

Effects of Auto-Suggest on the Usability of Search in eCommerce

*Katharina C. Furtner, Thomas Mandl,
Christa Womser-Hacker*

Department of Information Science & Natural Language Processing
University of Hildesheim, Germany
mandl@uni-hildesheim.de

Abstract

The following study analyses the effects of Auto-Suggest as implemented on mobile Websites on user's search in m-commerce-shops. A benchmark showed that many design guidelines regarding the implementation of Auto-Suggest have been fulfilled. For usability tests, different designs of Auto-Suggest have been systematically selected. The results showed that the Auto-Suggest function guarantees a faster interaction with the website for most of the tasks on average. For tasks which require more time, the additional interaction time is around 10% and, surprisingly, does not result from more interaction steps. Many qualitative findings explain how interacting with Auto-Suggest can be troublesome for users.

Keywords: Information seeking, Auto-Suggest, Auto-Complete

1 Motivation and background

The mobile Web and the constantly rising number of smartphones for all age groups change the behaviour of many users regarding communication, information enquiry and consumer spending. In order to be successful, sites

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 178–190.

need to engage in continuously improving the user experience and supporting search in eCommerce. Auto-Suggest and Auto-Complete are, among others, important value added functions for simplifying the query construction process for users. However, these additional functions also require more knowledge for the user and interaction steps which were previously unnecessary. Our study intends to identify optimal design solutions for Auto-Suggest and to measure their effect on the usability.

Few studies have been conducted so far from a usability point of view. Our research design follows the steps listed below. The state of the art analysis revealed several guidelines for the implementation of Auto-Suggest on mobile m-commerce websites. Altogether, 17 guidelines have been suggested and empirically justified in advance. Dellinger (2013) showed usability-problems with the implementation of Auto-Suggest on mobile websites. A benchmark for 23 websites showed that many of these guidelines are already widespread but not all of them are realized by all sites. We identified three diverse implementations of Auto-Suggest and conducted usability tests to find out how they affect the usability for users.

2 State of the art

Published research on Auto-Suggest and AutoCompletion is still limited. Most papers deal with the Information Retrieval aspects and the effective implementation of these functions. Jiang et al. (2014) showed that the context as observed in previous queries can support the algorithm in finding better suggestions. Li et al. (2014) analyzed a log from a commercial search engine and observed that suggestions low on the list received very few clicks.

Several authors have presented guidelines for designing Auto-Suggest (Quirnbach 2012, Lewandowski & Quirnbach 2013, Wilson 2011). In an overview study, Dellinger (2013) collected these and other guidelines for Auto-Suggest from a usability perspective and empirically validated them in a preliminary qualitative user test. On the basis of both positive and negative findings, he revealed guidelines for Auto-Suggest on mobile websites. The tests used 12 websites with 15 test persons, but the test persons had not to perform tasks on each of the websites.

The following table lists the guidelines for the implementation of Auto-Suggest on mobile websites (cf. Dellinger 2014).

Table 1: List of Auto-Suggest guidelines

#1	Auto-Suggest should not have a title.
#2	All relevant suggestions must be displayed.
#3	Suggestions which are not clearly related to the search term should be avoided.
#4	The fewer suggestions are shown, the less effort for the user (cognitive and scrolling), but several suggestions may be positive if there are similar products.
#5	Font size and clickable areas should be designed large enough so that users can read and accurately select them at any time.
#6	A match between query and Auto-Suggest is important – the suggestions should contain the search term.
#7	The suggestion corresponding best to the search term should be displayed first.
#8	Suggestions which correspond to the search term should be shown on top of a list of suggestions.
#9	At least a part of a list of automatic suggestions should be displayed at the visible area of the touchscreen.
#10	If a list is partly hidden by a delivered keyboard, this has to be conveyed to the user.
#11	It can be helpful to show the number of expected results for similar results (e.g. in brackets on the right of the suggestion).
#12	It is helpful for the user to mark the category to which the suggestion is related.
#13	The wording of the suggestion should speak for itself – there must not be any additional description.
#14	Symbols, flags or the like should be used only if they give further value.
#15	The possibility of deleting the whole search term with only one click is helpful for some users (icon “x” on the right of the input box).
#16	Geolocation can be placed on top of the list of suggestions if this has further value regarding the results.
#17	The optional function of adding a suggestion to the search term without directly starting the search makes it possible for the user to search in detail and typing less.

Little is known from published research about the quantitative aspects of usability for Auto-Suggest. We intended to measure how much time and typing amount could be reduced by Auto-Suggest. How many queries are needed with and without the use of Auto-Suggest to complete the tasks?

3 Methodology

Our study examined these guidelines by using quantitative methods. Usability-tests are most suitable for this kind of elicitation of data. Apart from video recording, the thinking aloud method is used. The test persons are asked to express their consideration and thoughts during their working on the tasks. Before and after the test as well as after each task, the test persons are interviewed concerning their individual usage of smartphones and the Internet and about the variations of Auto-Suggest occurring during the test. For these interviews, a questionnaire has been designed. As Auto-Suggest is a very small and special part of the system, no standardized questionnaire can be applied.

3.1 Analysis of the context of use of Auto-Suggest

The analysis of the context of use is important to determine properties of the test persons and test environment. First, it is to say that mobile users use websites in another context as desktop-users. There is no single mobile context, but several types in which users interact (Maurice, 2012: 7). Furthermore, the mobile context can be everywhere; most of the interactions within the mobile context are situated in a very dynamic and unpredictable environment (Hinman, 2012: 41).

In a study conducted by Google, three types of mobile users are named (Google, b): The group of users called *Repetitive now* comprises users who carry out repetitive actions by searching for the same information again and again (e.g. the weather, updates of status in social networks). The second group, *Bored now*, consists of users who have some time available and want to use it e.g. to get through waiting at the bus stop or at the café (Maurice, 2012: 7). Users of the third group *Urgent now* however want to carry out their tasks as soon as possible with their mobile device. These tasks are often

associated with the situation in which they are. Online-shop-users likely correspond most with the user type Bored now – users who have some time to rummage or search in mobile online-shops – may it be at home or on the way. The analysis of the context is important with regard to the drawing up of the scenarios used in the usability tests as they should be embedded in typical, everyday situations for the users.

3.2 Benchmark and selection of the web sites for the study

A list of 42 m-commerce-websites which has been drawn up within a study by Furtner and Pölderl (2013) for the investigation of navigational elements on m-commerce-websites was the basis for selecting the websites. In the preliminary study by Furtner and Pölderl, those online-shops were chosen of a list of the largest German online-shops 2012 Internet World Business which had implemented a mobile website and showed the characteristics of an online-shop. In a benchmark, these websites were examined for the fulfilment of the guidelines for Auto-Suggest. #4 and #16 were left out of account as they could not be examined as “fulfilled” or “not fulfilled”. #2 was sampled. #5 was fulfilled even when the size of clickable buttons fell below the recommended size of at least 1 cm × 1 cm. Most of the buttons on the websites had a height of circa 0.5 cm, #5 was not fulfilled when the size of buttons was below this value or when the text was too large for the screen.

The result of the benchmark was rather positive. On average, 9.7 out of 15 guidelines were fulfilled on the websites. On the other hand, only 23 of 42 online-shops had an Auto-Suggest function – this shows that many companies have a lot to catch up. Even though on some of the websites most of the guidelines were fulfilled, the Auto-Suggest function was not usable as the suggestions often were not clickable or lead nowhere. Another problem was that the suggestions were not clickable separately or imposed themselves on the user as he could use only the suggestions for the search (there was no button to search for his own queries).

For this study, websites which showed the largest differences regarding the guidelines for Auto-Suggest were chosen as objects of investigation. Hence, ebay.de with 13 fulfilled guidelines and zooplus.de with 7 fulfilled guidelines were chosen. The test scenarios would be taken on two websites for a direct comparison, so that one additional website was chosen which should show largest differences to ebay.de regarding the number of fulfilled

guidelines but still should have diverse products like ebay.de. Weltbild.de was then chosen.

3.3 Interview and test guideline

A test guideline with tasks and questioning was designed to define the course of the individual tests so that it was the same for each of the tests. Five tasks were formulated in which the test person had to search for a product on two websites one after another. The Tasks were embedded in everyday situations. As the tasks concerning the free-text search were formulated very general, no background knowledge on the part of the test persons was needed. Also, the concrete aim of the usability-evaluation – namely the evaluation of the Auto-Suggest-function – was hidden from the test persons until they had performed all tasks. Two versions of the test guideline, which differ only in the order of the retrieved websites, were drawn up, so that the comparison could be made between the websites and not between the test persons, furthermore the learning effects of the test persons could be taken into consideration. In a pre-test the test persons were asked about their individual usage of smart-phones, the mobile Internet and m-commerce. After the task, the test person was asked how difficult or easy the task was. Besides that, the test person could comment on the task. Having finished the last task, the test person was questioned on Auto-Suggest during the test and about their everyday usage of Auto-Suggest.

3.4 Test persons and test environment

If possible, the test persons should be real users so that they are as representative as possible for the group. In the context of this work, 20 test persons of different age groups were recruited. 13 of them were female and 7 of them were male. 16 test persons were aged under the age of 26 (Digital Natives) and 4 test persons belonged to the group aged between 26 and 49, so that they neither belonged to the Digital Natives nor to the Silver Surfers.

The usability-tests took place at the test person's house and took at the most 30 minutes. The equipment – a mobile usability-labor – consisted of a laptop, two cameras, a microphone and a smartphone. The test took place under realistic conditions. The screen of the smartphone and user input were taped by a camera which has been attached on the table. At the beginning of

the test session, the smartphone was placed in front of the user on the table, but he could also pick it up during the test. A smaller camera which had a microphone integrated taped the user's expressions and comments. The usability suite MORAE was used during the test to record audio, video and display data. The tests were carried out on the Smartphone XPERIA P (Model LT22i), Android-version 4.1.2 (Jelly Bean) and Chrome. After each test session, cookies and cache of the smartphone were deleted to make sure that stored data would not influence the next test.

4 Results and discussion

As some experience in the use of smartphones and the mobile Internet was a precondition for the recruitment of test persons, all of them owned a smartphone and used it for online-activities. 80% of the test persons said they would use the mobile Internet several times a day, 15% said they would use it every day and 5% several times a week.

Particularly younger test persons belonging to the group of Digital natives declared to use it for Social Media while the older people declared to use it for Messaging, general information and smaller enquiries. 45% of the test persons already had experience with m-Commerce.

4.1 Post-test

When the test persons were asked if they had seen Auto-Suggest already before the test, all of them declared to have known and used Auto-Suggest already before and would continue using it. Likewise, all of the test persons could remember at least one website with Auto-Suggest implemented. In that connection most frequently called were the websites by Amazon, Google, Ebay, Youtube, Zalando and Facebook.

The test persons found the Auto-Suggest-function to be helpful on all of the used websites, as the suggestions appeared mostly after having typed only a few characters. The displaying of the category to which a product belonged was found to be helpful as well as the fact that the suggestions could give inspiration. Zooplus.de was estimated as being well-structured. On all used websites as negative was found that not all relevant or irrelevant sug-

gestions were displayed. Some test persons had problems with the spreading of the products on ebay.de. In general, the test persons emphasized the profitability of Auto-Suggest to avoid spelling mistakes and reduce typing amount. Irrelevant suggestions were found to be very annoying.

4.2 Quantitative results

A first approach of measuring the saving of time by the use of Auto-Suggest was to measure the time users needed before and after the first use of Auto-Suggest. The problem with this approach is that the users already knew Auto-Suggest before the test and thus used it very different, not from a certain time on. Besides that, the level of difficulty differed from task to task so that it is not possible to compare them with regard to the time needed. For that reason, we compared for each task the average time needed by users who used Auto-Suggest with the average time needed of users who did not use it.

Whereas for two tasks, the users took slightly longer, the two first tasks took 20% respectively 45% less time. Auto-Suggest can save much time in cases when the list of suggestions contains good candidates and the user does not have to scroll a lot. We can interpret Auto-Suggest as an intermediate step to first retrieve the query term before starting the actual retrieval of documents. When this first step is successful, then the user can be highly effective by selecting a suggested candidate term and continue by saving much typing time. The cost for this efficiency is that the user has to learn a new function and interrupt or at least slow down his typing activity for judging the candidates for completion. Instead of continuing to type the user needs to switch to a cognitive activity, read the candidates and switch to another activity. He needs to change from the typing mode to the clicking mode. Obviously, making a decision about the switching and carrying it out, takes cognitive effort and time. Only if the saved effort is larger than the additional effort, the value added function Auto-Suggest will lead to higher efficiency overall.

Table 2: Task completion time

Task	Time needed using AS	Time needed without AS
1	38.24 sec	47.76 sec
2	No suggestions	32.7 sec
3	28.23 sec	52.15 sec
4	34.19 sec	30.07 sec
5	47.78 sec	46.15 sec

In contrast, if suggestions are not relevant, reading and judging them costs the user some time. She can be distracted or has to scroll a lot in the list of suggestions. If the user selects a misleading candidate or even makes a usage error while working with the Auto-Suggest, time will be lost. If the user switches cognitively to the clicking mode and takes away his fingers from the keyboard and cannot profit from the Auto-Suggest, even more time is lost.

The reduction of typing effort is measured the same way as the saving of time, so by the average amount of typing. Each keystroke and click was counted as an action. In all tasks, the test persons saved keystrokes and clicks by the use of Auto-Suggest. Some of the test persons commented on this effect of Auto-Suggest already during the test sessions.

Table 3: Key Strokes

Task	Amount of typing using AS	Amount of typing without AS
1	18	22.14
2	No suggestions	17.48
3	11.62	21.9
4	11.85	15.07
5	14.44	19.73

Interestingly and contrary to the first finding, the users did not need more key strokes even for tasks which required more time with Auto-Suggest.

For all tasks and on all sites, the test persons needed one or two queries regardless of the use of Auto-Suggest. No differences were observed between the use of Auto-Suggest and manual query input.

Table 4: Number of Queries

Task	Using AS	Without using AS
1	1.17	1.15
2	No suggestions	1.03
3	1.05	1.29
4	1.19	1.2
5	1.22	1.18

In the following, selected qualitative observations are given. They give hints on the design problems of Auto-Suggest but do not fully explain the quantitative outcomes. Further results can be found in Furtner (2014).

4.3 Overcoming uncertainty

The tests showed that Auto-Suggest can help overcoming uncertainty and avoid spelling mistakes. In one of the tasks given to the users, there was a spelling mistake in the product the user had to search for. Most of the users who chose a suggestion did not even notice the spelling mistake as they were given the right product in the list of suggestions.

4.4 Auto-Suggest can introduce new ideas

During the search for the products, some test persons got inspired by Auto-Suggest and were given new ideas. The search term which is entered by the user can also be expanded by the suggestions.

4.5 Implementation of Auto-Suggest

The test persons had different explanations on the implementation of Auto-Suggest. One of the most frequently uttered explanations was that the suggestions are based on the most frequently bought or searched products.

4.6 Acceptance and use of Auto-Suggest

Except from one test person, all of them used Auto-Suggest during the test sessions at least once. 7 of them chose the first suggestion immediately (without looking at the other suggestions), 12 test persons looked at the list before choosing an entry. All of the test persons knew and had used Auto-Suggest already before the test.

4.7 Completeness of suggestions

30% of the test persons thought that they would leave a website without Auto-Suggest. 70% of them said that they would continue searching on the website even when there would be no Auto-Suggest. During the tests, the users showed a completely different behaviour: 90% of the test persons hesitated typing when there was no Auto-Suggest which shows that Auto-Suggest is present in the user's mind and also expected. Therefore, when there is no Auto-Suggest or when the list of suggestions is incomplete or con-

tains irrelevant suggestions, this can lead to uncertainty, loss of time and higher amount of typing.

4.8 Number of displayed suggestions

#4 (of the guidelines from table 1) can be verified as it is a fact that the more suggestions are displayed, the more interaction this means for the user. During the test sessions, one of the test persons even adapted the search term to get fewer suggestions. The completeness of the suggestions (#2 and #3) must be taken into consideration.

4.9 Matching of search term and suggestion

The test results showed that it can lead to usability-problems and uncertainty of the user if the search term does not match the suggestion.

4.10 Order of suggestions

The order of the suggestions in a list is of importance. If irrelevant suggestions appear at the top of the list, this and the fact that users rarely scroll the list of suggestions can lead to a heavy usability-problem: 90% of the test persons would leave the website in this instance. #7 and #8 thus can be verified.

4.11 Scrolling the list of suggestions

Only 55% of the test persons scrolled at least once during the test sessions, which means that 45% may not have noticed suggestions which were hidden under the virtual keyboard. In connection with the order of suggestions and a mobile context, this can lead to usability-problems. On weltbild.de, scrolling was not possible as there were no suggestions hidden under the keyboard. Some test persons nevertheless tried to scroll the list because this fact was not clear for them. #9 and #10 (from table 1) can be verified.

4.12 Appearing and disappearing suggestions

The appearing and disappearing of suggestions can lead to heavily usability-problems when the users see the suggestions but miss the time to click on them and then have to type the whole search term because the suggestions have disappeared.

4.13 Input box should be placed on the homepage

Some of the test persons claimed it to be a negative aspect when the input box was not placed on the homepage. This led to additional effort.

5 Conclusion and future plans

The test results showed that Auto-Suggest supports the user on m-Commerce-websites as users know and use Auto-Suggest and many of the guidelines are already implemented. The results showed as well that a good implementation of Auto-Suggest can mean a saving of time and typing for the user. Nevertheless, for some tasks, Auto-Suggest leads to a longer interaction time. Auto-Suggest can help to avoid uncertainty and spelling mistakes. Beyond that users can even get inspiration and new ideas from Auto-Suggest.

Auto-Suggest is also a very recent function. A lot of change takes place during the optimization of mobile websites. Even during this study, some websites changed their function completely. Currently, eye-tracking experiments are developed to find out exactly how users save or lose time when interacting with Auto-Suggest.

References

- Callender, Jeffrey; Morville, Peter (2010): *Search patterns*. O'Reilly.
- Dellinger, Kai F. (2013): *Usability von Auto-Suggest: Benchmark bestehender Ansätze und deren Evaluierung im mobilen Kontext*. University of Hildesheim, MA Thesis.

- Dellinger, Kai F. (2014): Usability von Auto-Suggest. "UP14 – German Usability Professionals". <http://www.germanupa.de/events/usability-professionals-up14/shortpaper/usability-von-autosuggest.html> <26.12.2014>.
- Furtner, Katharina (2014): *AutoSuggest auf mobilen Websites: Evaluierung anhand von M-Commerce-Angeboten*. BA Thesis. University of Hildesheim.
- Furtner, Katharina; Pölderl, Friederike (2013): Verwendung von Navigationselementen auf mobilen Webseiten im Bereich des E- bzw. M-Commerce. University of Hildesheim, project report.
- Google: Think Insights. <http://www.thinkwithgoogle-com/insights/emea/featured/latest-insights-for-mobile> <4.3.2014>.
- Hinman, Rachel (2012): *The mobile frontier: A Guide For Designing Mobile Experiences*. Brooklyn, NY: Rosenfeld Media.
- Jyun-Yu Jiang, Yen-Yu Ke, Pao-Yu Chien, & Pu-Jen Cheng (2014): Learning user reformulation behavior for query auto-completion. In: *Proc. 37th Intl. ACM SIGIR conference*. ACM, New York, NY, USA, pp. 445–454.
- Krannich, Dennis (2010): *Mobile System Design: Herausforderungen, Anforderungen und Lösungsansätze für Design, Implementierung und Usability-Testing Mobiler Systeme*. Books on Demand.
- Lewandowski, Dirk, & Quirnbach, Sonja Monika (2013): Suchvorschläge während der Eingabe. In: D. Lewandowski (Ed.). *Handbuch Internet-Suchmaschinen 3: Suchmaschinen zwischen Technik und Gesellschaft*. Berlin: Akademische Verlagsgesellschaft AKA, pp. 273–298.
- Maurice, Florence (2012): *Mobile Webseiten: Strategien, Techniken, Dos und Don'ts für Webentwickler*. München: Carl Hanser.
- Quirnbach, Sonja Monika (2012): *Suchmaschinen: User Experience, Usability und nutzerzentrierte Website-Gestaltung*. Springer.
- Russell-Rose, Tony; Tate, Tyler (2013): *Designing the search experience: The information architecture of discovery*. Amsterdam: Morgan Kaufmann.
- Wilson, Max (2012): *Search User Interface Design*. San Rafael, Calif.: Morgan & Claypool Publishers.
- Yanen Li, Anlei Dong, Hongning Wang, Hongbo Deng, Yi Chang, & Cheng Xiang Zhai (2014): A two-dimensional click model for query auto-completion. In: *Proc. 37th Intl. ACM SIGIR conference*. ACM, NY, USA, pp. 455–464.