Product Information Retrieval on the Web: An Empirical Study

Sabri Bouzidi, Damir Vandic, Flavius Frasincar, and Uzay Kaymak

Erasmus University Rotterdam
Erasmus School of Economics
PO Box 1738, NL-3000 DR, Rotterdam, the Netherlands
307228sb@student.eur.nl, vandic@ese.eur.nl, frasincar@ese.eur.nl

Abstract. In this paper, we investigate the consumers' perception of on-line product search using a questionnaire-based survey. We identify that the information retrieval activity of the purchase process can be performed with three Web applications: a search engine, a price comparison service, and a Web shop. The study underlines the need for linked product data as proposed by the Semantic Web. We argue that linked data will result in easier product search on the Web for the consumer.

1 Introduction

The Web makes it easier for consumers to find information on products compared to traditional product information sources. More and more users search on-line for products [14], whether it is to make a purchase, obtain information, or for pleasure [4]. Without the Web, it has become almost impossible to properly orient yourself for a product purchase. Everything one needs can be found on-line, usually with lots of information and against competitive prices. This luxury also brings some issues with it. For example, from which of the dozens stores will one order, should one visit a particular store, use a price comparison site, or use Google, and what should one do to explore the available products.

The most common activity on the Web is searching for information [8]. Besides search types related to fact-finding and research-based searches, there is also a significant share of search sessions which are related to product search. On-line shoppers are faced with a multitude of choices: search engines (e.g., Google, Yahoo!, Bing, etc.), price comparators (e.g., PriceGrabber, BizRate, Google Shopping, etc.), and Web shops (e.g., Amazon, eBay, Books-A-Million, etc.). This study aims to understand how these searches take place, in order to identify the current bottlenecks. The hypothesis that we are testing in this paper is that the lack of linked data makes the current search for products suboptimal.

Each store and comparison service operates its own database with product information. Executing a sophisticated query against such a dataset often does not lead to desired results. The current lack of linked data for product information makes it difficult to compare and to find relevant products on the Web. A linked data approach to product information would allow one to navigate to

the desired information at Web-wide level [3]. In this way, it becomes easy to find missing or complementary information about a product. There are three possible benefits of this improvement:

1. The availability of the product information increases due to replicating and complementary product data;
2. It makes possible to use complex queries based on product metadata;
3. The search costs will decrease as it will cost less time and effort to locate products.

The paper is organized as follows. Section 2 discusses the conceptual framework for on-line product search and the potential benefit that linked data can bring for the Web activity. Section 3 describes the survey that we carried out with respect to the current product search experiences on the Web. Section 4 presents and discusses the results of the survey. Last, Sect. 5 gives our conclusion and identifies future work.

2 Conceptual Framework

This section presents our conceptual framework for product (information) search. First, we define information retrieval in the context of product (information) search. Then, we discuss the search process and the need specificity of a consumer. Then, the Semantic Web data model for products is conceptualized. The identification of the need for linked data for product search will be examined by considering the way customers search for products on the current Web.

2.1 Information retrieval

On-line shopping can be divided into three stages: information retrieval, information transfer, and the actual product purchase [15]. In this study we focus on the first stage, because at this stage the consumer is searching for product information. This is important for us, because the consumer can then specify whether the search task is difficult or easy. An overview of the purchase process is presented in Fig. 1. The first stage, the information retrieval activity, represents basic data exchange between any product related Web site and consumers. The last two stages are not considered in this study, because the consumers in these stages already found their desired product. This study solely focuses on how consumers find products on-line, as outlined in the previous section.

Information retrieval tools We argue that the search outcomes strongly depend on the used tool. The activities in the information retrieval stage can be performed with several tools. There are three types of tools available: a search engine, a price comparison service, and a Web shop. When it comes to general information search, the most common way to do this is by using a search engine [16]. Looking at the most visited websites we find a majority of search

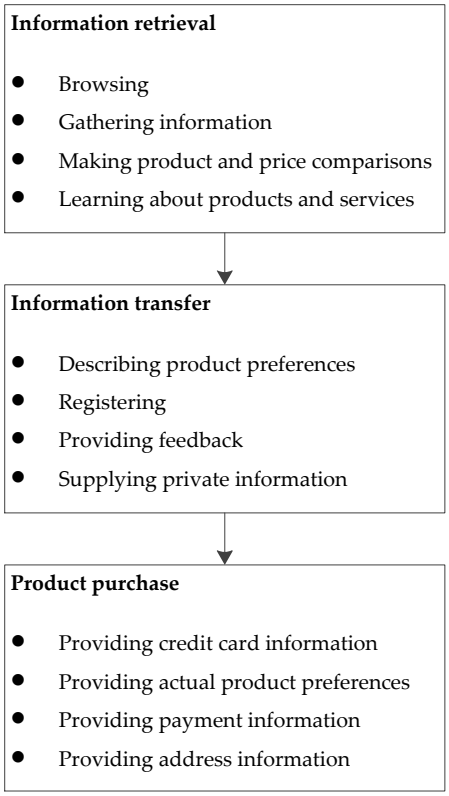

Fig. 1. On-line purchase process, adopted from [15]. This is the entire process for making purchases on the Web. Our study focuses on the first stage. This is because at stages 2 and 3, consumers already found their desired product and Web shop, and thus will no longer search for product information and Web shops.

engines, e.g., Google, Yahoo!, Bing, etc. [9]. Besides search engines, consumers also use the so-called price comparison services. A price comparison service enables individuals to find the many prices that exist for specific products [19]. An example of such a comparison service is Google Shopping, which is used worldwide, and Kieskeurig.nl, which is used in the Netherlands. Also, consumers can go directly to a Web shop, e.g., BestBuy. Most web shops aim to build a relationship with a customer, which should result in customers that prefer a particular Web store and come back for another purchase [13].

Search strategies The search strategy is dependent on the purpose of the search. When one knows what their desired product is, it will require a different way of searching than when the desired product is not known. This is also referred to as need specificity. The task type that is related to a low need specificity (i.e., we do not know what we are looking for) is known as product search. Prod-

uct search aims to find a product candidate that was previously unknown [10]. In contrast, the task type that is related to high need specificity (i.e., we exactly know what we are looking for) is known as consumer information search. Consumer information search aims to find product information [10]. This distinction has important implications for the search process, because it affects consumer behaviour. In case of high need specificity, the consumer is more likely to use non-value-added search mechanisms [13], which is defined as objective information which is publicity available. For example, when searching for a smart-phone, this can be the brand and model. In case of low need specificity, the consumer is more likely to use value-added search mechanisms, this is defined as subjective information which is not publicity available. In the smart-phones example, this would include product reviews. When consumers have high need specificity, they would like to go directly to the product without having to go through subjective information. In contrast, consumers with low need specificity would like the subjective information, since that can help them better define their needs.

On-line information search has often been studied using micro-economic and decision-making theory. Accordingly, previous studies have determined that people search for information taking into account the trade-off between costs (e.g., time and effort) required to obtain relevant information, and the benefits obtained from the search outcome [1]. Furthermore, it is has been shown that individual factors, such as experience and cognitive skills, have a significant impact on the on-line search behaviour of a user [20].

In [7] the distinction between searching and browsing is examined. The research results suggest that searching requires more detailed product information, while browsing places greater emphasis on information from a retailer. Furthermore, consumers spend more time on searching when the product is higher priced, more visible, and more complex [2]. In addition, consumers search differently because of individual factors, such as the perceived benefits of search (e.g., enjoyment, self-confidence, etc.), demographic aspects, and product knowledge possessed [2].

In order to identify the motives of searching for products, we distinguish between pre-purchase search and ongoing search [4]. The ongoing searchers are consumers that gather product information, even if they do not plan to buy anything. The motives for this activity are to experience pleasure, and to obtain information for future use. The outcome of ongoing search is an increased product and market knowledge, which leads to less time and effort needed for searching, and an increased impulse of buying products [4]. Pre-purchase searchers are consumers who search in order to make a better purchase decision. The main outcome of the pre-purchase activity is also an increased product and market knowledge, and an increased satisfaction with the found product.

In Fig. 2 the factors that affect the consumers' search strategy are shown. The human factors adopted from [8], task types adopted from [10], and the motives adopted from [4] are important factors that affect the search strategies. The focus of this study is on the relation between the task type and the used

tool, with respect to the difficulty of finding products in the perception of the consumer.

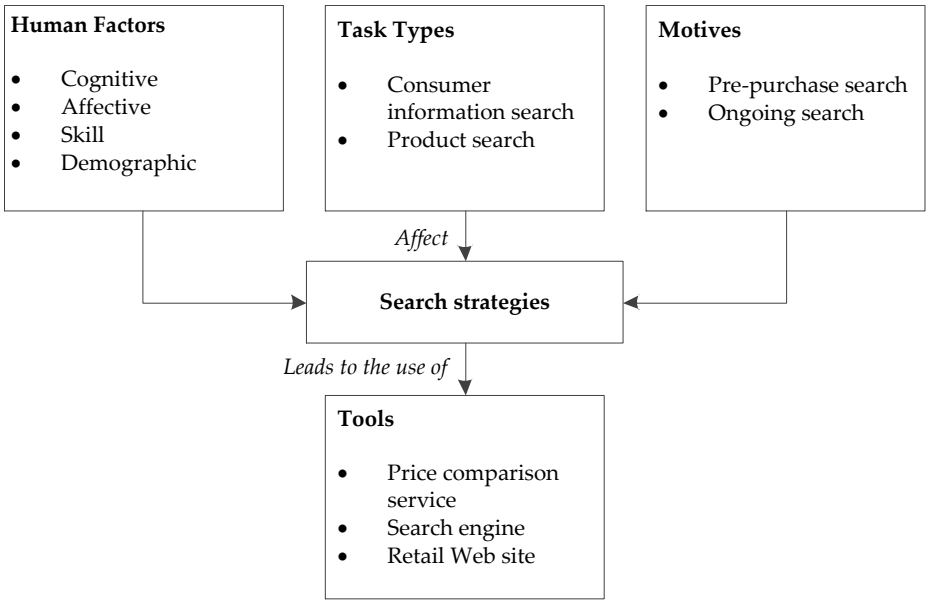

Fig. 2. Factors that affect the search strategies. The human factors are adopted from [8], the task types are adopted from [10], and the motives are adopted from [4]. We argue that the search strategy leads to the use of different tools.

2.2 Semantic Web: Linking product data

The Semantic Web makes it possible to find information based on its meaning, not just its text [5]. When employing semantics for products, consumers can use a system that can find products with more ease. Such a system can also determine which products relate to each other. An important issue the Semantic technology can cover is the information provision, and the ability to deal with missing and complementary information. The Semantic Web has been developed to cope with incomplete, possibly inconsistent information, and product data is no exception. Price comparison services are early examples of the Semantic Web [18]. Most comparison services get their data from data feeds generated by retailers. Such a service only centralizes data; it does not link this data, making it difficult to use queries through product-related information. Linking product data with semantics is yet in an early stage. To our knowledge, the only system that links product data is ProductDB [6]. ProductDB claims to create a page for every product in the world, for the purpose of connecting underlying

structured data into one interlinked dataset. However, the project seems to have been discontinued.

We believe that using linked data for products can have benefits for consumers. The quality of the product information increases when the data is linked, as missing and complementary information can be corrected. Also, the use of linked data makes possible to use sophisticated queries, as the user has access to product metadata over large product information datasets. The advantage of linked data is that it becomes easier to find the right information. The search costs will decrease for the consumer, as it will cost less time and effort to locate information. Linked data can facilitate aggregators of product information by means of a shared product ontology [17]. A popular high level ontology to describe product information is GoodRelations [11]. This ontology has been refined for various product domains. GoodRelations is being used by more than 10,000 small and large shops world-wide, including Google, Yahoo!, and Best Buy [11].

3 Empirical Study

We conduct an empirical study to investigate the perception of consumers with respect to product search on the Web. The results of the survey are aimed at underlining the need for linked product data. The survey has the following objectives:

1. To determine the ease-of-search for product information, with low need specificity;
2. To determine the ease-of-search for product information, with high need specificity;
3. To determine the ease-of-search for Web shops;
4. To examine the relationship between the ease-of-search and the used information retrieval tool.

3.1 Product Classes Studied

The primary criteria for choosing products are their availability on the Web. In addition, the products must belong to 'search goods'. Search goods are defined as goods for which the most important attributes for assessing product quality are discoverable without the consumer interacting with the product [12]. Also, the price of the products must not be too low, since consumers engage in more search when the price is higher [2]. Based on these considerations, three product classes were chosen:

- Clothing;
- Smart-phones;
- Vacations.

These product groups have the similarity that information, and sale points are available on the Web. They have attributes which can be compared with each

other. Linked data can be of value to these product groups, because of the large amount of information currently unorganized for these products on the Web. With linked data this information can be extracted and presented to the user by making use of queries.

3.2 Participants

In our research we use a field experiment. In order to control the sample we contacted responders. Participants could only participate if they declare that they have consulted the Web to search for the three product groups for their latest purchases. Eventually 32 responses are obtained. The demographic statistics are shown in in Table 1. Most responders are male, between the ages of 19 and 35, and indicated to have a professional education.

Table 1. Participants. This table contains the characteristics of the participants.

Gender		Age			Education		
Female	Male	0-18	19-35	36-50	High school	Professional	University
9	23	4	22	6	8	16	8

In order to determine the sample power, we perform a power analysis. This analysis is done with the software package IBM SPSS SamplePower. The results of the analysis show that with a sample size of 32 subjects, the means of the answers to the questions have a margin error of plus/minus 0.25 points from the true value. In computing the sample size we assume that we want to be 95% certain that the observed value falls within the margin of error.

3.3 Evaluation measures

In order to measure our research objectives, we used an excerpt from the questionnaire presented in [15]. The questions from the 'perceived ease-of-use' are adapted, so that the search process on Web-wide level is examined for the three product groups. For the questionnaire details we refer to the Appendix. We ask the respondents how easy they locate product information in case of high and low need specificity. Also, how easy it is to find Web shops in case of high need specificity. A five-point Likert-scale is used, where the item 'strongly disagree' means that it is very hard to locate a product, and the item 'strongly agree' means that it is very easy to locate a product. The second measurement examines which of information retrieval tools are used for each product group. A five-point Likert-scale is used, where the item 'strongly disagree' means that the tool is never used, and the item 'strongly agree' means that the tool is always used.

The reliability of the measures is tested by using the Cronbach's alpha method. With IBM SPSS Statistics, nine variables are combined, i.e., ease-of-search for information with high and low need specificity, and ease-of-search for Web shops, for the three product groups. The result shows that Cronbach's Alpha is 0.94, therefore the measures are reliable. The measures are assumed also to be valid, because they are adapted from other studies.

4 Analysis and Results

Table 2 shows the descriptive statistics of the items used in this study. The means shows that for all three product groups, finding information is perceived as an easy-to-medium task, since the means are between 3 and 4. Finding information that has high need specificity is perceived as slightly easier than low need specificity, but the standard deviation is higher. This can be explained by the fact, that with high need specificity there are better terms on which the user can search. For example, entering a specific product in a search engine often gives good search outcomes, however, the user has to master the search engine query language well. Overall, finding Web shops is perceived as an easier task than finding information, since the means on locating stores are higher than the means of locating information.

Table 2. Descriptive statistics. This table contains the means and standard deviations of the responses to the questions about the ease-of-search and the use of information retrieval tools. The scale used is from 1 to 5, with 1 being 'very hard to find' or 'never used', and 5 being 'very easy to find' or 'always used'.

	Clothing		Smart-phones		Vacation	
	μ	σ	μ	σ	μ	σ
Ease of search						
Product information (low need specificity)	3.31	1.03	3.41	0.98	3.28	1.08
Product information (high need specificity)	3.75	0.76	3.84	0.68	3.75	0.88
Web shops	3.81	0.78	3.91	0.69	3.78	0.79
Information retrieval tools used						
Search engines	4.31	0.59	4.25	0.72	4.38	0.55
Price comparison services	2.41	0.84	3.25	1.22	2.44	0.98
Web shops	3.56	1.16	2.69	1.09	3.06	1.22

Based on our study, search engines are most commonly used as product information tool. The price comparison service for clothing and vacations are

less used than for smart-phones. This can be explained by the fact that smart-phones are more present on comparison services than the other groups, which are usually less present. The Web shop is used more for clothing and vacations, and less often for smart-phones. However, the average is close to 'neutral' for Web shops. When ease-of-search has 4 points or higher, then the search task is defined as easy. As shown in Table 2, the means of locating information and locating shops is lower than 4. This implies that the first part of our hypothesis is supported, that is that the current search of products on the Web is suboptimal.

Table 3 shows significant Pearson correlations between search engines, and both the ease-of-search for Web shops, as well as product information. Price comparison services are correlated with the search for smart-phones; this applies both for information as to Web shops. As for the use of Web shops for information, we observe that this is only correlated with vacations. This can be explained by the fact that vacations are not as common products as clothes and smart-phones, so for convenience users prefer the travelling agency Web site.

Table 3. Correlations between ease-of-search and product information retrieval tools. The correlations are calculated using the Pearson method in the software package IBM SPSS Statistics.

Product group	Perceived ease of search	Search engine	Price comparison	Web shop
Clothing	Product information (low)	0.522**	0.147	0.28
	Product information (high)	0.536**	-0.038	0.091
	Web shops	0.480**	-0.028	-
Smart-phones	Product information (low)	0.493**	0.561**	0.244
	Product information (high)	0.547**	0.518**	0.063
	Web shops	0.505**	0.528**	-
Vacations	Product information (low)	0.517**	-0.028	0.451**
	Product information (high)	0.464**	-0.243	0.377**
	Web shops	0.414*	-0.205	-

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

In this section, we have analysed the outcomes of the survey, first by comparing the means of the answers to the questions, followed by the computations of the correlations between the ease of search and the information retrieval tools.

The analysis shows that the subjects experience searching for products and product information on the Web as an easy-to-medium task. Smart-phones seem to be less difficult to find than clothing and vacations. We find this an indication of a good potential for linked product data, especially for the latter categories. In conclusion, we see opportunities for linked product data, because in the perception of consumers finding products and product information on the Web is not as easy as it can be with linked product data.

5 Conclusion

This study examines the way consumers search on the Web for products. We identified that the information retrieval activity can be performed using three tools: search engines, price comparison services, and Web shops. The tools are used in the search strategy of consumers in order to browse products, gather information, and make product and price comparisons. In addition, we discussed literature that investigates the factors that affect the search strategy of a consumer, which are human factors, task types, and motives.

With our empirical study we found that locating products and Web shops is perceived by the consumers as an easy-to-medium task. With regard to the product groups, the smart-phones are found with more ease than vacations and clothing. Smart-phones is the only group that has a significant correlation with price comparison services. The other groups, vacations and clothing, often lack of a comparison service, therefore these groups would benefit most from linked data. Linked product data can support consumers to make the task of searching product information on the Web more efficient, allowing navigation over Wide-wide product information. The perceived ease-of-search will improve, when more parties link their data, such as retailers, manufactures, review sites, comparison services, etc.

Future studies can focus on possibilities to encourage third parties to make use of linked data. For example, one can investigate how an incentive can be created for third parties to join the project. Linked product data would also be beneficial for consumers, as they will have customizable access to product information and offers on a Web-wide level.

References

1. Alba, J., Lynch, J., Weitz, B., Janiszewski, C., Lutz, R., Sawyer, A., Wood, S.: Interactive home shopping: Consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. The Journal of Marketing 61(3), 38–53 (1997)
2. Beatty, S.E., Smith, S.M.: External search effort: An investigation across several product categories. Journal of Consumer Research 14(1), 83–95 (1987)
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. Int. J. Semantic Web Inf. Syst. 5(3), 1–22 (2009)
4. Bloch, P.H., Sherrell, D.L., Ridgway, N.M.: Consumer search: An extended framework. Journal of Consumer Research 13(1), 119–126 (1986)

5. Davies, J., Studer, R., Warren, P. (eds.): Semantic Web Technologies: Trends and Research in Ontology-Based Systems. Wiley, Chichester, UK (2006)
6. Davis, I.: Productdb (2010), http://productdb.org/, [Online; accessed 24-February-2012]
7. Detlor, B., Sproule, S., Gupta, C.: Pre-purchase online information seeking: Search versus browse. Journal of Electronic Commerce Research 4(2), 72–84 (2003)
8. Flavián-Blanco, C., Gurrea-Sarasa, R., Orús-Sanclemente, C.: Analyzing the emotional outcomes of the online search behavior with search engines. Comput. Hum. Behav. 27(1), 540–551 (2011)
9. Google: Doubleclick ad planner by google (2011), http://www.google.com/adplanner/static/top1000/, [Online; accessed 12-February-2012]
10. Häubl, G., Dellaert, B.G.C., Donkers, B.: Tunnel vision: Local behavioral influences on consumer decisions in product search. Marketing Science 29(3), 438–455 (2010)
11. Hepp, M.: Goodrelations— the most powerful web vocabulary for e-commerce (2011), http://www.heppnetz.de/projects/goodrelations/, [Online; accessed 16-February-2012]
12. Huang, P., Lurie, N.H., Mitra, S.: Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods. Journal of Marketing 73(2), 55–69 (2009)
13. Koufaris, M., Kambil, A., LaBarbera, P.A.: Consumer behavior in web-based commerce: An empirical study. Int. J. Electron. Commerce 6(2), 115–138 (2001)
14. Li, B., Ghose, A., Ipeirotis, P.G.: Towards a theory model for product search. In: Proceedings of the 20th international conference on World Wide Web WWW 2011. pp. 327–336. ACM, New York, NY, USA (2011)
15. Pavlou, P.A.: Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. Int. J. Electron. Commerce 7(3), 101–134 (2003)
16. Rangaswamy, A., Giles, C.L., Seres, S.: A strategic perspective on search engines: Thought candies for practitioners and researchers. Journal of Interactive Marketing 23(1), 49–60 (2009)
17. Vandic, D., Van Dam, J., Frasincar, F.: Faceted product search powered by the semantic web. Decision Support Systems (2012), http://dx.doi.org/10.1016/j.dss.2012.02.010
18. Wikipedia: Comparison shopping agent— Wikipedia, the free encyclopedia (2012), http://en.wikipedia.org/wiki/Comparison_shopping_agent, [Online; accessed 24-February-2012]
19. Wikipedia: Price comparison service— Wikipedia, the free encyclopedia (2012), http://en.wikipedia.org/wiki/Price_comparison_service, [Online; accessed 12-February-2012]
20. Yee, P.L., Hsieh-Yee, I., Pierce, G.R., Grome, R., Schantz, L.: Self-evaluative intrusive thoughts impede successful searching on the internet. Computers in Human Behavior 20(1), 85–101 (2004)

A Questionnaire

The table below contains the questions from the survey. For questions 1 to 6, for each product group the subjects indicate whether they agree with the statement on a five-point Likert-scale, with 1 being 'strongly disagree', 3 being 'neutral', and 5 represents 'strongly agree'. The questions 1 to 3 refer to the ease-of-search, while questions of 4 to 6 relate to the use of the information retrieval tools. Questions 7 to 9 are intended for the identification of characteristics of the respondents.

Questions	Variable
1. When I don't know exactly what to buy, I find it easy to locate the relevant product information when searching for:	Clothing (locate product information, low need specificity)
	Smart-phones (locate product information, low need specificity)
	Vacations (locate product information, low need specificity)
2. When I do know what product to buy, I find it easy to locate the relevant product information when searching for:	Clothing (locate product information, high need specificity)
	Smart-phones (locate product information, high need specificity)
	Vacations (locate product information, high need specificity)
3. When I do know what product to buy, I find it easy to locate the shops that sell it when searching for:	Clothing (locate shops)
	Smart-phones (locate shops)
	Vacations (locate shops)
4. I use a search engine (e.g., Google) when searching for:	Clothing (search engine)
	Smart-phones (search engine)
	Vacations (search engine)
5. I use a price comparison service (e.g. Google Shopping) when searching for :	Clothing (price comparison service)
	Smart-phones (price comparison service)
	Vacations (price comparison service)
6. I use a specific Web store that I know of when searching for:	Clothing (Web shop)
	Smart-phones (Web shop)
	Vacations (Web shop)
7. Your gender	Gender
8. Your age	Age
9. Your education level	Education