Measuring Web Searching Task Efficiency: An Imprecise Data Envelopment Analysis Approach

Štefan Lyócsa[a], Cyril Závadský[a], Tomáš Výrost[a], Eduard Baumöhl[a]*

[a] University of Economics, Faculty of Business Economics, Tajovského 13, 040 01 Košice, Slovakia

[*] Corresponding author.
e-mail address: stefan[dot]lyocsa[at]gmail[dot]com
tel. no.: +421 55 722 31 11
fax. no.: +421 55 623 06 20

Abstract

In this paper we propose an application of Imprecise Data Envelopment Analysis model (IDEA) in the setting of measuring web searching task efficiency. The presented approach had been applied to a sample of 337 students within two years, conducting a web-searching experiment consisting of domain-specific and domain-neutral search tasks in both single and multiple correct answer scenarios. We discuss the benefits of using DEA as a performance measure in experiments involving the respondent in performing multiple tasks, as well as the extension to imprecise DEA to account for the ordinal nature of some of the results. Finally, we present some preliminary results on efficiency of web searching.

Keywords: web searching; web searching efficiency; data envelopment analysis; imprecise data envelopment analysis; web searching task types

1. Introduction

The goal of the presented paper is to advance our understanding in measuring web searching performance. In our approach, we perceive web searchers as being the so called Decision Making Units (DMUs), who spend different inputs (e.g., time), to produce multiple outputs (e.g., information or a measure of success in getting the information searched on the web). The Data Envelopment Analysis (DEA) is used to measure performance within a set of DMUs. According to Cooper et al. (2004, p. 3): "*A DMU is to be rated as fully (100%) efficient on the basis of available evidence if and only if the performances of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs.*" In this study, we define web searching efficiency with regard to a given set of inputs and outputs as a ratio of outputs to inputs. Since the seminal work of Charnes et al. (1978), DEA has been used for studies on measuring efficiency of hospitals, cities, banks, or even countries. This study contributes to the literature on web searching by using a new approach to measure web searching task efficiency, but also to the literature on DEA as we apply the methodology to a new set of problems. We also utilize a more recent DEA model developed by Zhu (2003).

Given the large number of studies conducted on web searching, there are many possible approaches to measure the search performance. In Table 1, we include a brief overview of the variables that can be used as performance measures. As the studies differed significantly both in objectives and methodology, so did the measures that reflected their focus. The common choices include the success rate in completing tasks (e.g., Lazonder et al., 2000) and the total time spent completing tasks (e.g., Tsai et al., 2011). Other measures included the number of links found, the number of keywords used and the number of pages visited. We refer to Terai et al. (2008), who analysed eye movement data while performing web searches, as an example of the diversity of approaches.

Using log files, videotapes and verbal descriptions of the searching behaviour of nurses, Jenkins et al. (2003) found that domain and web experts had better search performance than others. Experts required less time to locate the necessary information, were more confident, conducted more in-depth searches and were less disoriented. Although similar web searching efficiency was observed between neuroscientists and life scientists, domain knowledge was shown to be the discriminating factor that differentiated the behaviour of participants in Vibert et al. (2009). Corredor's study (2006) also showed that prior domain and general knowledge influences searching behaviour. For studying domain knowledge, the sample used by Jenkins et al. (2003) proved beneficial, as nurses can be more easily classified according to their domain expertise. In our rather homogenous sample of students, it was much more difficult to identify domain-specific knowledge experts. Therefore, we created both kinds of tasks: those that lie outside the expected domains of expertise of our participants (e.g., genetics, mechanics), as well as more general tasks.

2. Research purposes

This study aims to propose a new approach for measuring web searching task efficiency utilising the Data Envelopment Analysis (DEA) framework and to examine the differences between web searching efficiency scores for various task types. DEA is a nonparametric method used to evaluate production efficiency among several entities, called DMUs (Decision Making Units, i.e., participants). It allows for multiple inputs and outputs. For our purposes, DEA was attractive for two reasons. First, it is possible to use it on ordinal data. Second, it produces an overall measure of efficiency, based on all four tasks given to each participant without having to arbitrarily choose weights for individual questions as this is done endogenously by DEA.

3. Methodology

3.1 The Imprecise Data Envelopment Analysis Model

To calculate the performance of web searching conducted by the participants of our experiment, we employed DEA to provide the overall efficiency measure. The inputs represent the time spent to complete each task, and the outputs represent the validity of the answer for *EN* (domain-neutral/one answer) and *ES* (domain-specific/one answer) and rank among the test group for *ON* (domain-neutral/more correct answers) and *OS* (domain-specific/more correct answers). The efficiency is expressed as a single number between 0 and 1 and is interpreted as a percentage of the efficiency of the selected DMU compared to the optimal ones. These are identified endogenously; thus, the performance measure for a specific participant depends both on individual and group performances, which serve as a basis for comparison. Technically, DEA is usually transformed into a linear programming problem.

In our case, the basic DEA model could not be used, as our outputs were ordinal variables. This is mostly notable in the case of *ON* and *OS*, in which we posed questions with more than one possible answer. For example, when evaluating the question where participants were asked to find the lowest price for a certain book, taking the price found for output is not very meaningful, as it can only take discrete values corresponding to the prices in bookstores in Slovakia. As DEA assumes that the variables are on interval/ratio scale, the use of data that is inherently ordinal would violate one of the model's assumptions. We have therefore interpreted the results as ranks based on the prices found by the respondents. To account for the ordinal nature of the data, we have employed Zhu's procedure (2003), using the following proposed model:

$$\min \theta_o \qquad (1)$$

subject to

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_o x_{io} \qquad (2)$$

$$\sum_{j \neq 0} \lambda_j \underline{y}_{rj} + \lambda_o \bar{y}_{ro} \geq \theta_o \bar{y}_{ro}, \qquad r \in BO \qquad (3)$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq \theta_o y_{ro}, \qquad r \in BO \qquad (4)$$

$$\lambda_j \geq 0, \qquad j = 1, 2, \ldots, n \qquad (5)$$

Here, θ_o is the efficiency of the selected DMU*o* (the participant), x_{ij} is the amount of input *i* used by participant *j*, y_{rj} is the output *r* achieved by participant *j* and *n* is the number of participants. The set *BO* consists of indices for outputs *r* that are imprecise (bounded, ratio bounded, or – as in our case – ordinal).

As can be seen above, in the case of outputs not belonging to *BO*, the data are used directly. When the outputs are imprecise, Zhu (2003) proposed a transformation of data, which allows the standard DEA procedure to be used. In our case, this means setting y_{rj} to $\underline{y}_{rj} = 0$ for all participants *j* whose rank in output *r* is lower than the rank of the analysed participant *o*, and setting it to $\underline{y}_{rj} = 1$ for the rest (that is, the value is set to 0 for all participants who were worse, and 1 for those who were at least as good as *o* when considering output *r*).

3.2 Participants

Our research was conducted during two years (January 2010 and January 2011). The participants included students from a public university in the Slovak Republic (namely the Faculty of Business Economics, University of Economics in Bratislava) finishing their first semester. The 2010 group included 202 students (151 females and 51 males), and the 2011 group included 136 students (90 females and 46 males). The average age was 19.3 (2010) and 19.4 (2011). The participation of students in this study was obligatory as it was part of their course in Introductory Informatics. Students participating in 2010 had no prior information about the scientific goals, or about the specific web searching tasks performed during the experiment other than the initial instructions, which we describe in more detail below. Although our instructions to the students remained the same in 2011, we cannot rule out that some students may have previously learned from others that they would be performing web searching tasks.

3.3 Data collection and analysis

The questionnaires and web searching tasks were carried out using an e-learning web browser-based module familiar to all students, as it was used in courses they had already taken. At the beginning of the study, we screened a short video in which we presented the details of the experiment, including instructions for the questionnaires and web searching tasks. After the presentation, students were asked to pose questions in case they felt the instructions they received were not thoroughly understood.

First, participants filled out several short questionnaires covering various topics, such as socio-demographics, familiarity with a PC and Internet self-efficiency (with very similar questions[1] as those used in Tsai and Tsai 2003, p. 44). All students used the same mainstream operating system, and their desktops showed only shortcuts to seven web browsers (allowing them to choose their favourite), with no browsing history stored. Following these short questionnaires (in most cases, it took less than 10 min to complete all of them), they were asked to complete four searching tasks. For each year, two groups were formed, with each group having different web searching tasks. Students were randomly assigned to these groups (Test A and Test B in 2010, and Test C and Test D in 2011). The order of tasks was again randomised. Each participant performed four searching tasks as described in the next section. After the searching tasks were finished, they completed one more questionnaire, the Online Information Searching Strategy Inventory (OISSI) of Tsai (2009). Including the presentation at the beginning, the whole procedure took approximately 50 to 60 min.

[1] Although better instruments have been developed (Tsai and Tsai, 2010), Internet self-efficiency was not the primary goal of this study. In addition, we did not want to overwhelm participants with excessive (or extensive) questionnaires.

3.4 Searching tasks

Like Navarro-Prieto et al. (1999) and Thatcher (2008), we specified four tasks for each test. The sample sizes for these tests were as follows: Test A n = 100, and Test B n = 102, Test C n = 74 and Test D n = 62. Technically, Test A and Test B were performed under the same conditions, as were Test C and Test D. The only difference was in web searching tasks. Therefore, we initially expected similar results for both pairs.

The searching tasks were divided into:

- Domain-specific searching tasks, which were designed to include a highly specialised topic, which we assumed to be unfamiliar to the participants so they would have no "prior knowledge" that would help them to accomplish the task more successfully.
- Domain-neutral searching tasks, which included general topics.
- Web searching tasks with one or more correct answers, where participants were required to provide only one answer. If they succeeded, their outcome was coded as 1, and if not, it was coded as 0.
- Web searching tasks with a) more correct answers of which participants were asked to give as many as possible (listing features of…) or b) several not equally correct answers of which participants were asked to give the best they could locate (lowest price of a book)[2]. The answers were recorded and for each participant a corresponding rank was assigned according to the a) number of listed answers, or b) closeness of the answer to the best answer reported.

In each test following combinations of web searching tasks were used (with examples below):

- *Domain-neutral / one answer (EN)*

  "Phillips's curve describes the relationship between inflation and unemployment. The author W. Phillips published his research paper at the end of the 60's. Your task is to find the title of this paper."
- *Domain-neutral / more correct answers (ON)*

  "You want to buy the book "*Language of Mathematics*" written by Keith Devlin and published in 2002. Your main criterion is price, and you want to buy it online from a

[2] Our stratification of task types was similar to those used by Navarro-Prieto et al. (1999), who employed fact-finding tasks (which correspond to our task requesting one answer) and exploratory tasks (in which more than one correct answer was possible).

Slovak web site (Slovak domain). Your task is to find the book with the lowest price. (Ignore the shipping costs and indicate the price in €, including VAT).”

- *Domain-specific / one answer (ES)*

  “Who was the first physician to describe the process of blood circulation?”

- *Domain-specific / more correct answers (OS)*

  “Find as many shortcomings of the Wankel engine as possible.”

Questions were posed in Slovak, the only language used at the faculty where the study was conducted. In addition to the answers themselves, students had to upload the links for web pages they used to provide them. Answers were evaluated by the two researchers who designed the searching tasks. The links uploaded by students were also verified. Students had 8 minutes to complete each search task. This is considerably less time when compared to most studies. However, prior to the study the searching tasks were performed by four other users and have showed 8 minutes to be sufficient for completing each task.

Two types of data were recorded for each task and participant: (1) the number of correct answers and (2) time spent. Time was an important component of our study as it allowed us to discriminate between users with an equal number of correct answers. One important note should be reported at this point. The *EN* task in Test A was too difficult, as only one participant answered it correctly. This might slightly distort our results.

Direct comparison of outcomes from different tests is not sensible. Therefore, we calculated the efficiency of completing web searching tasks in each test. The final sample size used for measuring web searching tasks efficiencies were as follows: Test A = 91, Test B = 99, Test C = 50 and Test D = 56.

Table 1 An overview of measures used in selected web search-oriented studies

Lazonder et al. (2000)	• Percentage of successfully completed tasks • Performance time • Performance efficiency, defined as a ratio of the number of completed task s to the time needed to complete them • Performance effectiveness, defines as a number of actions needed to successfully complete a task
Stronge et al. (2006)	• Four categories, depending on whether the participant provided a correct answer, incorrect answer, quit or ran out of time.
Chevalier and Kicka (2006)	• Time necessary to find information • Number of steps/hyperlinks visited • Amount of cognitive resources used (measured by the dual task technique) • Participants' usability satisfaction with regard to the visited site
Terai et al. (2008)	• Number of web pages browsed • Reading time • Eye movement
Thatcher (2008)	• Time to completion • Number of search steps • Number of queries • Number of terms per query
Tu et al. (2008)	• Number of keywords • Number of visited pages • Maximum depth of exploration • Average number of words per keyword • Number of words used in first keyword
Ford et al. (2009)	• Perceived search difficulty (as reported by the participant on a scale of 1 to 5)
Gwizdka and Lopatovska (2009)	• Number of pages visited • Number of visits to first search result pages (without revisits) • Number of content pages visited, including within page navigation, without revisits • Number of bookmarked individual results pages • Ratio of revisits to web pages • Linearity of navigation path • Total time on each task
Willoughby et al. (2009)	• Number of times search engines accessed • Number of terms used in search • Number of first links • Number of seconds links • Number of third and subsequent links • Number of relevant/good sites • Number of fair sites • Number of irrelevant sites • Time spent on irrelevant sites
Tsai et al. (2011)	• Number of keywords • Maximum depth of exploration • Total depth of exploration (sum of all levels during searching) • Number of revisited web pages • Total time for browsing result pages • Total time for browsing pages for less than seven seconds • Total time for browsing pages for more than seven seconds

4. Results

Table 2 indicates the number of valid observations by web searching task outcomes per test, by searching task with number of correct answers for the *EN* and *ES* tasks, and average time spent on each task. For tasks *OS* and *ON*, the range corresponds to the ordinal variables assigned to each answer. (For example, the lowest price of a book from among 15 different prices was assigned number 15; the next was assigned number 14, and so on. We proceeded similarly in cases where one needed to list as many features as possible).

Table 2 Descriptive statistics of web searching outcomes

		TEST A (n = 91)	**TEST B** (n = 99)	**TEST C** (n = 50)	**TEST D** (n = 56)
TYPE *EN*	**success**	n = 1, (0.010 %)	n = 17, (17.17 %)	n = 38, (76.00 %)	n = 19, (33.93 %)
	average time	393.97 s.	418.72 s.	240.54 s.	359.10 s.
TYPE *ES*	**success**	n = 14, (15.38 %)	n = 39, (39.39 %)	n = 30, (60.00 %)	n = 38, (67.86 %)
	average time	423.94 s.	392.86 s.	284.64 s.	309.46 s.
TYPE *OS*	**range**	0 - 7	0 - 5	4 - 29	1 - 6
	average time	406.18 s.	426.02 s.	285.66 s.	305.35 s.
TYPE *ON*	**range**	1 - 9	1 - 11	1 - 24	1 - 13
	average time	318.64 s.	289.62 s.	287.84 s.	353.92 s.

Note: EN corresponds to domain-neutral/one answer tasks, ES to domain-specific/one answer tasks, ON to domain-neutral tasks with more, ordered answers, OS to domain-specific tasks with more, ordered answers.

Note, that the maximum time for task completion was set to 480 seconds (8 minutes). The average time of participants suggests that this limit was sufficient. Table 3 presents average efficiency and standard error of efficiency for each test and task type. Within each test, we observe some differences. For example in test B the average efficiency for the task *EN* was only 0.26 compared to the *OS* task efficiency of 0.44. Similar differences were found in test D and smaller in test A and test C. However, from methodological point of view it is difficult to compare task types in general, because the differences in efficiencies may be attributed not only to a characteristic of a task, but also to an inherent difficulty of the task itself. These two effects are coupled and it is difficult, if not impossible, to discriminate between them. Therefore, one should interpret such results with great caution.

Table 3 Descriptive statistics of web searching efficiency for each test and task type

test A	*E*	*O*	test B	*E*	*O*
S	0.40 (0.016)	0.49 (0.020)	*S*	0.43 (0.016)	0.44 (0.015)
N	0.48 (0.011)	0.46 (0.021)	*N*	0.26 (0.020)	0.41 (0.019)
test C	*E*	*O*	**test D**	*E*	*O*
S	0.43 (0.031)	0.38 (0.038)	*S*	0.47 (0.026)	0.54 (0.027)
N	0.45 (0.029)	0.56 (0.031)	*N*	0.25 (0.018)	0.54 (0.024)
overall efficiency		**test A**	**test B**	**test C**	**test D**
		0.88 (0.013)	0.82 (0.018)	0.86 (0.027)	0.87 (0.016)

Note: Standard error of efficiency in parenthesis

5. Conclusions and discussion

The limitations of this study are twofold. First, as our sample consists of a specific group of students, generalizations to a wider population should be drawn with caution. Second, it is not possible to attribute differences in web searching efficiency solely to differences of web searching task types. The difficulty of tasks itself may be different as well. More importantly, it is not sensible to compare efficiencies calculated from different sets of inputs, outputs and DMUs. Future studies should be therefore designed to minimize such effect. One possible solution might be to prepare 8 – 12 web searching tasks for each task type. Then for each participant, 2 – 4 web searching tasks should be assigned each task type i.e., 8 – 16 web searching tasks per participant together. Of course, such a study would be more time consuming.

We believe that the new measure of web searching task outcomes might contribute to the growing literature on web searching. The analysis of inputs and outputs of the model may be enhanced by other variables as well. Researchers can experiment with various setups. More importantly, this measure offers possibilities for evaluating the web searching outcomes of many different web searching tasks. However, it also has its limitations. One important limitation is that removing one participant from the efficiency calculation sample can influence the efficiency scores of other participants. Thus, results from such a sample are not comparable to any other sample. We used time as an input variable. Partitioning time into components (e.g., Tsai et al., 2011) would likely be more advantageous.

References

Charnes, A., Cooper, W. W., Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(*6*), 429-444.

Chevalier, A., & Kicka, M. (2006). Web designers and web users: Influence of the ergonomic quality of the web site on the information search. *International Journal of Human-Computer Studies*, *64*(*10*), 1031-1048.

Cooper, W. W., Seiford, L. M., Zhu, J. (2004). Data Envelopment Analysis: Models and interpretations. In Cooper, W. W., Seiford, L. M., Zhu, J. (Ed.), *Handbook on Data Envelopment Analysis* (pp. 1-39). Boston: Kluwer Academic Publisher.

Corredor, J. (2006). General and domain-specific influence of prior knowledge on setting of goals and content use in museum websites. *Computers and Education*, *47*(*2*), 207-221.

Ford, N., Eaglestone, B., Madden, A., & Whittle, M. (2009). Web searching by the "general public": an individual differences perspective. *Journal of Documentation*, *65*(*4*), 632-667.

Gwizdka, J., & Lopatovska, I. (2009). The role of subjective factors in the information search process. Journal of the American Society for Information Science and Technology, *60*(*12*), 2452-2464.

Jenkins, C., Corritore, C. L., & Weidenbeck, S. (2003). Patterns of information seeking on the Web: A qualitative study of domain expertise and Web expertise. *IT & Society, 1*(*3*), 66-89.

Lazonder, A. W., Biemans, H. J. A, & Wopereis, I. G. J. H. (2000). Differences between Novice and Experienced Users in Searching Information on the World Wide Web. *Journal of the American Society for Information Science*. *51*(*6*), 576-581.

Navarro-Prieto, R., Scaife, M., & Rogers, Y. (1999). Cognitive strategies in web searching. (2011). (online). http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/.

Stronge, A. J., Rogers, W. A., & Fisk, A. D. (2006). Web-Based Information Search and Retrieval: Effects of Strategy Use and Age on Search Success. *Human Factors*, *48*(*3*), 434-446.

Terai, H., Saito, H., Egusa, Y., Takaku, M., Miwa, M., & Kando, N. (2008). Differences between informational and transactional tasks in information seeking on the web. *Proceedings of the second international symposium on Information interaction in context*, 152-159.

Thatcher, A. (2008). Web search strategies: The influence of web experience and task type. *Information Processing and Management*, *44*(*3*), 1308-1329.

Tsai, M.-J., & Tsai, C.-C. (2003). Information searching strategies in web-based science learning: The role of Internet self-efficacy. *Innovations in Education & Teaching International*, *40*(*1*), 43-50.

Tsai, M.-J., & Tsai, C.-C. (2010). Junior high school students' Internet usage and self-efficacy: A re-examination of the gender gap. *Computers and Education*, *54*(*4*), 1182-1192.

Tsai, M.-J., Hsu, C.-Y., & Tsai, C.-C. (2011). Investigation of High School Student's Online Science Information Searching Performance: The Role of Implicit and Explicit Strategies. *Journal of Science Education and Technology*, Doi: 10.1007/s10956-011-9307-2

Tu, Y.-W., Shih, M., & Tsai, C.-C. (2008). Eight grader's web searching strategies and outcomes: The role of task types, web experiences and epistemological beliefs. *Computers and Education*, *51*(*3*), 1142-1153.

Vibert, N., Ros, C., Le Bigot, L., Ramond, M., Gatefin, J., & Rouet, J.-F. (2009). Effects of domain knowledge on reference search with the PubMed database: An experimental study. *Journal of the American Society for Information Science and Technology*, *60*(*7*), 1423-1447.

Willoughby, T., Anderson, S. A., Wood, E., Mueller, J., & Ross, C. (2008). Fast searching for information on the Internet to use in a learning context: The impact of domain knowledge. *Computers and Education*, *52*(*3*), 640-648.

Zhu, J. (2003). Imprecise data envelopment analysis (IDEA): A review and improvement with an application. *European Journal of Operational Research*, *144*(*3*), 513-529.