Tailoring and the Efficiency of Information Seeking

Nathalie Colineau and Cécile Paris

CSIRO - ICT Centre
Locked Bag 17, North Ryde NSW 1670, Australia
{nathalie.colineau,cecile.paris}@csiro.au

Abstract. We present an empirical study assessing the impact of tailoring on information seeking tasks. Our aim was to evaluate whether providing tailored information would help people find the information they need more quickly and more accurately. Our results show that tailored documents have an impact on information seeking, at least when the information to be found is spread over a number of sources and needs to be synthesised.

1 Introduction

With the increasing amount of available information, finding what one needs is not an easy task. Tailoring information to people's needs seems like an effective way to deliver information to people (e.g., [4], [3], [1], [5]). There have been experiments to test whether users prefer information customisation over generic information (e.g., [6], [7]), or whether tailored information leads to more behavioural changes when information is meant to influence behaviour (e.g., [2]). However, the impact of tailoring on the efficiency of information seeking has not been tested.

This is the specific question we addressed in this work: Is tailoring effective in an information seeking task? Our aim was to evaluate whether providing tailored information would help people find the information they needed, allowing them to find the information more quickly and more accurately than when using generic information. We exploited one of our applications, SciFly [8], a system that automatically generates brochures tailored to the user's stated interests (i.e., their query). The availability of both the generic (existing) brochures and the SciFly tailored brochures allowed us to compare the effectiveness of tailored brochures over generic brochures in information seeking tasks. We describe our experimental setting and our results.

2 Impact of Tailoring Content on an Information Seeking Task

The objective was to assess the usefulness of tailoring in answering people's information needs, especially in situations when the topic of interest is spread over multiple documents, necessitating users to consult more than one document to get the information. (Note that this is what currently happens when someone asks for a brochure about one or more topics, at least at CSIRO, as there cannot be a brochure for every possible combination of topics a customer may have.) We wanted to assess whether

C. Conati, K. McCoy, and G. Paliouras (Eds.): UM 2007, LNAI 4511, pp. 440–444, 2007.

producing a tailored brochure would help people get to the information they needed more quickly and more accurately.

2.1 Experimental Design

We did a two-group post-test only randomised experiment. Subjects were given brochures about a topic of research undertaken at CSIRO and asked to answer a set of factoid questions[1] (e.g., who is the project leader of T1?). Some questions required browsing the entire document to find the answers, others looking for a specific paragraph. There were 4 questions and 12 facts to be found. The design (illustrated in Table 1) was done such that each group (of 12 subjects) performed the task twice: once with a tailored brochure and once with a set of non-tailored brochures.

Table 1. Experiment setup

	Non-tailored	Tailored
Topic1: Image Analysis	Group A	Group B
Topic2: Environmental Informatics	Group B	Group A

For repeatability purposes, we chose two different topics with two sets of similar questions. This allowed us to verify that the results observed in the first task were not due to a group but to the type of brochure (i.e., tailored *vs*. non-tailored). In the group performing the task with the tailored material (our treatment group), subjects were given only one brochure, a brochure generated by Scifly, tailored to the chosen topic. In the group performing the task with the generic material, subjects were given the set of existing (generic) brochures that covered the topic. As a result, there was additional content, in particular, content related to other projects carried out in the division's laboratories. It is important to note, however, that the text fragments (paragraphs) used in the tailored brochures were the same as those used in the generic brochures; both were manually written by the communicators.

Our aim was to find out whether one group would perform the task better (more correct answers, fewer incorrect answers) and whether one group would complete the task faster. Our hypotheses were that the group with the tailored brochures would perform better and faster, as the other group had to search for the relevant facts in more material and had irrelevant information, and that the group with the generic brochures would retrieve more incorrect answers. We used the following measures:

- *Time performance*: We recorded the time spent to complete the task.
- *Recall of correct items of information*: We counted for each subject the number of correct items of information retrieved out of the 12 items to be found. We calculated the mean of correct items for each group and computed a t-Test. We calculated the recall rate by computing the proportion of correct answers retrieved by the subjects out of the 144 correct answers (12 correct answers for 12 subjects).
- *Precision of information retrieved*: We counted for each subject the number of incorrect items of information brought back instead of (or in addition to) the

[1] The questions had clear factual answers to avoid ambiguity in the judgment of correctness.

correct answers. (Sometimes, the subjects answered the question correctly but included additional irrelevant and thus incorrect information.) We calculated the mean of incorrect items of information retrieved for each group and computed a t-Test. We calculated the precision rate by taking the proportion of correct answers retrieved by the subjects (a maximum of 144) to all answers retrieved by the subjects (which could be more than 144, if they retrieved incorrect items as well as correct ones).

2.2 Results

We used the tools available at http://faculty.vassar.edu/lowry/VassarStats.html. Table 2 shows the results on time performance. They confirm the direction of our hypothesis: for both topics, the group performing with the non-tailored brochures took more time to perform the task than the group with the tailored brochures: most subjects needed between 400 and 600 seconds to complete the task with the tailored brochure, while they needed between 100 and 200 seconds more with the non-tailored brochures. The difference observed is statistically significant.

Table 2. t-Test results for the time performance (mean time for each group, in seconds)

	Non-tailored brochures		Tailored brochures	
Topic 1	Group A	**626.5**	Group B	**409.75**
	t	**+3.83**	df	22
	P one-tailed		**0.000456**	

Topic 2	Group B	**602.08**	Group A	**483.5**
	t	**+1.76**	df	22
	P one-tailed		**0.0461535**	

Table 3. t-Test results for the number of correct items of information retrieved (mean)

	Non-tailored brochures		Tailored brochures	
Topic 1	Group A	**10.5**	Group B	**10.41**

Topic 2	Group B	**8.16**	Group A	**10.5**
	t	**+2.85**	df	22
	P one-tailed		**0.0046565**	

As shown in Table 3, when the task was performed the first time (i.e., on topic 1), there was hardly any difference on the number of correct items of information retrieved. The difference in mean did not confirm the direction of our hypothesis, as group A (with generic brochures) retrieved more correct items on average than group B (with tailored brochure). The difference is tiny (overall, one more correct item was found), but enough to reject our hypothesis. Thus, there was no need to perform the t-Test to determine the level of confidence. Considering their closeness in performance, we consider that the two groups performed equally on that measure.

When performing the task for the second time (i.e., on topic 2), however, the difference was statistically significant (at and beyond 0.005). Considering that the subjects were less familiar with topic 2 (which we know from a questionnaire), the customised brochure may have provided a real difference here, reducing considerably the search space and facilitating the retrieval of information. Note that the recall performance for the group working with the tailored brochure is constant across topics, averaging the retrieval of 10 items out of 12. In contrast, the recall performance for the group working with the non-tailored brochures dropped significantly from an average of 10 to 8 correct items retrieved. (See Table 5.)

Table 4 shows the number of incorrect items of information retrieved[2]. In both tasks, the group performing with the tailored brochure performed better, bringing back less irrelevant content. However, this is statistically significant in the first case only. We have no explanation as to why this is the case. We did notice that, in most cases, it was not so much people answering the questions wrongly, but more people including additional (irrelevant) information.

Table 4. t-Test results for the number of incorrect items (mean)

	Non-tailored brochures		Tailored brochures	
Topic 1	Group A	**1.75**	Group B	**0**
	t	**+3.78**	df	22
	P one-tailed		**0.000515**	

Topic 2	Group B	**2.91**	Group A	**1.83**
	t	**+1.05**	df	22
	P one-tailed		**0.152558**	

Table 5. Recall rates, precision rates and F-measures

	Non-tailored brochures		Tailored brochures	
Topic 1	Recall	**87.5%**	Recall	**86.7%**
	Precision	**85.7%**	Precision	**100%**
	F-measure	**0.86**	F-measure	**0.92**

Topic 2	Recall	**68%**	Recall	**87.5%**
	Precision	**73.6%**	Precision	**85.1%**
	F-measure	**0.70**	F-measure	**0.86**

Putting these results in perspective using the F-measure (see Table 5), we observe that, in terms of overall performance, the group with the tailored brochures did better.

Referring back to our original hypotheses, we can conclude that we have shown that the results observed were not due to the performance of a specific group, but due to the tailoring. With content tailored to their needs, people can find the information they are looking for more quickly, more accurately, and thus, overall, demonstrate better performance.

[2] For example, for the question about naming the divisions carrying out some research on topic 1, if the names given were incorrect, we counted 1 point by incorrect name.

3 Conclusion

Our aim was to understand the impact of tailoring on information seeking tasks. We presented an experiment and its results, showing that tailored documents have an impact on information seeking, at least when the information to be found is spread over a number of sources. With documents tailored to their needs, people find the information they seek more quickly, and overall, more accurately. We can conclude that tailoring is indeed useful in information seeking tasks.

Acknowledgement

We want to thank all members of our team for their contribution to the development of SciFly and all the people who participated in our experiment.

References

1. André, E., Müller, J., Rist, T.: WIP/PPP: automatic generation of personalized multimedia presentations. In: Proc. of ACM Multimedia, pp. 407–408 (1996)
2. Campbell, M.K., DeVellis, B.M., Strecher, V.J., Ammerman, A.S., DeVellis, R.F., Sandler, R.S.: Improving dietary behavior: The effectiveness of tailored messages in primary care settings. American Journal of Public Health 84, 783–787 (1994)
3. Chedrawy, Z., Sibte, S., Abidi, R.: An Adaptive Personalized Recommendation Strategy Featuring Context Sensitive Content Adaptation. In: Wade, V., Ashman, H., Smyth, B. (eds.) AH 2006. LNCS, vol. 4018, pp. 61–70. Springer, Heidelberg (2006)
4. De Carolis, B., De Rosis, F., Andreoli, C., Cavallo, V., De Cicco, M.L.: The dynamic generation of hypertext presentations of medical guidelines. The New Review of Hypermedia and Multimedia 4, 67–88 (1998)
5. Elhadad, N., McKeown, K., Kaufman, D., Jordan, D.: Facilitating physicians' access to information via tailored text summarization. In: AMIA Annual Symposium, 2005, Washington DC (2005)
6. Kaplan, C., Fenwick, J., Chen, J.: Adaptive Hypertext Navigation Based on User Goals and Context. In: User Modelling and User Adapted Interaction, 3, pp. 193–220 (1993)
7. Paris, C., Wu, M., Vercoustre, A-M., Wan, S., Wilkins, P., Wilkinson, R.: An Empirical Study of the Effect of Coherent and Tailored Document Delivery as an Interface to Organizational Websites". In: Brusilovsky, P., Corbett, A.T., de Rosis, F. (eds.) UM 2003. LNCS, vol. 2702, pp. 133–144. Springer, Heidelberg (2003)
8. Paris, C., Colineau, N.: Scifly: Tailored Corporate Brochures on Demand. CSIRO Technical Report 06/268 (2006)