Information Processing & Management Vol. 26, No. 6, pp. 719–738, 1990
Printed in Great Britain.
0306-4573/90 $3.00 + .00

INTEGRATION OF USER PROFILES: MODELS AND EXPERIMENTS IN INFORMATION RETRIEVAL

SUNG H. MYAENG
4-206 Center for Science & Technology, School of Information Studies,
Syracuse University, Syracuse, NY 13244

and

ROBERT R. KORFHAGE*
Department of Information Science, University of Pittsburgh,
Room 737 LIS Building, Pittsburgh, PA 15260

(*Received* 27 *May* 1988; *accepted in final form* 26 *April* 1990)

Abstract — One difficult problem in information retrieval (IR) is the proper interpretation of user queries. It is extremely hard for users to express their information needs in a specific yet exhaustive way. In an effort to alleviate this problem, two theoretical models have been proposed to utilize user characteristics maintained in the form of a user profile. Although the idea of integrating user profiles into an IR system is intuitively appealing, and the models seem viable, no research to date has established a foundation for the roles of user profiles in such a system. Aiming at the investigation of the roles of user profiles, therefore, this study first identifies and extends various query/profile interaction models to provide a ground upon which the investigation can be undertaken. From a continuum of models characterized on the basis of interaction types, metrics, and parameters, nearly 400 models are chosen to investigate the "model space." New measures are developed based on the notion of user satisfaction/frustration. In addition, three different criteria are used to guide users in making judgments on the quality of retrieved items. Analysis of the data obtained from the experiments shows that, for a wide variety of criteria and metrics, there are always some query/profile interaction models that outperform the query alone model. In addition, preferable characteristics for different criteria are identified in terms of interaction types, parameters, and metrics.

1. INTRODUCTION

The problem of retrieving information from natural language databases has been studied during the past quarter century. In traditional context, retrospective information retrieval (IR) systems are those in which a user initiates the search process by means of a set of active queries and receives a set of references to items of potential interest.

One difficult problem in such systems is the transformation of the user's information need to the form of an explicit query which accurately matches the original intention, and retrieves all items of interest in the database being searched, and only those. Therefore, users often have great difficulty in using an IR system successfully regardless of the query language implementation (e.g., a vector form, a boolean expression of terms, a combination of both [1,2,3], or other retrieval models [4,5,6,7,8]). As a result, user queries are not completely satisfactory in expressing the needs in most retrieval situations. It seems natural that the output of a system based on such a query is necessarily incomplete and unsatisfactory.

One reason underlying this query formulation problem is the mismatch between terms used in a query and those used in documents. Blair and Maron [9] analyzed the poor performance of a large-scale IR system and contended that it is exceedingly difficult for users to predict the exact words, word combinations, and phrases that are used by all (or most) relevant documents and only (or primarily) by those documents. This difficulty is usually

*All editorial correspondence should be sent to Robert R. Korfhage.

reflected by the inverse relationship between precision and recall, the two most commonly used measures of retrieval effectiveness. To alleviate this difficulty, various schemes have emerged. Thesauri have been used to expand the vocabulary and to allow users to modify queries. By utilizing many-one and one-many relationships between words and their referents, precision and recall can be enhanced, respectively [10]. In interactive retrieval situations, the relevance feedback technique has been used to allow users to reformulate queries based on the relevance assessment supplied by themselves for previously retrieved items. It has been shown experimentally that this technique can achieve improvements in retrieval effectiveness [11].

The difficulty of adequate query formulation also seems related to the subtlety of the human information seeking behavior. Widely recognized is the fact that different users usually expect different sets of items from the same query and make different relevance judgments on the same retrieved items. This means that user variability should be considered as a factor in information seeking process [12] and incorporated into the system design in some way. However, since the typical communication achieved between a user and a system is only through a set of queries and a set of retrieved items, this somewhat narrow and restricted channel inhibits the system from catering to the individual's variability in terms of information needs.

It is conceivable that by maintaining characteristics of an individual user in the form of a profile, the bandwidth of the communication channel can be widened. Used as a way of improving the level of user/system communication effectiveness, the profile information is expected to allow the underlying system to understand users better and to improve the quality of a retrieval output. For instance, the profile information allows a different interpretation of a query to produce a different result, and helps the initial output to be tailored to the user's particular needs and ranked appropriately, based on the user's preference. While the use of tools such as thesauri and stemming algorithms for a priori processing of a query aims at better query interpretation by depersonalizing the query in a sense, profiles are used for the same purpose by personalizing the query [13].

The influence of the user profile on the quality of output depends on various factors. One important and immediate consideration is how to modulate the interaction between a query and a profile, so that reasonable quality of information is maintained. Some models of query/profile interaction have been developed and their theoretical foundations have been established [14,15,16]. Another aspect to be considered is how to maintain user profiles. Assuming that reasonably well-constructed profiles increase the system effectiveness, the nature and quality of the information in user profiles should determine the degree of improvement. Recognizing that people tend to be poor at self-description, a method of automatically and dynamically updating user profiles has been proposed to facilitate an intelligent and personalized IR system [17].

Researchers have recognized directly or indirectly the need for user modeling in various information systems. Given that information seeking is part of the problem solving process, it is difficult to study information seeking apart from a particular context or process [12]. In particular, IR system outputs need to be produced based not only on the topicality of documents and queries, but also on informativeness, often affected by such factors as novelty, understandability, the order of output presentation, and the suppression of redundancy [18], which are dependent on individual users. If an IR system is to be designed to take into account individual variability in backgrounds, interests, preferences, or other significant characteristics, it becomes obvious to develop a form of user models for individuals. Nonetheless, the possibilities for user representations have been explored only to a limited extent in experimental IR systems [19], and uncertainty about how to incorporate knowledge about users into system design is a major stumbling block in designing effective IR systems [20]. Indirect uses of the concept of user modeling in IR are found in [21] and in [22,23].

This study aims at demonstrating the superiority of IR systems with profiles, a limited form of user models, to those without profiles, and investigating the query/profile "model space" in order to develop a theory. In this paper, we first present the "model space" constructed by identifying and extending the existing query/profile interaction models and then report the results of a series of experiments conducted to meet the objectives.

2. THE MODEL SPACE

Since this research aims at investigating the roles of user profiles in a general sense, various interaction models have been reviewed and extended to serve as a ground on which the investigation can be undertaken. Given that a profile contains information about a user's (or a group of users') interest, it may be used in three distinct ways, depending on when and how it is applied to the retrieval process. First, the interest profile can play a role in preprocessing a query to produce a modified query to be used in the subsequent retrieval process. Second, the profile and the query can be considered as the same kind of entity directing the retrieval process. The third possibility is to treat the profile as a filter to post-process outputs retrieved based on the query alone. Although each method possesses its own potential merit, the first two have been the focus of this research; they lend themselves to the theoretical framework developed to date.

Even with two methods of using interest profiles, there is a continuum of models from which 396 different models have been identified and selected to investigate the "model space." For ease of manipulation and theory development, they are organized along three different dimensions:

1. modes of query/profile interaction,
2. parameters embedded in the interaction modes, and
3. metrics used to discriminate among documents.

2.1. *Representation*

The basic representational scheme adopted and used in this research is a vector model. This not only makes it possible to handle various information retrieval objects easily, but also opens the possibility of using well-established mathematical properties of a general vector space. In this representation, documents, queries, and profiles can be regarded as points in an n-dimensional space where n is determined by the number of terms or descriptors in a retrieval system. Given any two points in the vector space, the distance between them can be computed and used in a number of different ways. For example, Euclidean distance can be computed between a query point and a document point so that a spherical shell around the query point can determine whether the document should be retrieved (Fig. 1).

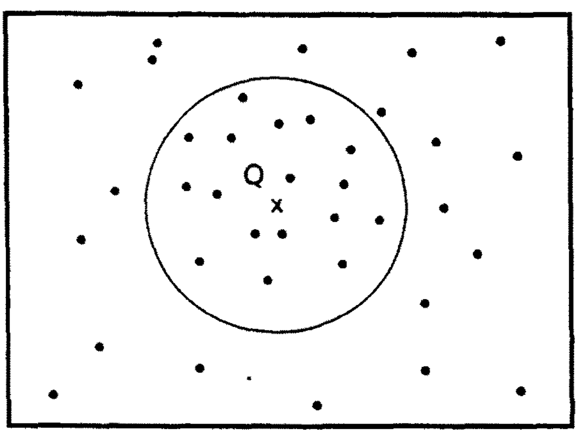

Fig. 1. Conceptualization of document space.

2.2. *Modifying the query*

Given that a user query is more or less inaccurate and incomplete, it seems desirable to adjust the component values of the query vector based on profile vector, thereby moving the query point in space. This means that the two original entities (query and profile) are no longer considered, and only the modified query affects the retrieval process. While this modification of the query can be achieved in a number of ways, two modification methods using linear models are studied for their simplicity and intuitive appeal.

In the first method, it is assumed that the effect of using a profile is uniform across all components. Hence, the modified vector Q' can be defined in a simple linear equation as follows:

$$Q' = tQ + (1 - t)P$$

for some value t, $0 \leq t \leq 1$. Obviously the value of t determines the relative importance of the profile information in the interaction.

The second method is based on the observation that zero-valued elements of P should not reduce the value of the corresponding elements of Q because they do not necessarily mean disinterest in the concept represented by the term; they may simply indicate lack of information in the profile. Instead of reducing the value of zero-valued terms in Q by the factor of t in the first method, this method only considers non-zero P terms. With two parameters, α and β, and p_i and q_i denoting the value of the i-th element in P and Q, respectively, the value of the i-th element in Q', q_i', is defined as follows:

$$\begin{array}{ll} q_i + (1 - |q_i|)p_i & \textit{if } q_i p_i > 0 \\ q_i + \alpha p_i & \textit{if } q_i p_i < 0 \\ \alpha p_i & \textit{if } q_i = 0 \text{ and } p_i > \beta \\ q_i & \text{otherwise.} \end{array}$$

It is clear in this method that no change is made to the zero value of an element in Q unless the value of the corresponding element in P is sufficiently large. Because this method allows different effects for different components, the term "piecewise model" is used in the sequel to refer to a model using this method.

2.3. *Models with two focal points*

An alternative way of using the profile lies in the conceptualization of the profile and query points as two separate entities that exist throughout the retrieval process and interact to form a more complex and flexible shell in the vector space. In other words, the "goodness" of a document is determined as a function of the two distances from P and Q to the document. Two methods of using these points as foci in shaping the shell have been developed [13,14,16].

As a natural extension to retrieval based on a spherical shell, the first method uses an ellipsoidal shell with P and Q as two focal points. The retrieval decision for documents is therefore made by considering the sum of two distances from P and Q to those document points:

$$\|D, Q\| + \|D, P\|,$$

where $\| \ \|$ denotes the distance between any two points in the space, measured by any metric. While the shell defined by an ellipsoid around the two reference points seems reasonable with two points close together, a problem surfaces as they get farther apart. For instance, if a query happens to be far different from the current interest reflected by the profile, a substantial number of documents retrieved will be close to neither the query nor the profile (Fig. 2).

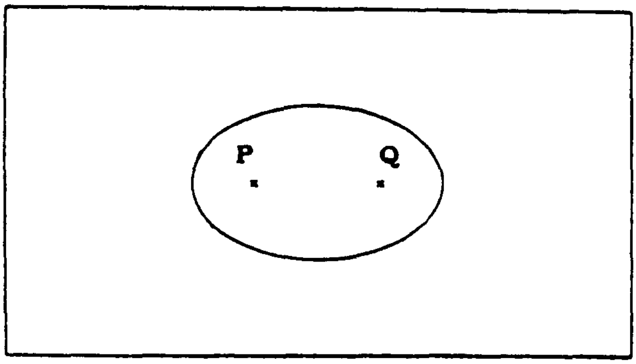

Fig. 2. An ellipsoidal shell.

The second method developed by Liu [16] remedies this problem through the use of Cassini ovals, providing an entirely different family of shells. More precisely, the retrieval decision is based on the following multiplicative inequality:

$$\|D,Q\| * \|D,P\| \leq k$$

where k is a threshold that determines the size and shape of a shell within which documents to be retrieved are found. As shown in Fig. 3, the shape of an Cassini oval is similar to that of an ellipse when the threshold is large (or, with a constant threshold, when two points P and Q are reasonably close to each other). As the threshold gets smaller, however, two longer sides gradually become caved in, resulting in a peanut shape and eventually forming two separate shells around two points. This property of Cassini ovals has direct bearing on information retrieval situations because the progressive change of the shapes seems to reflect the possible variations in the closeness of profile and query. Specifically, for example, when a query is not closely related to the profile, two separate clusters of documents are retrieved, one based on each of the query and the profile, to exclude those documents irrelevant to both of them.

2.4. *Q' as a reference point*

Assuming that the user's current information need is better reflected in a modified query Q' than in the initial query Q, it seems natural to investigate aforementioned models with Q' as one of the focal points. Since Q' must be located somewhere in the "middle" between Q and P in the space, the substitution of Q' in the distance computation for P or Q can generate two overlapping shells, each of which is smaller than that based on

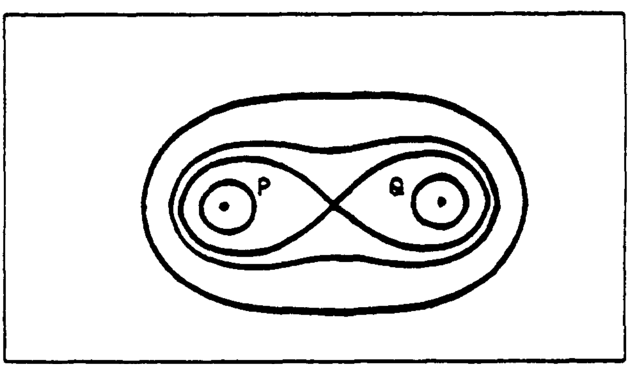

Fig. 3. Cassini oval shells.

P and Q (Fig. 4). With this kind of substitution, four different groups of models are identified with the following retrieval criteria:

ellipse on Q' and P:	$\|D,Q'\| + \|D,P\|$
Cassini oval on Q' and P:	$\|D,Q'\| * \|D,P\|$
ellipse on Q' and Q:	$\|D,Q'\| + \|D,Q\|$
Cassini oval on Q' and Q:	$\|D,Q'\| * \|D,Q\|$.

These groups of models are particularly interesting in that the distance between two focal points is always shortened and the shell is moved toward either P or Q. These seem most appealing when P and Q are initially far apart. Instead of retrieving documents encompassed by a large ellipse or by two separate shells, these models can retrieve documents encompassed by a single shell of a smaller size, which is formed around either P or Q and shifted toward Q or P depending on the different emphases (Fig. 4).

Since three distinct reference points are available in the space now, it is also possible to consider models retrieving documents based on all three at once, not two at a time. Although it is worthwhile to investigate this possibility, the current research does not include such models in the model space. The three-point models would become more attractive if nonlinear models were used for the purpose of generating Q'.

2.5. *The metrics*

The second facet of defining the model space lies in the use of various metrics to compute the distance between a document and any of the three reference points in the n-dimensional document space. For this research four different metrics were used. It is assumed that similarity between a document and a query (or profile) is inversely related to the distance between them. The cosine similarity measure has been widely used for its simplicity and effectiveness [1]. Hence the inverse of the cosine similarity measure is one of the metrics used. Its effect on retrieval is compared with the effects of three metrics from the L_p family. For the distance between a document $D = \langle d_1, d_2, \ldots, d_n \rangle$ and a query $Q = \langle q_1, q_2, \ldots, q_n \rangle$, the class of L_p metrics is defined as:

$$\|D,Q\|_p = (\Sigma_{i=1 \text{ to } n} |d_i - q_i|^p)^{1/p}.$$

Among the infinite number of possible p values, L_1, L_2, and L_∞ are chosen because of their extremity and interpretation. In addition, the selection is justified by recent experimental results on the application of these metrics to the extended Boolean information retrieval system [1,2].

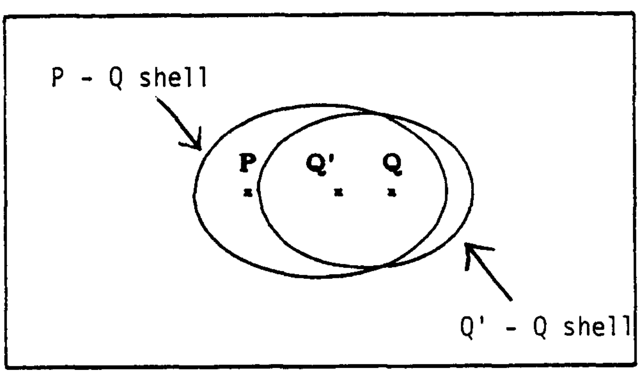

Fig. 4. Q' as a reference point.

The L_1, L_2, and L_∞ metrics are defined by:

$$\|D,Q\|_1 = \Sigma_{i=1 \text{ to } n} |d_i - q_i|$$

$$\|D,Q\|_2 = (\Sigma_{i=1 \text{ to } n} |d_i - q_i|^2)^{1/2}$$

$$\|D,Q\|_\infty = max_{i=1 \text{ to } n} |d_i - q_i|$$

respectively. On the other hand, the distance measured by "inverse cosine" is computed by:

$$\text{invCos}(D,Q) = 1 - \frac{\sum_{i=1}^{n} d_i \cdot q_i}{\left(\sum_{i=1}^{n} d_i^2 \cdot \sum_{i=1}^{n} q_i^2\right)^{1/2}}.$$

All the interaction modes discussed in the previous section can now employ four different metrics introduced in this section. To put it differently, the model space investigated can be partitioned into four groups according to the metric used by each model. Early experiments showed that models using the L_1 metric did not produce improved retrieval results. Hence that metric was dropped, and we will consider it no further.

2.6. *Weighting and parameters*

In considering the interaction between the query and the user profile, we may wish to give one or the other more significance. We can modify the ellipsoidal and Cassini oval models to modulate the relative importance among Q, P, and Q' by using weighting factors in the distance computations.

In the case of ellipsoidal models, six parameters (a_1, b_1, a_2, b_2, a_3, and b_3) are introduced for each metric used:

$$a_1 \|D,Q\| + b_1 \|D,P\|$$

$$a_2 \|D,Q'\| + b_2 \|D,P\|$$

$$a_3 \|D,Q'\| + b_3 \|D,Q\|.$$

While multiplicative parameters seem reasonable for elliptical models, their roles are not distinctly identifiable in Cassini oval models because they can be factored out together; instead, exponential parameters are used as follows:

$$\|D,Q\|^{c_1} * \|D,P\|^{d_1}$$

$$\|D,Q'\|^{c_2} * \|D,P\|^{d_2}$$

$$\|D,Q'\|^{c_3} * \|D,Q\|^{d_3}.$$

As shown already in an earlier section, several parameters come into play when a modified query Q' is generated: t in the simple linear case and α and β in the piecewise case. Obviously, the value of t or α can vary depending on the relative importance of Q or P, whereas β limits the unconditional inclusion of a profile term in the new query when the corresponding term is absent in the original query.

In sum, there are 15 different types of parameters included in various interaction modes. Under the assumption that there is a certain smoothness and monotonicity in the effects of parameter changes [24], only some extreme values are considered, generating a feasible number of models to be investigated. The 396 models investigated enumerated in Table 1 (99 models in each of four metrics) are shown explicitly in two groups: one-point

Table 1. Enumeration of models in each metric, m

One Point Models

P	Q	Q'						
		Simple Linear(SL)			Piecewise(PW)			
		t=.9	t=.1	t=.5	$\alpha>.5$ $\beta>.5$	$\alpha>.5$ $\beta<.5$	$\alpha<.5$ $\beta>.5$	$\alpha<.5$ $\beta<.5$
m01	m02	m03	m04	m05	m06	m07	m08	m09

Two Point Models

			Ellipsoidal			Cassini Oval		
			W_a =.1	W_a =.9	W_a =.5	W_a =.1	W_a =.9	W_a =.5
Q & P			m10	m11	m12	m13	m14	m15
Q' & P	Q' (SL)	t=.9	m16	m17	m18	m37	m38	m39
		t=.1	m19	m20	m21	m40	m41	m42
		t=.5	m22	m23	m24	m43	m44	m45
	Q' (PW)	>>*	m25	m26	m27	m46	m47	m48
		><	m28	m29	m30	m49	m50	m51
		<>	m31	m32	m33	m52	m53	m54
		<<	m34	m35	m36	m55	m56	m57
Q' & Q	Q' (SL)	t=.9	m58	m59	m60	m79	m80	m81
		t=.1	m61	m62	m63	m82	m83	m84
		t=.5	m64	m65	m66	m85	m86	m87
	Q' (PW)	>>	m67	m68	m69	m88	m89	m90
		><	m70	m71	m72	m91	m92	m93
		<>	m73	m74	m75	m94	m95	m96
		<<	m76	m77	m78	m97	m98	m99

* This indicates $\alpha > .5$ and $\beta > .5$.

m=0: L_1; m=1: L_2; m=2: L_∞; m=3: Inverse Cosine

models and two-point models. It should be noted that the parameter W_a determines the weight of the first term in any of the two-point models. For instance, $m16$ represents a model in any of the four metrics in the following form:

$$0.1\|D,Q'\| + 0.9\|D,P\|$$

where $t = 0.9$ shows that Q' is the same as in $m03$, that is, $q_i' = 0.9\ q_i + 0.1\ p_i$, $i = 1,\ldots,n$. Similarly, $m19$ represents a model in any of the four metrics in the form:

$$0.1\|D,Q'\| + 0.9\|D,P\|$$

where Q' is as in $m04$; that is, $q_i' = 0.1\ q_i + 0.9\ p_i$, $i = 1,\ldots,n$.

3. EXPERIMENTS

With the theoretical and preparatory work described, a series of experiments was conducted in a laboratory situation, but with many operational components. Ideally, the decision whether an experiment is conducted in operational or laboratory environment should be made in such a way that all and only meaningful variables are manipulated by the experiment. A laboratory test is more desirable when it is necessary to control factors that are not immediate concerns of the experiment but are expected to have some influences on the results, whereas an operational test would be more appropriate if, for example, human cognitive activity is a substantial component of the testing purpose [25]. Since this research has a strong connection with individual users' underlying needs, the laboratory setting was designed to provide a "semi-operational" experimental environment in which users are induced to use real queries and profiles and generate subjective judgments.

An experimental retrieval system called PBS (profile-based system) has been developed for this research. In addition to common features such as accepting a query, searching a database, and retrieving document surrogates, it provides capabilities to handle profiles and evaluate different models based on a query and a profile. The database consists of 3703 abstracts of *Communications of the ACM* from 1958 to 1985. Some standard methods have been employed to analyze and prepare the database for the retrieval purpose. For example, a stemming algorithm was used for both database processing and query processing, and the methods of computing discrimination values and term frequency information [1] were adopted to compute weights on term-document pairs. Details of the structure and components of the PBS as well as methods used for the database process are found in [26].

Considering the large number of models being tested, the goal of the experimental design was to maximize the efficiency of available human resources and minimize the error variances, especially those which might be incurred from uncontrolled individual differences. To this end, every query was processed by all models against the document database so that systematic differences among queries, and hence among users, could hardly mask the actual differences among models. The experimental design had to overcome two difficulties. It is well known that sequencing of the output affects a user's judgment. That is, if document D_2 is seen after document D_1 then the user's judgment of D_2 is affected by the judgment already made on D_1. A similar sequencing effect pertains across models: judgment of the output of a given model is affected by the judgement of prior models. To minimize sequencing effects two strategies were used: the output from all models was merged into a single set, and the documents were presented to the user in a randomized order rather than in an order related to their presumed relevance.

It was expected that there would be a large, but not total, overlap in the document sets retrieved by the different models. To keep the final document set of reasonable size, only the top twenty-five documents from those retrieved by each model were selected for merger into the final set. Thus the sequence of operations was:

1. For each model, retrieve the document set and select the top twenty-five.
2. Merge the selected documents into the final set.
3. Randomize the order of documents within the final set for presentation to the user.

This sequence is shown in Fig. 5. Each experimental subject was asked to review at least 60 documents from the final set. This number was determined on the basis of pilot experiments that indicated excessive user fatigue beyond 60 documents.

Subjects (users) were drawn from the Department of Information Science (DIS) and the Department of Computer Science (DCS) at University of Pittsburgh. Since subjects should be mature enough to specify the technical interests in a profile and to interpret information contained in the highly technical database, only graduate students or those with senior standing were allowed to participate in the experiments. The actual group of volunteered subjects was composed of nine graduate students in DIS and two seniors in DCS, whose contributions varied depending on their availability. The number of queries submitted by each subject ranged from two to four, constituting the total of 30 queries.

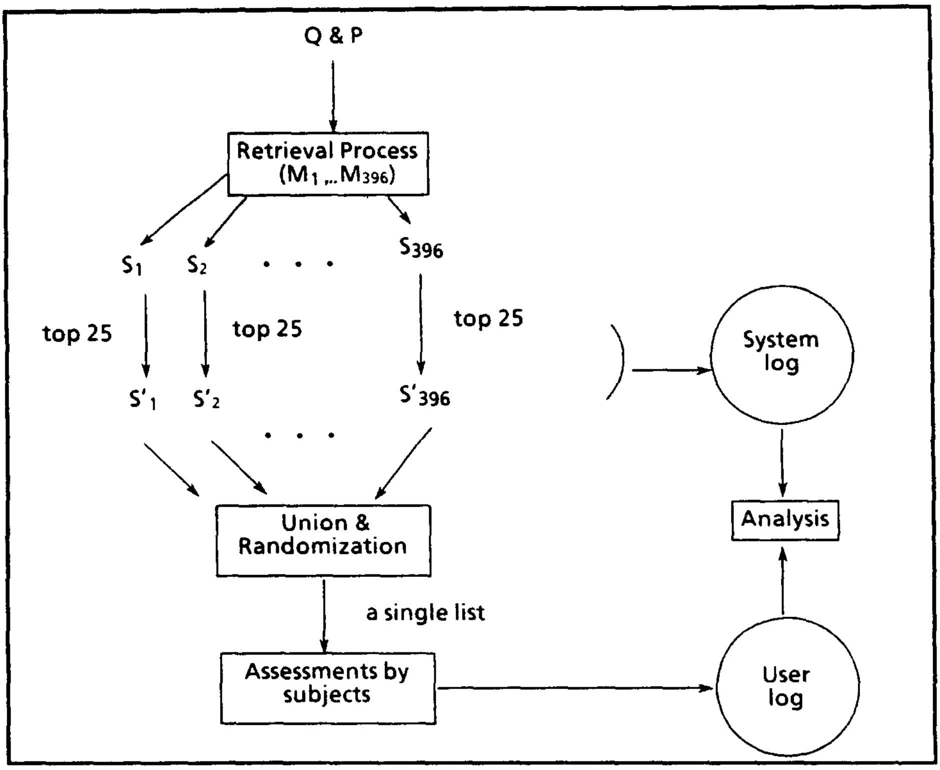

Fig. 5. Schematic diagram for the experimental design.

After an introductory session [26], the first non-trivial task for a subject was to construct a profile as a list of weighted terms that represent his or her real-life interests within the discipline of information and computer science and engineering. The following is an example of a profile constructed by a subject whose main interest lies in AI in general and human/computer communication interface in particular:

((artificial 7) (intelligence 7) (communication 10) (interface 7)
(human 3) (factors 3) (network −2)).

The last term 'network' with a weight of −2 was used to explicate her disinterest in the area of networking, which otherwise might be implied by the inclusion of the term communication. Unspecified weights defaulted to a value of 1.

Subjects were then asked to formulate a query to be searched against the database in the PBS. There was a time interval of at last one day between profile construction and query formulation, which supposedly reduced any unnecessary dependence of a query on the content of a profile. Although they were encouraged to bring their own current information needs for queries to be submitted to the PBS, a pool of real questions drawn from comprehensive examinations given by the DIS at University of Pittsburgh was available as a guide to help them in conceptualizing and defining an information need and thus a query, not as a depository from which they should select an information need.

When the subjects were given a randomized list of documents, they went through documents in that order and determined the quality of each document based on three criteria—relevance, pertinence, and usefulness. These fine-grained criteria were used to forcefully avoid confusion as to how the general term 'relevancy' can be interpreted, as well as to observe what aspects of 'relevancy' are affected by the use of profiles. *Relevance* was to be judged objectively based on how closely a document was related to a 'stated query', regardless of the user's expectation or intention. *Pertinence*, in contrast, was to be judged on how much a document satisfied the current information need or desire that was supposed to be reflected in the query. Obviously this is a more subjective measure in which pragmatics of documents and queries play an important role. If the user's intention is not well embedded in a query, for example, a retrieved document could be relevant but not pertinent. *Usefulness*, finally, was related to the user's short-term and/or long-term interests, regardless of the current need embedded in the query. Thus a pertinent document is expected to be more or less useful, whereas a useful document may not be pertinent at all.

4. ANALYSIS

In view of some limitations in existing measures, including the recall/precision pair and the utility measure, a new measure has been developed for this research. Based on an intuitive interpretation of effectiveness as the combination of users' satisfaction and frustration, the new measure, called *SFT*, satisfies the following properties desired by the current experiments:

1. A single number is produced to make it possible to compare the various models.
2. The number is sensitive to scaled judgments by the user as well as to binary (accept-reject) judgments.
3. The number is sensitive to document ranking differences produced by different models.

The *SFT* measure, T, of a model M is defined by:

$$T(M) = \delta S(M) - (1 - \delta)F(M),$$

where S and F are *satisfaction* and *frustration* measures, respectively, and δ, $0 \le \delta \le 1$, sets the balance between satisfaction and frustration in determining T. The measures S and F are both bounded by 0 and 1, and are computed from the value that the user assigns to each document and the ranking of the documents in the model M. The distinction between S and F is that S uses only those documents that the user rates positively (accepts) and F uses only the negatively related (rejected documents). The parameter δ can be adjusted to give more or less credence to the ranking of accepted documents in comparison to the ranking of rejected ones. In practice, δ might be determined by the purpose of evaluation or individual differences in the perception of effectiveness criteria. In our experiments, $\delta = 0.5$ was used across all subjects, with no emphasis on either the satisfaction or frustration aspect of effectiveness.

The method of computing S and F values is similar to that of normalized recall [27], used for the SMART system. In the case of satisfaction, the S value for a given retrieval session is determined by the cumulative sum of individual degrees of satisfaction obtained every time a "good" document is found. At the end of the judgment process, the S value for the session is computed as the area difference between two staircase functions representing an ideal case (i.e., user-generated rankings) and an actual case (i.e., system-generated rankings). Unlike the normalized recall case where increments are made uniformly, the amount of increase of satisfaction is equal to the relative weight of the "good" document in proportion to the sum of the weights of all "good" documents considered in the evaluation process, which does not necessarily require that all documents in the collection be reviewed. The F value is computed analogously except that an "ideal" case has a counter-intuitive meaning; it represents the worst retrieval situation, where all the "bad" documents are retrieved before any "good" documents. It should be noted that the precise interpretation of "goodness" or "badness" depends on the evaluation criterion such as relevance, pertinence, or usefulness. The formulas and other details are found in [26].

The investigation of the overall model space began by first formulating a null hypothesis and an alternative hypothesis:

H0:

The integration of user profiles into a conventional IR system has no impact in the effectiveness of such a system.

H1:

The integration of user profiles into a conventional IR system enhances the effectiveness of such a system.

This general hypothesis was broken down into many refined ones because the effectiveness was evaluated by three different criteria (relevance, pertinence, and usefulness) and measured in terms of S, F, and T measures and three different metrics, L_2, L_∞, and inverse cosine that survived the initial screening process [26].

Because it was of primary interest to compare the performance of any two models, especially the performance of all models against that of the query alone model (M_q), two statistical tests were used to ensure significance of differences observed: the paired T test and the Wilcoxon signed-rank test. While the paired T test, a variation of the well-known T test used commonly in information retrieval experiments [1], was chosen on the ground that the samples (queries) are not considered as independent, the Wilcoxon signed-rank test was employed to reinforce the result by the paired T test, which assumes normal populations.

5. RESULTS

Results were summarized separately based on pertinence, relevance, and usefulness, and discussed with respect to S, F, and T values in each metric. Numbers in the tables represent an S, F, or T value: the greater an S value, the more satisfaction and the less difference between the user's ranking and a particular model's ranking of "good" documents; the greater an F value, the more frustration and the less difference between a model's ranking and the worst possible ranking of "bad" documents. As described earlier, a T value merely represents a linear combination of an S and an F value of a particular model, permitting easy comparison among many models. The only values shown in each table are the M_q value and those model values that are significantly better than it. This paper describes only the most salient results; all details are found in [26].

5.1. *Results on pertinence*

Because clear differences were observed with the use of different metrics for the same set of interaction models and parameters, the performance differences among different models within each metric group were first analyzed. In L_2, the results suggest that pertinence is most enhanced when the balance between Q and P is maintained properly, not when a heavy emphasis is placed on either of them. However, only five models in S, zero in F, and two in T are better than M_q (see Table 2 for T).

The behavior of the models in the L_∞ metric group shows a substantial improvement compared to that of M_q. There are two distinctive patterns observed in Table 3. First, models in the Q & Q' category look more attractive than any others in increasing overall pertinence because most models in the category are outstanding, whereas no models in other categories appear to be better than M_q. The other pattern is related to a comparison between *SL* and *PW* models. Excluding those carrying insufficient query information in *SL* models, general superiority of *SL* models to *PW* models is observed. This is especially true for those using equal query and profile weights in constructing Q', which outperformed other models, confirming the observation made in the case of L_2.

Although the performance of models in the inverse cosine group was expected to be similar to that of L_2 because of the similarity in distance computation, there are at least two major differences (see Table 4). First, 10 models in this metric group show some improvements in reducing user frustration, whereas no model in the L_2 and L_∞ cases performs better than M_q in this respect. Second, although a balanced mix of profile and query information seemed to be a requirement for models to be outstanding in the L_2 and

Table 2. Statistically better models: pertinence, total, $L_2 (m = 1)$

One Point Models

P	Q	Q'						
		Simple Linear(SL)			Piecewise(PW)			
		t=.9	t=.1	t=.5	$\alpha>.5$ $\beta>.5$	$\alpha>.5$ $\beta<.5$	$\alpha<.5$ $\beta>.5$	$\alpha<.5$ $\beta<.5$
-	.1358	-	-	-	-	-	-	-

Two Point Models

			Ellipsoidal			Cassini Oval		
			W_a =.1	W_a =.9	W_a =.5	W_a =.1	W_a =.9	W_a =.5
Q & P			-	-	-	-	-	-
Q' & P	Q' (SL)	t=.9	-	-	-	-	-	-
		t=.1	-	-	-	-	-	-
		t=.5	-	-	-	-	-	-
	Q' (PW)	>>*	-	-	-	-	-	-
		><	-	-	-	-	-	-
		<>	-	-	-	-	-	-
		<<	-	-	-	-	-	-
Q' & Q	Q' (SL)	t=.9	-	-	-	-	-	-
		t=.1	-	-	-	-	-	-
		t=.5	-	-	-	-	-	.1770
	Q' (PW)	>>	-	-	-	-	-	-
		><	-	-	.1634	-	-	-
		<>	-	-	-	-	-	-
		<<	-	-	-	-	-	-

* This indicates $\alpha > .5$ and $\beta > .5$.

L_∞ cases, among the models determined to be superior to M_q in this metric group, those seem to perform better where heavier weights are placed on query components by setting a large value for the parameter t or by emphasizing Q in forming an n-dimensional shell.

Several remarks on general trends across different metrics are in order. First, the experimental evidence shows that integration of the user profile is effective in increasing user satisfaction but not in decreasing user frustration, except for some models in the inverse cosine metric group. Because the increases in S values are more significant than the increases in F values, substantial increases in T values are observed. Second, two point models with Q' and Q as two focal points seem more attractive and worthy of further investigation than any other models. Third, differences among PL models seem quite insensitive to the value of the parameter α. This seems to indicate that a judicious introduction or deletion of terms is far more critical to effectiveness than the change of a term weight.

5.2. *Results on relevance*

As defined previously, relevance is an objective criterion to measure how closely a retrieved document is related to a stated query that does not necessarily represent user's in-

Table 3. Statistically better models: pertinence, total, $L_\infty (m = 2)$

One Point Models

P	Q	Q': Simple Linear(SL) t=.9	t=.1	t=.5	Q': Piecewise(PW) $\alpha>.5$ $\beta>.5$	$\alpha>.5$ $\beta<.5$	$\alpha<.5$ $\beta>.5$	$\alpha<.5$ $\beta<.5$
-	.0931	.2078	-	-	.1571	.1571	.1525	.1525

Two Point Models

			Ellipsoidal W_a =.1	W_a =.9	W_a =.5	Cassini Oval W_a =.1	W_a =.9	W_a =.5
Q & P			-	-	-	-	-	-
Q' & P	Q' (SL)	t=.9	-	-	-	-	-	-
		t=.1	-	-	-	-	-	-
		t=.5	-	-	-	-	-	-
	Q' (PW)	>>*	-	-	-	-	-	-
		><	-	-	-	-	-	-
		<>	-	-	-	-	-	-
		<<	-	-	-	-	-	-
Q' & Q	Q' (SL)	t=.9	.2079	.2071	.2075	.2079	.2071	.2075
		t=.1	-	-	-	-	-	-
		t=.5	.2247	.2332	.2359	.2240	.2334	.2332
	Q' (PW)	>>	.1665	.1701	.1668	.1677	.1712	1667
		><	.1665	.1701	.1668	.1677	.1712	.1678
		<>	.1619	.1654	.1621	.1630	.1654	.1632
		<<	.1619	.1654	.1621	.1630	.1666	.1632

*** This indicates $\alpha > .5$ and $\beta > .5$.**

tention. Therefore, the primary purpose of using this criterion is not to select better models, but to differentiate between objective and subjective judgments that could be intermixed without it. This separation of relevance from pertinence is essential to the identification of models or their characteristics that will best satisfy a particular user's information needs, not those of a hypothetical user.

As one can expect, the experimental evidence shows little or no improvement in terms of relevance: in L_2 metric no models appear better than M_q in terms of any measure; although several models in inverse cosine metric perform better than M_q, the differences in actual values are almost negligible (see Table 5).

In L_∞ metric, however, many models appear to be better than M_q by significant differences. This anomalous result seems attributable to the idiosyncrasy of L_∞ metric and the inadequacy of using the *SFT* measure in association with this metric [26].

5.3. *Results on usefulness*

Another aspect of the models is examined by means of user assessments on the general usefulness of retrieved documents. As summarized in Tables 6 and 7, two prominent

Table 4. Statistically better models: pertinence, total, inverse cosine ($m = 3$)

One Point Models

P	Q	Q'						
		Simple Linear(SL)			Piecewise(PW)			
		t=.9	t=.1	t=.5	α>.5 β>.5	α>.5 β<.5	α<.5 β>.5	α<.5 β<.5
-	.0566	.0869	-	-	-	-	-	-

Two Point Models

			Ellipsoidal			Cassini Oval		
			W_a =.1	W_a =.9	W_a =.5	W_a =.1	W_a =.9	W_a =.5
Q & P			-	.0869	-	-	.0762	-
Q' & P	Q' (SL)	t=.9	-	.1086	-	-	.1018	-
		t=.1	-	-	-	-	-	-
		t=.5	-	-	-	-	-	-
	Q' (PW)	>>*	-	-	-	-	-	-
		><	-	-	-	-	-	-
		<>	-	-	-	-	-	-
		<<	-	-	-	-	-	-
Q' & Q	Q' (SL)	t=.9	.0690	.0839	.0865	.0869	.0839	.0916
		t=.1	.0882	-	-	.0779	-	-
		t=.5	.0894	-	.1038‡	.0818	-	-
	Q' (PW)	>>	-	-	-	-	-	-
		><	.0810	-	.0954§	.0805	-	-
		<>	-	-	-	-	-	-
		<<	.0809§	-	.0957‡	-	-	-

* This indicates $\alpha > .5$ and $\beta > .5$.
‡ Only for Wilcoxen signed-rank test.
§ Only for paired T test.

trends are observed across different metric groups (L_∞ is excluded because of its anomaly, as indicated earlier.) First, almost all profile-based models appear to be better at retrieving useful documents than M_q, regardless of measures and metrics. This experimental evidence that a profile alone retrieves more useful documents than a query alone, which is supposed to represent more direct and short-term information needs, seems counterintuitive; but it supports the premise that it is difficult to formulate a query that will reject useless but relevant documents. Thus, if an IR system is designed to meet a user's general interests as well as temporary needs, a query alone does not seem sufficient to satisfy both demands.

Second, although a profile alone can achieve relatively high performance in usefulness, it does not necessarily follow that the existence of query information always reduces satisfaction (or increases frustration). Instead, it seems essential for models to include and be guided by some query information in their retrieval process. As shown in the tables, the models in the Q' & P category in the L_2 and the inverse cosine metric groups always per-

Table 5. Statistically better models: relevance, satisfaction, inverse cosine ($m = 3$)

One Point Models

P	Q	Q'						
		Simple Linear(SL)			Piecewise(PW)			
		t=.9	t=.1	t=.5	α>.5 β>.5	α>.5 β<.5	α<.5 β>.5	α<.5 β<.5
-	.4419	.4493‡	-	-	-	-	-	-

Two Point Models

			Ellipsoidal			Cassini Oval		
			W_a =.1	W_a =.9	W_a =.5	W_a =.1	W_a =.9	W_a =.5
Q & P			-	.4473‡	-	-	-	-
Q' & P	Q' (SL)	t=.9	-	.4591‡	-	-	-	-
		t=.1	-	-	-	-	-	-
		t=.5	-	-	-	-	-	-
	Q' (PW)	>>*	-	-	-	-	-	-
		><	-	-	-	-	-	-
		<>	-	-	-	-	-	-
		<<	-	-	-	-	-	-
Q' & Q	Q' (SL)	t=.9	.4457	-	.4509‡	.4548	-	.4514‡
		t=.1	.4475‡	-	-	-	-	-
		t=.5	.4504‡	-	-	.4453‡	-	-
	Q' (PW)	>>	-	-	-	-	-	-
		><	-	-	-	-	-	-
		<>	-	-	-	.4483§	-	-
		<<	.4521§	-	-	-	-	-

*** This indicates $\alpha > .5$ and $\beta > .5$.**
‡ Only for Wilcoxen signed-rank test.
§ Only for paired T test.

form better than the profile-alone model when W_a is 0.5. In other words, unless the modified query is very close to the profile, documents retrieved by a well-balanced retrieval shell are more useful than those retrieved by a profile or a query alone, or by a shell distorted by emphasis on the query or profile.

5.4. *Ellipsoidal vs. Cassini oval*

Intuitively, Cassini oval models should perform better than ellipsoidal models when P and Q are located far apart in a document space, whereas little difference would be observed when they are close to each other. To test the hypothesis that the performance of Cassini oval models would be better than that of ellipsoidal models on average, all two-point models were grouped in pairs and compared as illustrated for a selected group of models in Table 8.

For each measure within each metric group, the pairs that showed a significant differ-

Table 6. Statistically better models: usefulness, total, $L_2(m = 1)$

One Point Models

P	Q	Q'						
		Simple Linear(SL)			Piecewise(PW)			
		t=.9	t=.1	t=.5	$\alpha>.5$ $\beta>.5$	$\alpha>.5$ $\beta<.5$	$\alpha<.5$ $\beta>.5$	$\alpha<.5$ $\beta<.5$
.2058	.0067	.0295‡	.2172	.2156	.1225	.1786	.0810§	.1417

Two Point Models

			Ellipsoidal			Cassini Oval		
			W_a =.1	W_a =.9	W_a =.5	W_a =.1	W_a =.9	W_a =.5
Q & P			-	.0360	.2099	-	.0330	.2048
Q' & P	Q' (SL)	t=.9	-	.0553‡	.2080	-	.0559‡	.2043
		t=.1	-	-	-	-	-	-
		t=.5	-	.2252	.2300	-	.2247	.2236
	Q' (PW)	$>>$*	-	.1495	.2335	-	.1565	.2392
		$><$	-	.1897	.2181	-	.1931	.2245
		$<>$	-	.1038	.2313	-	.1053	.2391
		$<<$	-	.1698	.2237	-	.1726	.2385
Q' & Q	Q' (SL)	t=.9	.0143‡	.0202‡	.0195	.0144‡	.0205‡	.0199
		t=.1	.0335	-	.1991	.0322	-	.1998
		t=.5	.0241	.1923	.1215	.0270	.2035	.1385
	Q' (PW)	$>>$	.0141	.1143	.0653	.0140	.1099	.0616
		$><$	.0175	.1677	.0856	.0154	.1636	.0819
		$<>$	.0142	.0723§	.0488‡	.0133	.0705§	.0465‡
		$<<$	.0158	.1292	.0823	.0149	.1253	.0588

* This indicates $\alpha > .5$ and $\beta > .5$.
‡ Only for Wilcoxen signed-rank test.
§ Only for paired T test.

ence in paired T tests were chosen and listed in a table. Because the superiority directions shown in the fifth column were not always consistent within tables, sign-tests were run to make sure any unidirectional superiority was statistically significant. The test results are summarized in Table 9 for all metrics, measures, and criteria.

Superiority of Cassini oval models or ellipsoidal models, ensured by 0.05 significance level, is indicated by a "+" symbol or a "−" symbol, respectively. When a difference is not statistically significant, an "=" symbol is shown in the corresponding cell. For instance, the table shows that Cassini oval models in the L_2 metric group are generally better than ellipsoidal models in the same group when compared in terms of pertinence measured by S values. Although it is difficult to draw a hard conclusion as to the original hypothesis, two local generalizations can be made based on the regularity: Cassini oval models are preferable when L_2 metric is used, whereas ellipsoidal models are more attractive when the inverse cosine metric is used.

Table 7. Statistically better models: usefulness, total, inverse cosine ($m = 3$)

One Point Models

P	Q	Q'						
		Simple Linear(SL)			Piecewise(PW)			
		t=.9	t=.1	t=.5	$\alpha>.5$ $\beta>.5$	$\alpha>.5$ $\beta<.5$	$\alpha<.5$ $\beta>.5$	$\alpha<.5$ $\beta<.5$
.1456	-.0903	-.0550	.1558	.1650	-.0004	.0580	-.0372[§]	.0367

Two Point Models

			Ellipsoidal			Cassini Oval		
			W_a =.1	W_a =.9	W_a =.5	W_a =.1	W_a =.9	W_a =.5
Q & P			-	-.0592	.1240	-	-.0630	.1072
Q' & P	Q' (SL)	t=.9	-	-.0365	.1483	-	-.0315	.1220
		t=.1	-	-	-	-	-	-
		t=.5	-	.1641	.1637	-	.1664	.1653
	Q' (PW)	>>*	-	.0367	.1623	-	.0334	.1598
		><	-	.0707	.1833	-	.0898	.1823
		<>	-	.0003	.1652	-	-.0074	.1560
		<<	-	.0682	.1727	-	.0606	.1620
Q' & Q	Q' (SL)	t=.9	-.0858	-.0615	-.0623	-.0857	-.0615	-.0589
		t=.1	-.0591	-	.1240	-.0628	-	.0997
		t=.5	-.0638	.1554	.0294	-.0658	.1513	.0261
	Q' (PW)	>>	-.0742	-.0105	-.0611[§]	-.0741	-.0088	-.0529[§]
		><	-.0703	.0282	-.0242	-.0697	.0299	.0183
		<>	.0142	-.0389	-.0656	-.0776	-.0372	-.0633[§]
		<<	-.0721	.0106	-.0449	.0149	.0130	-.0416

* This indicates $\alpha > .5$ and $\beta > .5$.
§ Only for paired T test.

6. CONCLUSION

There is little doubt about the importance and potential advantages of integrating user information into underlying systems. Especially in information retrieval, the difficulty of interpreting user queries, which are often incomplete and inaccurate, necessitates the adaptation of a system to their characteristics. This research aims at investigating the idea of integrating user interests in the form of user profile, and establishing a foundation that will justify further development in this direction.

The analysis of the experimental results has demonstrated the superiority of profile-based models over a wide range of criteria and metrics used for evaluation; there were always some models that outperformed the query alone model. Although overall effectiveness was improved for those better models, a dual phenomenon similar to the recall/precision relationship, which often characterizes information retrieval, also occurred; user satisfaction evaluated in terms of pertinence appeared to be increased by integrating a profile, but user frustration was also increased. However, the integration of user profile improved use-

Table 8. Comparison between ellipsoidal and Cassini oval models: pertinence, total, L_2

Model Number		Mean		Superiority	Significance Level
Ellipsoidal	Cassini Oval	Ellipsoidal	Cassini Oval		
118	139	.0793	.1410	<	< .01
123	144	.0141	.0247	<	< .05
124	145	-.0714	-.0559	<	< .10
126	147	.1407	.1529	<	< .05
132	153	.1404	.1513	<	< .05
133	154	.0263	.1587	<	< .05
159	180	.1408	.1410	<	< .10
161	182	.1425	.1375	>	< .10
169	190	.1397	.1266	>	< .10

fulness in both satisfaction and frustration. It was particularly noteworthy that for usefulness almost all profile-based models outperformed the query alone model. Relevance was used as a device that isolated the subjective assessments related to the user's intention from the objective ones. In spite of the theoretical and intuitive appeal of Cassini oval over ellipsoidal models, it was difficult to prove the superiority of the former in general. Instead, Cassini oval models appeared to be attractive in the L_2 metric group, whereas ellipsoidal models seemed better in the inverse cosine metric group.

Although the results support the main hypothesis and make it possible to select promising models for more detailed study, the strong regularity in connection with different parameters and different types of interactions also suggests further investigation of some aspects of the model space. There are numerous possible extensions and improvements to be made in the future. They can be categorized into three groups: methodological improvements, extensions in query/profile interactions, and exploration of using profiling tools.

In retrospect, the limitation of resources precluded possibilities of strengthening the validity of the experimental results; more human resources could have extended the cutoff point imposed on the number of documents reviewed by subjects. In addition, by using multiple, heterogeneous databases and subjects with diverse background, the query interpretation problem is more likely to surface, and it will be possible to investigate the roles of user profiles in more realistic and interesting situations.

While there is room for improvement in terms of more realistic query/profile interaction models, it seems necessary to connect different user groups with different features of models. This will make it possible to map different interaction models to different groups

Table 9. Overall comparison between ellipsoidal and Cassini oval models

	L_2			L_x			Inverse Cosine		
	P	R	U	P	R	U	P	R	U
S	+	=	=	=	+	+	−	−	−
F	−	+	=	−	−	−	=	−	−
T	−	=	=	+	−	−	−	−	−

significance level = 0.05

+: Cassini Oval models preferred

−: ellipsoidal models preferred

=: no significant difference

of users and to develop a system that will adapt its query processing to user characteristics. On the other hand, it would also be interesting to see relationships between models and the proximity of a query and a profile in the document space.

The third area of research is concerned with enhancing the quality of user profiles by means of profiling tools. Two approaches have been explored and are to be developed further. One is to update user profiles automatically based on the interaction with users. In this way, more accurate and up-to-date user information is expected to be maintained [17]. Another approach is based on the finding in psychology that people are better at recognition than at recall performance [28]. With relationships among terms available in a given database, the task of formulating a profile is expected to become less difficult and more effective in that the task becomes a recognition process rather than a recall [29].

REFERENCES

1. Salton, G.; McGill, M.J. Introduction to modern information retrieval. New York: McGraw-Hill; 1983.
2. Salton, G.; Fox, E.A.; Wu, H. Extended boolean information retrieval. Communications of the ACM, 26(11): 1022–1036; 1983.
3. Salton, G. The use of extended boolean logic information retrieval. SIGMOD Record, 14(2): 277–285; 1984.
4. Radecki, T. Fuzzy set theoretical approach to document retrieval. Information Processing and Management, 15(5): 247–259; 1979.
5. Bookstein, A. Fuzzy requests: an approach to weighted boolean searches. Journal of ASIS, 31(4): 240–247; 1980.
6. Crawford, R.G. The relational model in information retrieval. Journal of ASIS, 32(1): 51–64; 1981.
7. Crawford, R.G. Bibliographic retrieval using a relational database system. Proc. of the Annual Conference, Canadian Association for Information Science, (11): 108–116; 1983.
8. MacLeod, I.A. SEQUEL as a language for document retrieval. Journal of ASIS, 30(2): 243–249; 1979.
9. Blair, D.C.; Maron, M.E. An evaluation of retrieval effectiveness for a full-text document-retrieval system. Communications of the ACM, 28(3): 289–299; 1985.
10. Svenonius, E. Unanswered questions in the design of controlled vocabularies. Journal of ASIS, 37(5): 331–340; 1986.
11. Noreault, T. User directed relevance feedback. Ph.D. dissertation, Syracuse University, School of Information Studies, New York; 1979.
12. Rouse, W.B.; Rouse, S.H. Human information seeking and design of information systems. Information Processing & Management, 20(1-2): 129–138; 1984.
13. Korfhage, R.R. Query enhancement by user profile. Proc. of ACM/BCS Symposium on Research & Development in Information Retrieval: 111–121; 1984.
14. Chavarria Garza, H. A user profile-query model for document retrieval. Ph.D. dissertation, Southern Methodist University, Dallas, Texas; 1982.
15. Korfhage, R.R.; Chavarria Garza, H. Retrieval improvement by interaction of queries and user profiles. Proc. COMPSAC'82, Sixth International Conference on Computer Software & Applications, Chicago: 470–475; 1982.
16. Liu, J.A. A distance approach toward an ideal information retrieval systems. M.S. Thesis, Southern Methodist University, Dallas, Texas; 1982.
17. Myaeng, S.H.; Korfhage, R.R. Towards intelligent & Personalized information retrieval systems. Proc. of the ACM SIGART International Symposium on Methodologies for Intelligent Systems; 121–129; 1986.
18. Boyce, B. Beyond topicality. Information Processing & Management, 18(3): 105–109; 1982.
19. Smith, L.C.; Warner, A.J. A taxonomy of representations in information retrieval system design. Journal of Information Science, 8: 113–121; 1984.
20. van Rijsbergen, C.J. A discrimination gain hypothesis. Proc. Sixth Annual International ACM SIGIR Conference on Research & Development in IR: 101–104; 1983.
21. Hausen, H.-L. Outline of a dynamic self-tuning and adaptive information retrieval system. SIGIR Forum XVI: 132–144; 1981.
22. Danilowicz, C. Relative indexing on the basis of users' profiles. Information Processing & Management, 19(3): 159–163; 1983.
23. Danilowicz, C. Users in experts in the document retrieval system model. Int. J. Man-Machine Studies, 21: 245–252; 1984.
24. Korfhage, R.R. Unified models for profile-query interaction in information retrieval. SMU Technical Report CSE-85-16. Department of Computer Science & Engineering, Southern Methodist University, Dallas, Texas; 1985.
25. Oddy, R. Laboratory tests: automatic systems. In K. Sparck-Jones (Ed.), Information Retrieval Experiments. London: Butterworth; 1981.
26. Myaeng, S.H. The roles of user profiles in information retrieval. Ph.D. dissertation, Department of Computer Science and Engineering, Southern Methodist University, Dallas, TX; 1987.
27. Rocchio, J.J. Document retrieval systems – optimization and evaluation. Harvard University Doctoral Thesis, Report No. ISR-10 to the National Science Foundation, Harvard Computational Laboratory, March, 1966.
28. Klatzky, R.C. Human memory: structure and processes; Freeman, 1980.
29. Myaeng, S.H.; Korfhage, R.R. A concept version. In Z.W. Ras; M. Zemankova (Eds.), Methodologies for Intelligent Systems. New York: North-Holland: 185–192; 1987.