the user. Yet this hypothesis is not fully confirmed if the users never changed the default settings. The situation with the ranking flexibility is much clearer – users appreciated the ability to change the ranking and used it frequently. Figure 9 shows the frequency of the ranking preset change and the amount of time subjects used presets during their exploratory searches. They used the half-and-half preset most of the time (46% in frequency and 54% in time). The next-favored preset was alpha=0.0, which removes the effect of the task model completely (35% in frequency and 30% in time). The least favored preset was alpha=1.0, which considers the task model only. While subjects could consider this preset an extreme because it ignores the effect of their own queries, its use was quite considerable (19% in frequency and 17% in time).

Because we had provided the preset alpha=0.5 as a default, we also need to consider whether it was favored by the subjects simply because it was the default. In Figure 10, we discarded the default use of this preset (alpha=0.5). The data shows that the frequency of direct preset change to 0.5 (from 1.0 or 0.0) is 24% and the amount of time subjects spent with this preset when it was selected directly was 14%. In total, we registered 124 explicit preset changes – more than 12 for each subject at average! Even though users favored the alpha=0.0 preset when making explicit choices, we should not disregard that they were still using the task model more (alpha=1.0 or 0.5) than the query only preset (alpha=0.0), where the frequency was 43% and the time spent was just 30% of the whole sessions.

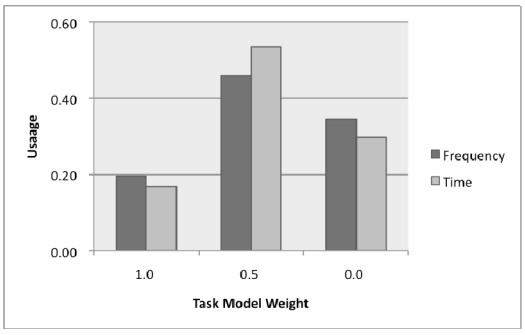

Figure 9 Task model weights

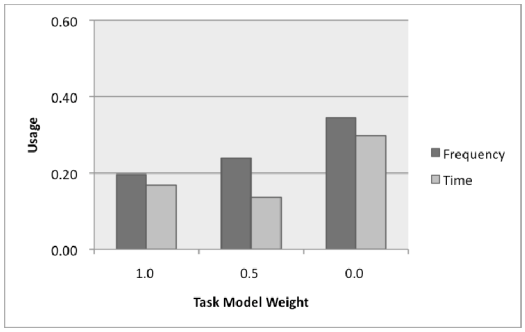

Figure 10 Task model weights (removed defaults)

We also tried to compare the user and system performances during each of these three presets. In Table 6 and Table 7, we separated the user/system performance measures discussed in the previous sections by the change of the presets. It clearly shows that the system performance (Table 7) was better when the subjects used the task model (when alpha=1.0 or 0.5) than when they just used the system in the query only mode (alpha=0.0). Also, alpha=1.0 task model weight mode was the best when we examined the note precision even though alpha=0.5 showed relatively poorer performance. User performance in terms of the open precision was about the same among the three modes, but the subjects showed higher rate of activity counts while they were taking notes, opening documents, and searching (217 vs. 110, 198 vs. 66, and 80 vs. 41 respectively). It is also interesting to note that the document access precision with preset alpha=0.0 (query only) was much higher than the average access precision in the baseline system. It provides some evidence that the users really mastered the preset manipulation picking up the most appropriate presets for different queries.

Table 6 User performance with varying task model weights

User Performance	1.0	0.5	0.0
Note Precision	0.97	0.80	0.90
Note Count	38	179	110
Document Access Precision	0.92	0.96	0.97
Document Access Count	51	147	66
Search Count	13	67	41

Table 7 System performance with varying task model weights

System Performance	1.0	0.5	0.0
Precision at Rank 10	0.98	0.95	0.87
Precision at Rank 5	0.96	0.95	0.87

Table 8 Post-questionnaire (*experimental system ONLY)

Q#	Text of Question
1	Were you familiar with this topic before the search?
2	Did the passages and their documents provide you sufficient information for your summary?
3	When choosing to view a full document, was it mostly because you found useful information in the passage?
4	Were you confident in the system's ability to find useful information on this topic?
5	Was it easy to mark up useful snippets using this system?
6*	Did you find the query-vs-profile weight adjustment helpful in finding useful information?
7*	Was displaying the terms in your task model helpful to you?
8*	Did you find the inclusion of passages with terms from your task model helpful in finding useful information?
9*	Did you find the highlighting of terms from your task model in the passages helpful?
10	Overall, did you have a positive experience with this system?

6. USER SUBJECTIVE ANALYSIS

Following each search task, subjects were given a post-questionnaire (Table 8) to assess their satisfaction with the version of TaskSieve used to complete the assigned task. Using a 5-point Likert scale, subjects were asked to rate their level of agreement (1=Not at All; 5 = Extremely) regarding their familiarity with the assigned task topic (Question 1); the sufficiency of news provided (Question 2); usefulness of the document summaries (i.e. surrogates) in the search results (Question 3); their ability to find useful passages (i.e. snippets) (Question 4), the system's ease of use (Question 5), and overall satisfaction with the system (final question.) For the experimental system only, subjects were asked to rate the utility of the features related to the task model – query-profile weighting adjustment, task model visualization, task-model based snippets, and highlighting of task model terms in snippets.

Chi-square tests were performed on the questionnaire data to determine significant differences in subject responses by system and by topic. Table 9 shows the mean post-questionnaire responses by system. There were no significant differences in any of the users'