		X		X
Category 3	control group 3	X				X		X	
	experimental group 3		X		X		X		X

- Tracking.
  In order to track the effects, we set up a web analytics tool (Google Analytics). Several settings and prerequisites have to be undertaken: event tracking and e-commerce tracking have to be configured and the tracking code has to be implemented in the web shop. The event tracking should detect that (i) the full text search was used, (ii) a product of one of the monitored categories was put into the basket from the search engine result page or from the subsequently loaded product page, (iii) a product of one of the monitored categories was put on the watch list. The e-commerce tracking logs the transaction data.
- Export and data analysis.
  The gathered data have to be exported from the analytics tool and merged with the exported product data from the ERP-system that contains the detailed configuration ranking settings.

Discussion of the setup and remaining challenges:

- Trade-off
  As mentioned above, the proposed measurement method is a trade-off between a scientifically sound measurement and a technically and economically possible measurement.
- Deleted cookies, different browsers
  Most web analytic tools rely on cookie tracking which means there is already an inaccuracy resulting from the use of different browsers or deleting existing cookies.
- No transfer of the referrer
  Our javascript event tracking needs for the tracking of "add to basket" clicks from the product page (after the use of the full text search) the referrer. Some company's firewalls don't transfer it, so this may lead to fuzziness of the tracking.
- Impact of situation in b2b-webshops
  Since the measurement takes place in a b2b-environment, we were able to find find an order scenario as follows: user A puts products in basket, user B approves the basket and places the order. Thus, the question as to how this scenario could be measured (or excluded) arises.
- Effects of users' behavior
  The web shop offers the users a watch list to collect products for later ordering; the impact on the usage is not easy to measure.
- Inaccuracies due to different browsers
  Most of the common web analytics tools use cookie tracking. Therefore inaccuracies can arise if different browsers are used or if existing cookies are deleted.
- Temporal connections
  We have to face order scenarios where a user searches a product und puts it in the basket, but orders the basket a couple of days later. In this scenario the question remains concerning how the temporal connection between the search event and the order event can be established.

4 Application of Measurement Method, Results

4.1 Application of Measurement Method

We applied the suggested measurement method to a specific web shop: the web shop is a b2b-web shop and contains office products. The b2b-scenario implies that different customers are offered a different product spectrum and that customers may have different prices, terms and conditions.

As described in the section above, we defined three different product categories: file folder (108 products), text highlighter (116 products) and correction products and correction fluids (30 products). These product categories were chosen because there are no seasonal fluctuations. The next step was to implement the tracking functions; we decided to use Google Analytics. In order to track the interesting measures, we had to implement the following functions:

- E-Commerce tracking
  The first step was to activate the e-commerce tracking option. After activation one can use the Javascript-functions "_addTrans()", "_addItem()" and "_trackTrans()" to track transactional data.
- Tracking site search
  This option is an elective one, but it is useful to record search terms.
- Event tracking
  Subsequently, the event tracking has to be implemented. Google Analytics offers the Javascript-function "_trackEvent(category, action, opt_label, opt_value, opt_noninteraction)" to track events. The following events have to be taken into consideration: (i) search leads to products in the defined categories, (ii) product is put into the basket (directly from the search engine result page or indirectly from the product-detail page), and (iii) product is put on the watch list. The function "_gaq.push" fires the tracking events to Google Analytics.

As described in chapter 3, the measurement took place during 8 weeks between mid-September and mid-November. Both the control group the experimental group contained the same product categories and products. The only difference was that different ranking weights ("spread configuration") were applied to the experimental group. The meaning of the spread is as follows:

- Spread 0: no devaluation
- Spread 1: devaluation of 0.33%
- Spread 2: devaluation of 0.66%
- Spread 3: devaluation of 0.99%

We hypothesize that products that are devaluated with a spread-configuration (i) are put into the basket less often and (ii) are less often ordered. The comparison takes place by using the control group (no spread configuration applied) and the experimental group (spread configuration is active).

4.2 General Results

As a first step some general statistics were examined. Fig. 1 exhibits the number of search terms used in full text search. 64% of the searches contain only 1 search term, 22% contain two terms; the remaining 14% contain more than two terms. The searches with one term can be distinguished in searches with a keyword (78%) and searches with a specific product ID (22%). These numbers lead to the assumption that the users do not have one specific product in mind when using the full text search.

Fig. 2 exhibits, how many users visit the second, third, etc. page of the search result pages. 80% of the visitors examine only the first search result page and about 9% click on the second page as well.

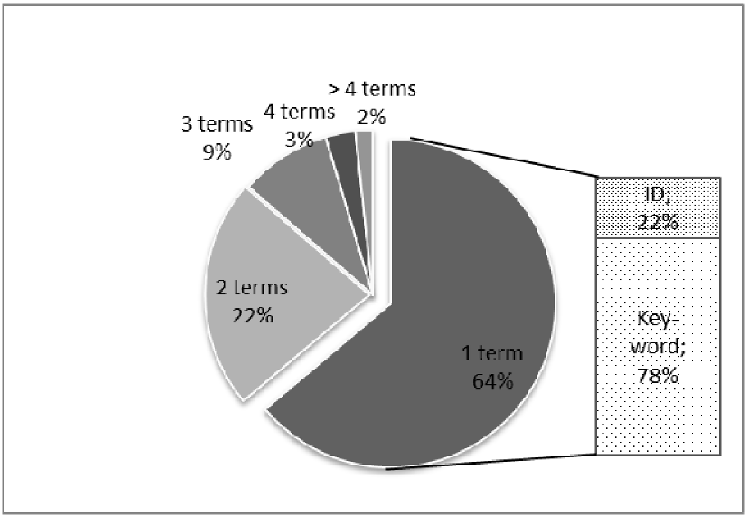

Fig. 1. Number of search terms

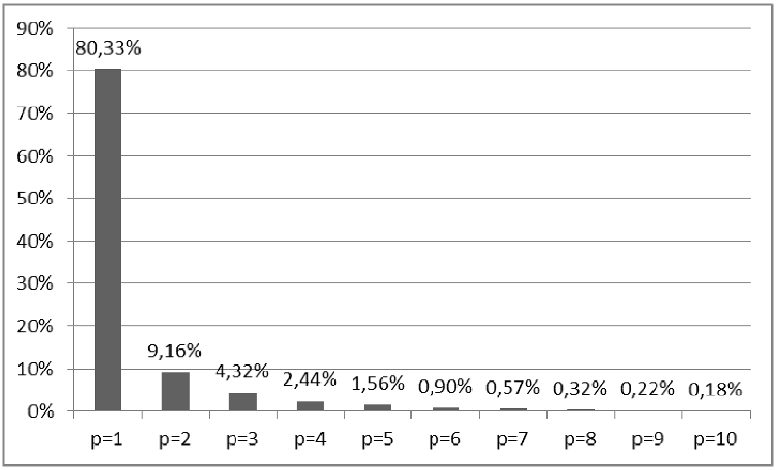

Fig. 2. Page views in search result page

The shop offers a “gallery” at the top of the search result page, where featured products are displayed and highlighted. This gallery contains 5 products, the rest of the page contains as the default setting 15 products (all following search result pages contain 20 products). The gallery (which contains 25% of the displayed products) generates 28% of the clicks, where products are put into the basket (resp. 33% unique clicks). Hence we can summarize that products displayed in a “featured area” are put in the basket more often than products from a list.

4.3 Specific Results

In the next step we compared two measures with the control group and the experimental group: clicks that put products into the basket (n = 13,981) and clicks that place the order (n = 8,876). Fig. 3 exhibits the differences between the clicks into the basket with the “spread on” and the “spread off” configuration. As mentioned before, we hypothesize that products that are devaluated with a spread-configuration (“spread on”) are put into the basket less often and less often ordered. However, as shown in Fig. 3, the differences are only marginal and not significant (tested with chi-square, e.g. spread 3: χ^2 = 0.0792). With spread 0,1 and 2 we actually found that the number of clicks into the basket was marginally higher than in the “spread on” configuration.

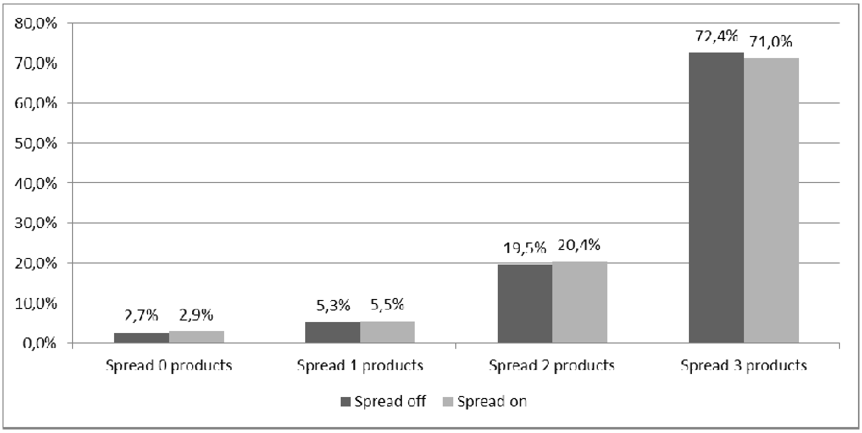

Fig. 3. Clicks into basket

As shown in Fig. 4 one can observe a similar situation; the differences are not very high, but with the exception of the “spread 0”-configuration - they are significant (χ^2<0,05). Again, with spread 0, 1 and 2 configuration we observe the opposite of our expectations.

In order to be able to go into more detail we split these data into the defined categories (file folder, text highlighter and correction products) and carried out the same evaluations. The detailed analyses show the same characteristics as in Figure 3 and 4.

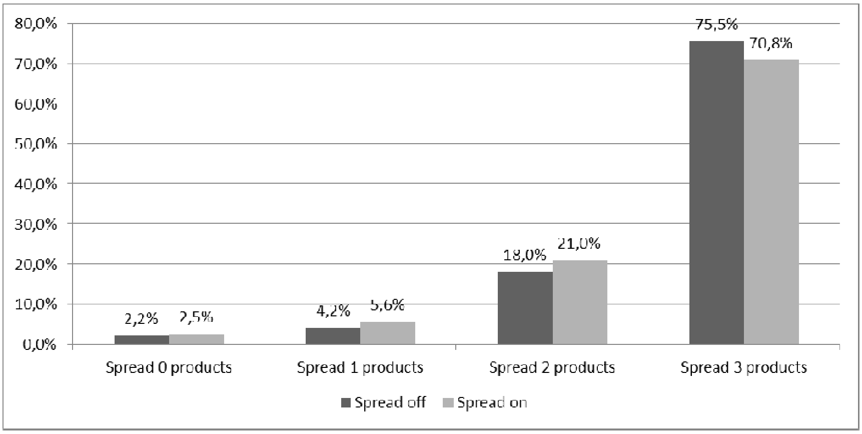

Fig. 4. Transactions

4.4 Discussion

The measurement setting still has some weaknesses. Unfortunately, we are not able to measure, estimate or exclude the possible biases, but the evaluation of random samples leads to the assumption that the impact of these biases is small. The measurement has a drawback: because of the b2b-environment (each customer has his own product selection) there is no technical possibility to determine by how many positions a product that has a spread configuration is ranked poorer.

The results indicate that higher devaluations (spread 3) have an impact on the number of clicks into the basket and the number of orders. The reason why products with a smaller devaluation (spread 0, 1 and 2) are ordered more often could be because the "unwanted" products (spread 3-configuration) are ordered less often and customers want or need to order alternative products.

These results were measured in a b2b-e-commerce-environment consisting of office materials and therefore certainly cannot be generalized to other scenarios.

5 Conclusion and Further Research

This paper discusses a method to measure the impact of variations of the rankings of the search result page of a web shop. The measurement method was applied to an e-commerce shop containing office products in the b2b-area. The results show that small variations of the positions of the research result page have no significant impact on clicks-to-basket or placed orders; more extensive variations of the positions have an impact on clicks-to-basket or placed orders – at least in the case of office materials and a b2b-e-commerce environment.

Further research work should be conducted on: (i) application of the measurement method in a b2c-scenario, which is less complex, (ii) in order to gain a deeper understanding of customers' behavior the measurement could be supplemented with usability-studies.

References

1. Goldman Sachs, Global Ecommerce to Accelerate (2013)
2. Wei, D.: Nothing But Net - Asia, Internet Investment Guide (2013)
3. Oracle (2013) B2C E-Commerce Trends for 2013
4. Oracle (2013) 2013 B2B Commerce Trends
5. Fredriksson, T.: E-commerce and Development. Key Trends and Issues. In: Workshop on E-Commerce, Development and SMEs (2013)
6. Stahl, E., Wittmann, G., Krabichler, T., et al.: E-Commerce-Leitfaden, 3rd edn. Noch erfolgreicher im elektronischen Handel.
7. Chopra, P.: The Ultimate Guide To A/B Testing (2010), http://www.smashingmagazine.com/2010/06/24/the-ultimate-guide-to-a-b-testing/ (accessed January 29, 2014)
8. Clifton, B.: Advanced Web metrics with Google Analytics, 3rd edn. Sybex serious skills. John Wiley & Sons, Indianapolis (2012)
9. Benbunan-Fich, R.: Using protocol analysis to evaluate the usability of a commercial web site. Information & Management 39(2), 151–163 (2001)
10. Chen, S.Y., Macredie, R.D.: The assessment of usability of electronic shopping: A heuristic evaluation. International Journal of Information Management 25(6), 516–532 (2005)
11. Lee, S., Koubek, R.J.: The effects of usability and web design attributes on user preference for e-commerce web sites. Computers in Industry 61(4), 329–341 (2010)
12. Deng, L., Poole, M.S.: Aesthetic design of e-commerce web pages – Webpage Complexity, Order and preference. Electronic Commerce Research and Applications 11(4), 420–440 (2012)
13. Salehi, F., Abdollahbeigi, B., Langroudi, A.C., et al.: The Impact of Website Information Convenience on E-commerce Success of Companies. Procedia - Social and Behavioral Sciences 57, 381–387 (2012)
14. Srinivasan, S.S., Anderson, R., Ponnavolu, K.: Customer loyalty in e-commerce: an exploration of its antecedents and consequences. Journal of Retailing 78(1), 41–50 (2002)
15. Hillier, M.: The role of cultural context in multilingual website usability. Electronic Commerce Research and Applications 2(1), 2–14 (2003)
16. McKnight, H.D., Choudhury, V., Kacmar, C.: Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. Information Systems Research 13(3), 334–359 (2002)
17. Gefen, D., Straub, D.W.: Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. Omega 32(6), 407–424 (2004)
18. Cui, G., Lui, H., Guo, X.: The Effect of Online Consumer Reviews on New Product Sales. International Journal of Electronic Commerce 17(1), 39–58 (2012)
19. Gretzel, U., Yoo, K.: Use and Impact of Online Travel Reviews. In: O'Connor, P., Höpken, W., Gretzel, U. (eds.) Information and Communication Technologies in Tourism, pp. 35–46. Springer, Vienna (2008)
20. Vermeulen, I.E., Seegers, D.: Tried and tested: The impact of online hotel reviews on consumer consideration. Tourism Management 30(1), 123–127 (2009)
21. Chitika, Inc., The Value of Google Result Positioning (2013), http://chitika.com/google-positioning-value (accessed January 28, 2014)