A Decision Theoretic Approach to Combining Information Filters: An Analytical and Empirical Evaluation

Yuval Elovici and Bracha Shapira
Department of Information Systems Engineering, Ben-Gurion University, Beer-Sheva, Israel.
E-mail: elovici@inter.net.il, bshapira@bgumail.bgu.ac.il

Paul B. Kantor
Department of Library and Information Science, School of Communication Information and Library Studies (SCILS), Rutgers University, 4 Huntington Street, New Brunswick, NJ 08901-1071.
E-mail: kantor@scils.rutgers.edu

The outputs of several information filtering (IF) systems can be combined to improve filtering performance. In this article the authors propose and explore a framework based on the so-called information structure (IS) model, which is frequently used in Information Economics, for combining the output of multiple IF systems according to each user's preferences (profile). The combination seeks to maximize the expected payoff to that user. The authors show analytically that the proposed framework increases users expected payoff from the combined filtering output for *any* user preferences. An experiment using the TREC-6 test collection confirms the theoretical findings.

Introduction

Many retrieval evaluation studies support the hypothesis that combining retrieval output from several systems enhances the benefits of the individual systems, resulting in improved effectiveness of combined results (Bartell, Cottrell, & Belew, 1994; Croft, 2000; Fox & Shaw, 1994; Lee 1997; Saracevic & Kantor 1988; see Croft, 2000 for a recent review of combination approaches and studies). Croft describes several aspects of information retrieval (IR) combination: (a) the combination of multiple representations of documents in a single search; (b) the combination of different queries as additional evidence of the searcher's information needs; (c) the combination of ranking algorithms, and (d) the combination of output from different search systems. In the current study we perform the last type of combination, that is, we combine the outputs of two information filtering (IF) systems to maximize user utility.

Received October 20, 2004; revised January 13, 2005; accepted January 28, 2005

 • Published online 7 December 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20278

Information filtering systems seek precisely the relevant documents in an incoming stream of information. This becomes a dual objective: to maximize the relevant information and minimize the nonrelevant information sent to users. Information filtering systems typically support users having long-term information needs, which may be expressed as a profile (Belkin & Croft 1992; Hanani, Shapira, & Shoval, 2001; Oard 1997).

We propose a combination framework consistent with the information structure (IS) model used in information economics to evaluate the value of information (Marschak, 1971). This model traces its origin to fundamental work in statistics and the theory of games, which viewed the design of experiments as a game against nature (Blackwell & Girshick, 1954/1979). The idea of modeling IR as a decision theory model was first presented by Kraft and Bookstein (1978), who also proposed several performance measures for overall retrieval performance. They showed that maximizing precision could be equivalent, under certain conditions, to maximizing the expected value of the IR system. In the IS model users represent their preferences by a payoff matrix (Ronen & Spector, 1995). The model seeks the optimal decision strategy based on those preferences (McGuire & Radner, 1986). Elovici, Shapira, and Kantor (2003) presented a model of IF systems as information systems. This model emphasizes the evaluation of IF systems from the point of view of the benefit to the user, leading to the problem of selecting the optimal decision strategy for each user, i.e., should the user follow the IF system's suggestions.

In this article, we present a comprehensive framework for deriving an optimal combination strategy, to maximize the users' expected payoff, based on the users' defined preferences. The expected payoff represents the welfare of the user from using the combined system, i.e., the cost versus the benefit. The cost is usually the time spent on reading information and the benefit is the value of the relevant knowledge to the user. This approach requires IS modeling of the combined IF systems, and consideration of some properties of the collection. The derived combination strategy, applied to the output of the individual systems, generates the optimal combined output. It is proved analytically and demonstrated empirically that for any two IF systems, and for any user preferences, the user-expected payoff from the properly combined IF systems is always higher than the expected payoff from each of the individual systems.

We have organized this article as follows. In the next section, we explain the relevant background including a brief review of the IS model and its application to IF systems; then we provide an overview of related combination studies. Following this, we present our new combination framework for IF systems and describe an expansion of the framework for dependent IF processes. We proceed to analyze the relation between user preferences and the combined system performance. We also illustrate an empirical application of the framework. We present conclusions and future work issues in the final section.

Background

Information Structure Model: A Brief Review

A brief review on the IS model and common notations are presented in this subsection, A more detailed description can be found in McGuire and Radner (1986) and Blackwell and Girshick (1954/1979).

Let S denote the finite set of Events: $S^t = \{s_1, \ldots, s_{ns}\}$. The a priori probabilities of the events are defined by $\pi = (\pi_1, \ldots, \pi_{ns})$ where, $\sum_{i=1}^{ns} \pi_i = 1, \pi_i \geq 0$. Let Π denote a square matrix where the vector π of a priori probabilities is placed in its main diagonal and zeros placed elsewhere. Let Y denote a finite set of signals: $Y^t = \{y_1, \ldots, y_{nz}\}$ and Q denote an information structure (IS); that is, a Markov (stochastic) matrix of conditional probabilities. Each element $q_{i,j}$ of Q represents the probability that a signal y_j is issued when an event s_i in fact occurs. Let A denote the actions that can be taken by the decision maker (DM): $A^t = \{a_1, \ldots, a_{na}\}$, and let U denote the payoff matrix $U : A \times S$ where each of its elements $u_{i,j}$ represent the payoff for a pairs of an action a_i and an event s_j. Finally, let D denote the DM decision rule that is a Markov matrix where each element of the matrix $d_{i,j}$ represents the probability that the DM will take action a_j when observing the signal y_i. In problems with linear formulation, D will always be deterministic, that is, each element is 0 or 1.

Using the standard von Neumann–Morgenstern formulation of utility theory (1944) The expected payoff to the DM is $EP = tr(QDU\Pi)$ where tr is the matrix trace operator (i.e., the sum of the elements over the main diagonal). The DM can maximize EP by finding, and then applying an optimal decision rule. For this linear utility situation, the optimal decision rule is found by linear programming (Ahituv & Ronen 1988).

In this framework we can compare two or more ISs operating on or responding to the same set of events. We say that one of the ISs is "generally more informative" (GMI) than the other if and only if its maximal expected payoff is never lower than the maximal expected payoff produced by the other IS for all possible payoff matrices and all possible a priori probabilities. The GMI relation imposes a partial ordering on the set of all information structures.

The partial ordering of two ISs is provided by the fundamental result called Blackwell's Theorem (Blackwell, 1951; Blackwell & Girshick 1954/1979; Hilton 1990; McGuire & Radner 1986). An IS Q is GMI than an IS T if and only if there exists a Markov matrix M (sometimes informally called the "garbling" matrix) with appropriate dimensions such that $Q \cdot M = T$. What this means is that Q is GMI than T if and only if the information provided by T can be reformulated as "take the information provided by Q, and degrade it by applying some stochastic process which "garbles the messages." This degradation weakens the connection between the reports provided by T, and the original information as represented by Q. In other words, the information in the messages given by Q is less tightly linked to the underlying reality than is the information in the messages given by T, because of the added stochastic process. Thus, it is not surprising that rigorous analysis, using concepts from game theory, can show the result of Blackwell's theorem.

Modeling Information Filtering Process With Information Structure

Elovici et al. (2003) showed that an IF process can be modeled as an IS. In fact, a filtering process that simply labels items as "relevant or not relevant" can be modeled as a 2×2 IS. Table 1 summarizes the representation of information filtering systems as an information structure.

Related Information Retrieval and Information Filtering Combination Studies

There is strong empirical and theoretical evidence that a combination of information retrieval or filtering systems does improve their performance (Croft, 2000). Different retrieval systems use different retrieval models, document representation, and ranking algorithms, resulting in different retrieved sets of relevant documents that barely overlap (Croft & Harper, 1979; McGill, Koll, & Norreault, 1979; Saracevic & Kantor, 1988). There is no single, ideal retrieval model, representation, or ranking algorithm that finds (a) only, and (b) all the relevant documents in a collection to a given query. A successful combination of retrieval systems would get the most out of each combined system and obtain better retrieval performance than any of them individually. As noted, Croft (2000) has listed several "levels" for combination of evidence. We note in passing that some literature refers to the process of combination as *data fusion*, guided by the structural isomorphism to the corresponding problem in sensor systems, where the procedures described here are generally known as *data fusion processes*.

DOI: 10.1002/asi

TABLE 1. Modeling information filtering (IF) systems with the information structure (IS) model.

IS Model	IF System modeled as information structure
Events (S)	S = {Relevant, Nonrelevant} where "Relevant" stands for a relevant document in the input stream and "Nonrelevant" stands for a nonrelevant document in the input stream. Assumed binary.
A Priori probabilities (π)	π = {G for a Relevant document, ($1-G$) for a nonrelevant document} where G is the measure of the density of relevant documents in the collection. (For IR systems Van Rijsbergen, 1979, defined the *Generality* parameter (G) as the measure of the density of relevant documents in the collection, i.e., the number of relevant documents in a collection divided by the total number of documents).
Signals (Y)	Y = {Flagged, Nonflagged} where "flagged" represents a document that the IF system believes to be relevant and "Nonflagged" stands for a document believed to be nonrelevant. Assumed binary.
Actions (A)	A = {Read, Disregard} where "read" stands for the user action of reading the document and "disregard" stands for not reading the document.
Information structure	The IS modeling the IF system is

IS Events	Signals	
	Read	Disregard
Relevant	R	$1-R$
Nonrelevant	$\frac{R \cdot G - R \cdot G \cdot P}{(1-G) \cdot P}$	$1 - \frac{R \cdot G - R \cdot G \cdot P}{(1-G) \cdot P}$

where R is the IF system recall, P is the system precision.

IS Model	IF System modeled as information structure
Payoff matrix	In the IS model the payoff matrix represents the user return from taking an action (reading/disregarding documents from the input stream). For an IF system modeled as an IS the payoff matrix represents the users estimate of the utility return realized by taking an action. The user defines the cost of reading nonrelevant documents, and of missing relevant documents, as well as the benefit from reading relevant documents. The payoff matrix can be thought of as a part of the user profile because it describes the users' preferences in respect to precision and recall of the filtered information.

U Actions	Events	
	Relevant	Nonrelevant
Read	u11	u12
Disregard	u21	u22

IS Model	IF System modeled as information structure
Decision rule	The decision rule defines what action (read or disregard) the user should take when he or she sees the IF system suggestion/signals (Flagged/Nonflagged) to maximize his or her expected payoff.

D Signals	Actions	
	Read	Disregard
Flagged	u11	u12
Nonflagged	u21	u22

Considering this decision rule is a new feature of the present framework. It enables us to adopt the rigorous models of IS to the filtering problem.

Combination of representation uses multiple sources about the same documents, represented by different methods and structures for a single search (Fuhr & Buckley, 1991; Turtle & Croft, 1991). A number of representation combinations has been investigated and found effective: (a) the combination of index terms and words (Cleverdon, 1967; Fisher & Elchesen, 1972), (b) the combination of citations (or links) and other representation (Croft & Turtle, 1989; Fox, Nunn, & Lee, 1988; Frisse & Cousins, 1989), (c) the combination of passages (parts of a document) with global (whole document) representations (Callan, 1994; Kaszkiel & Zobel, 1997; Knaus, Mittendorf, & Schauble, 1995; Salton & Buckley, 1993), (d) the combination of phrase-based and word-based representations (Bartell et al., 1994), and (e) various combinations of multimedia and text to improve multimedia retrieval (Croft & Turtle, 1992; Fagin, 1996; Harmandas, Sanderson, & Dunlop, 1997).

Moving from the representation to the user, it is also effective to combine different sources of the information needs (queries) because users, (for example, on the Internet) will express their information needs in ways that only approximate "what they exactly want." Various representations

DOI: 10.1002/asi

of an information need such as relevance feedback, expanded queries, or multiple queries for the same search can be combined to improve the representation of the need, leading to better retrieval results. This type of combination improves search results only when each individual query is effective. Thus, a genuinely bad query combined with a good one is less effective than the good query on its own (Belkin, Cool, Croft, & Callan, 1993; Rajashekar & Croft, 1995).

There have been extensive studies of combinations applied to ranking algorithms or to the output of different IR systems operating on the same data collection. This type of combination was introduced during the DAPRA TIPSTER project by Harman (1992) and was later investigated by Kantor (1994, 1995) who used different fusion rules to combine results from several search engines. Fox and Shaw (1994, 246) suggested six particular rules for combining the output of different systems:

1. Take the minimum of the individual scores.
2. Take the maximum of the individual scores.
3. Take the sum of the individual scores.
4. Take the sum of individual scores divided by the number of nonzero similarities.
5. Take the sum of the individual scores multiplied by the number of nonzero similarities.
6. Take the median of the individual scores.

It can be shown that the first two are the generalization to the case of scores of simply logical rules AND and OR. Lee (1995) examined the effectiveness of the above combination rules using six different systems that participated in TREC-3 (Third Text REtrieval Conference, Gaithersburg, Maryland, November 2–4, 1994). He found that the combination of two systems generally gives better retrieval results than either individual system. Lee reports that the sum of individual scores (Rule 3) was the best combination rule when the documents scores were not normalized, the sum multiplied by the number of nonzero scores (Rule 5) was better when the scores were normalized. More complex combination strategies include learning a weighted linear combination (Bartell et al., 1994) or taking a weighted sum (Vogt & Cottrell, 1998, 1999) or weighted average (Thompson, 1993), but results were only slightly better than simple combination (Hull, Pedersen, & Schutze, 1996).

It is also possible to combine results obtained from systems that operate on collection which are not the same. This form of combination is performed on single or distributed architectures and is a component of the operation of "meta search engines" in the Internet environment (Lee, 1995; Vorhees, Gupta, & Laird, 1995).

One important issue in studying fusion is to assess whether two result sets are suitable for fusion. That is, what characteristics of the result sets assure improvement of combined results? Lee (1997) found that overlapping of the result sets was an important factor in the combination success and defined a success measure based on the degree of overlap.

Vogt and Cottrell (1999) sought specific indicators to predict successful combination. They report that for a successful combination the sets of relevant items provided by the individual systems should be similar while the sets of nonrelevant items should be dissimilar. Ng and Kantor (2000) found that the effectiveness of combination is related to two measures: the degree to which the systems were "equally powerful," and the degree to which the specific rankings of documents that they provide are as different as possible. All of these studies, although identifying some characteristics of sets that are associated with successful combination, cannot yet predict (even having the result sets at hand that have not been judged) which queries will benefit from combination of result sets.

All of this work considered batch retrieval systems (the so-called ad hoc task, where a new query is presented to a standing collection). It is possible to model IF combination as a combination of the output of multiple classifiers producing estimates of the probability of relevance. These estimates can then be combined to improve performance (Tumer & Ghosh, 1999). Modeling as probabilistic classifiers is very well suited for filtering systems, which are usually evaluated by linear utility that is conceptually similar to the cost function applied to classifiers (Hull et al., 1996).

Here we discuss a new combination framework for information filtering based on decision theory. This approach is unique in enabling a user to define their preferences and receive in return combination results that maximize the user's expected payoff.

The Decision Theoretic Approach to Combining Information Filters

Assume two (or more) IF systems receiving the same input stream, processing it in parallel. Each system flags documents according to its filtering algorithm, achieving a certain performance measured by precision and recall. The goal is to combine the output of these systems to maximize the user benefit from the combination, as measured by the user's expected utility or "payoff." Note that a combination does not necessarily obtain higher precision or higher recall than each of the individual systems does but will obtain higher expected payoff. Assume, for example, a situation where the user's payoff leads to a combination that maximizes the recall of the combined system, i.e., the combined system returns every document that was returned by either of the systems. Assume a query with 40 relevant documents. System A returned 30 documents out of which 20 were relevant (precision $2/3$, recall $20/40$), system B returned some *other* 30 documents out of which 15 were relevant (precision $1/2$, recall $15/40$). The combined system will return the 60 documents that were returned by both systems (precision is $35/60$, and recall is $35/40$). Recall is better than the recall of either of the systems ($35/40 > 20/40$ of the better system), but the precision ($35/60$) is worse than the precision of the first system ($2/3$), which suits the user's preferences.

DOI: 10.1002/asi

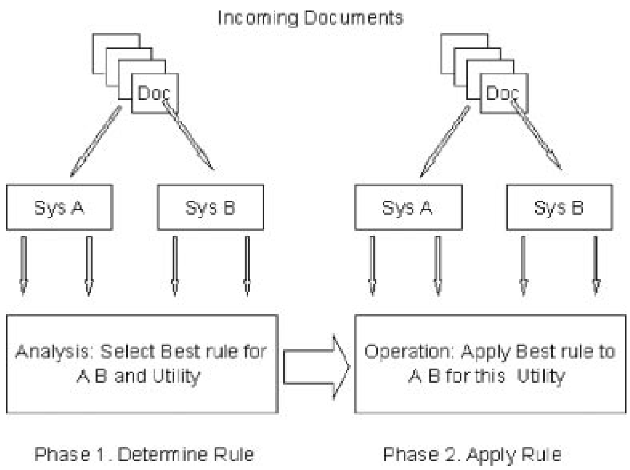

FIG. 1. Information structure-based combination framework.

is GMI than either of the combined ISs. Specifically, the orthogonal product of two ISs is an IS whose events (input) are the same as those for the individual systems (relevant, nonrelevant). The signals (output) of the Orthogonal Product IS are all possible combinations of the two combined ISs signals (That is, the four possible outputs are: Flagged A and Flagged B, Flagged A and Nonflagged B, etc.).

The orthogonal product of two IF systems modeled by IS A and IS B is described below. It will accurately represent the behavior of the combined system if the methods that generate the flags in the two systems are stochastically independent (the dependent case is described in The Effect of User Preferences on the Combined Output section). Recall the entries in the matrix represent the conditional probabilities $Pr(Flag|Truth)$:

$$A \;@\; B = \left[\begin{array}{c|cc} & \text{Flagged A} & \text{Nonflagged A} \\ \hline \text{Relevant} & \text{a11} & \text{a12} \\ \text{Nonrelevant} & \text{a21} & \text{a22} \end{array}\right] @ \left[\begin{array}{c|cc} & \text{Flagged B} & \text{Nonflagged B} \\ \hline \text{Relevant} & \text{b11} & \text{b12} \\ \text{Nonrelevant} & \text{b21} & \text{b22} \end{array}\right]$$

$$= \left[\begin{array}{c|cccc} & \text{Flagged A \& B} & \text{Flagged A \& Nonflagged B} & \text{Nonflagged A \& Flagged B} & \text{Nonflagged A \& B} \\ \hline \text{Relevant} & \text{a11} \cdot \text{b11} & \text{a11} \cdot \text{b12} & \text{a12} \cdot \text{b11} & \text{a12} \cdot \text{b12} \\ \text{Nonrelevant} & \text{a11} \cdot \text{b21} & \text{a11} \cdot \text{b22} & \text{a22} \cdot \text{b21} & \text{a22} \cdot \text{b22} \end{array}\right]$$

The combination framework has two main phases, a training phase and an online phase. During the training phase, the relation between the systems is explored and the optimal combination rule is computed. During the online phase, this rule is applied to the output of the individual systems. The IS-based combination framework is presented in Figure 1.

In the following subsections we review the different phases in detail.

Phase I: Training

During the training phase an optimal combination rule is computed based on the user preferences and the test collection (documents, queries, and relevant judgments for the queries). The test collection parameters are used to compute the precision and recall of each of the filtering systems to model the systems as ISs (see the subsection *Modeling Information Filtering Process With Information Structure*). Note that the computations can be done without ever reducing the data to precision and recall, but we present these for consistency with the general information retrieval literature. The structure models of the IF systems are matrices, as noted above, and are combined by applying the orthogonal operator which we represent by @. The orthogonal operator, which is an outer product of the matrices describing IF systems was used by Ahituv and Ronen (1988). It is a simple consequence of the Blackwell result that the orthogonal product of two ISs

The orthogonal product is represented as a 4×2 matrix, because there are now four possible outputs or flags. An optimal decision rule can be computed for the system represented by this matrix. It is 2×4 matrix, and the entries represent the probability that the combined systems flags, or does not flag, the given item. The sum of the numbers in each row is 1.

$$D = \left[\begin{array}{c|cc} \text{Input Signals} & \text{Flagged} & \text{Nonflagged} \\ \hline \text{Flagged A \& Flagged B} & d_{11} & d_{12} \\ \text{Flagged A \& Nonflagged B} & d_{21} & d_{22} \\ \text{Nonflagged A \& Flagged B} & d_{31} & d_{32} \\ \text{Nonflagged A \& Nonflagged B} & d_{41} & d_{42} \end{array}\right]$$

This optimal decision rule is "how the user should act" upon receiving the signals of the two systems to maximize the expected payoff. Hereafter, this optimal decision rule will be called the *optimal combination strategy*. This optimal combination strategy is applied during the second phase of the framework (online phase) to combine the outputs of the two ISs (as shown in the subsection *Related Information Retrieval and Information Filtering Combination Studies*).

A detailed description of the Phase 1 steps is given below.

Input

- Density of relevant documents (G) for test collection T.
- IF system A with average precision and recall (P^A, R^A) for test collection T

DOI: 10.1002/asi

- IF process B with average precision and recall (P^B, R^B) for test collection T
- User payoff matrix (U)

Output

- Optimal combination strategy (*Optimal Decision Rule* in the IS model terminology)

Step 1—Model IF processes as ISs

Step 1.1—Model the IF system A as an IS based on the precision P^A and the recall R^A:

$$A = \begin{bmatrix} R^A & 1 - R^A \\ \dfrac{R^A \cdot G - R^A \cdot G \cdot P^A}{(1-G)\cdot P^A} & 1 - \dfrac{R^A \cdot G - R^A \cdot G \cdot P^A}{(1-G)\cdot P^A} \end{bmatrix}$$

Step 1.2—Model the IF system B as an IS based on the precision $P^B$and the recall R^B:

$$B = \begin{bmatrix} R^B & 1 - R^B \\ \dfrac{R^B \cdot G - R^B \cdot G \cdot P^B}{(1-G)\cdot P^B} & 1 - \dfrac{R^B \cdot G - R^B \cdot G \cdot P^B}{(1-G)\cdot P^B} \end{bmatrix}$$

Step 2—Combine the IF systems (A & B) modeled as ISs by applying the orthogonal product: $A@B$.

Step 3—Compute the expected payoff for IS ($A@B$) that models the combined system:
$EP(D) = Tr((A@B) \cdot D \cdot U \cdot \pi)$, where D is a pure decision rule.

Step 4—Derive the optimal combination strategy:

Step 4.1—Find the coefficients of all the elements of D in $EP(D)$.

Step 4.2—For each row of D, find the maximal coefficient

$$\forall j \in [1, \ldots, n_y] \quad C_j^{Max} = Max(Coefficient(d_{jt}))$$

Step 4.3—Because the system is linear, the optimal combination rule D^* is a pure decision rule where for each row j of D^* the element d_{ji} with maximal coefficient is set to 1 while the rest of the elements in the row are set to zero. The output of phase 1 is the optimal combination strategy as illustrated below:

$$D^* = \left[\begin{array}{l|cc} \text{Input Signals} & \text{Flagged} & \text{Nonflagged} \\ \hline \text{Flagged A \& Flagged B} & d_{11} & d_{12} \\ \text{Flagged A \& Nonflagged B} & d_{21} & d_{22} \\ \text{Nonflagged A \& Flagged B} & d_{31} & d_{32} \\ \text{Nonflagged A \& Nonflagged B} & d_{41} & d_{42} \end{array}\right]$$

Here is an example that illustrates the training phase:

Assume two systems A and B whose precision and recall when trained on a test collection T are: $R^A = 0.9$, $P^A = 0.8$, $R^B = 0.8$, $P^B = 0.8$ and the density, G, of the documents in T is 0.2.

During Step 1 of Phase 1, the IF systems are modeled as ISs. The following tables present IS modeling of the IF systems A and B, based on their precision and recall.

IS A	Signals	
Events	Flagged	Nonflagged
Relevant	0.9	0.1
Nonrelevant	0.05625	0.94375

IS B	Signals	
Events	Flagged	Nonflagged
Relevant	0.8	0.2
Nonrelevant	0.05	0.95

Assume that the user has the following payoff matrix U that represents his or her preferences (user profile):

U	Events	
Actions	Relevant	Nonrelevant
Read	20	−5
Disregard	−10	0

The user's optimal expected payoffs using each of the IF systems individually is $EP^A = 3.175$, $EP^B = 2.6$.

During Step 2 the orthogonal operator @ is applied on systems A and B to obtain a combined IS.

The orthogonal product of the IF systems A and B modeled as ISs is:

$$A @ B = \left[\begin{array}{l|cc} & \text{Flagged A} & \text{Nonflagged A} \\ \hline \text{Relevant} & 0.9 & 0.1 \\ \text{Nonrelevant} & 0.05625 & 0.94375 \end{array}\right] @ \left[\begin{array}{l|cc} & \text{Flagged B} & \text{Nonflagged B} \\ \hline \text{Relevant} & 0.8 & 0.2 \\ \text{Nonrelevant} & 0.05 & 0.95 \end{array}\right]$$

$$= \left[\begin{array}{l|cccc} & \text{Flagged A \& B} & \text{Flagged A \& Nonflagged B} & \text{Nonflagged A \& Flagged B} & \text{Nonflagged A \& B} \\ \hline \text{Relevant} & 0.72 & 0.18 & 0.08 & 0.02 \\ \text{Nonrelevant} & 0.0028125 & 0.0534375 & 0.0471875 & 0.896563 \end{array}\right]$$

During Step 3 the expected payoff of the combined system ($A@B$) is computed as a function of the decision rule D.

DOI: 10.1002/asi

During Step 4 of Phase 1 the optimal combination strategy is computed for the orthogonal product combination. The result of this step is expressed by the following matrix:

$D^{A@B} =$

	Read	Disregard
Flagged A & Flagged B	1	0
Flagged A & Nonflagged B	1	0
Nonflagged A & Flagged B	1	0
Nonflagged A & Nonflagged B	0	1

The above optimal combination strategy is a logical OR, which is interpreted as follows: To maximize his or her payoff, the user should read a document when both systems recommend reading (Flagged A & Flagged B), or when only one of the systems recommends it. To maximize the expected payoff, the user should refrain from reading a document only when both systems recommend not reading (Nonflagged A & Nonflagged B). If the user follows the optimal combination strategy the expected payoff of the combined system is $EP(A@B) = 3.47$, which is higher than the expected payoffs of each of the individual systems A and B (recall that $EP^A = 3.175$, $EU^B = 2.6$). This result of course confirms the findings of Ahituv and Ronen (1988) regarding the superiority of a combined system $A@B$ over A or B.

Phase 2: Online Process

In this phase, the output streams of the two individual IF systems are combined by applying the optimal combination strategy computed in Phase 1. The following is a detailed description of Phase 2:

Input

- Signals of IF system A (Flagged A, Nonflagged A)
- Signals of IF system B (Flagged B, Nonflagged B)

An optimal combination strategy for the combined IS $A@B$ (the output of Phase 1).

Output

- Flagged and nonflagged signals computed by applying the optimal combination strategy

Step 1—Apply the optimal combination strategy on the output of the individual IF systems by filtering out the documents that are not recommended for reading by the combination strategy. The user is therefore exposed only to documents recommended by the combination strategy. The following table describes how the combination signals are formulated.

	Combination signals	
Signals of systems A & B	Flagged	Nonflagged
Flagged A, Flagged B	If d11 = 1	If d12 = 1
Flagged A, Nonflagged B	If d21 = 1	If d22 = 1
Nonflagged A, Flagged B	If d31 = 1	If d32 = 1
Nonflagged A, Nonflagged B	If d41 = 1	If d42 = 1

Corollary 1: The expected payoff from the combination of IF processes is at least as high as the expected payoff from each of the combined IF processes for every user payoff matrix (user profile).

Proof: Each IF system is represented by an IS. The ISs are combined into one IS using the orthogonal product. The orthogonal product is more informative than each of the ISs, which means that it yields at least as high an expected payoff to the user.

Q.E.D.

Precision and Recall of the Combined Output

In this subsection, we show how to derive the expected precision and recall of the combined output. First, we reduce the representation of the IS describing the orthogonal combination to two columns representing two signals (Flagged, Nonflagged) instead of four namely, Flagged A & Flagged B, Flagged A & Nonflagged B, Nonflagged A & Flagged B, Nonflagged A & Nonflagged B). The first column is a result of the accumulation of all the columns of the IS $A@B$ that represents signals recommending reading a document (based on the optimal combination strategy). The second column is a result of the accumulation all the columns of the IS $A@B$ that represents signals recommending the user to disregard the information.

For example, the orthogonal product of IS A and IS B ($A@B$) is reduced to IS AB

$A@B =$

	Flagged A & B	Flagged A & Nonflagged B	Nonflagged A & Flagged B	Nonflagged A & B
Relevant	0.72	0.18	0.08	0.02
Nonrelevant	0.0028125	0.0534375	0.0471875	0.896563

$D^{A@B} =$

	Read	Disregard
Flagged A & Flagged B	1	0
Flagged A & Nonflagged B	1	0
Nonflagged A & Flagged B	1	0
Nonflagged A & Nonflagged B	0	1

$\Rightarrow AB =$

	Flagged A & B Flagged A & Nonflagged B Nonflagged A & Flagged B	Nonflagged A & B
Relevant	0.98	0.02
Nonrelevant	0.103437	0.896563

DOI: 10.1002/asi

In this example, the combination of the output using the orthogonal product according to the optimal decision strategy, flags 98% of the relevant documents and misses 2%.

It is possible now to derive the precision and recall of the combined (reduced) IS AB. The recall for the above system *AB* is:

$$R^{AB} = Pr(Flagged/Relevant) = AB[1,1] = 0.98$$

The precision of the above system *AB*, is somewhat more complicated and depends on the additional information provided in the generality, *G*:

$$P^{AB} = Pr(Relevant/Flagged) = \frac{Pr(Flagged/Relevant)\cdot Pr(Relevant)}{Pr(Flagged/Relevant)\cdot Pr(Relevant) + Pr(Flagged/Nonrelevant)\cdot Pr(Nonrelevant)}$$

$$= \frac{R\cdot G}{R\cdot G + AB[2,1]\cdot(1-G)} = \frac{0.98\cdot 0.2}{0.98\cdot 0.2 + 0.103437\cdot 0.8} = 0.70314$$

In this subsection we computed "theoretical" (predicted) values for the combined output precision and recall based of the average precision and recall of the individual systems as computed during the training phase.

System Dependence and Information Filtering Combination

The combination framework presented in the above section is based on the assumption that the two combined systems are independent. This is not a realistic assumption because most systems use the same basic retrieval algorithms and document representations with some enhancing modifications. In this section, we present the combination of dependent systems. We show how to derive the orthogonal product when the two systems are dependent.

Assume IF system A and IF system B modeled as ISs A and B.

IS A Events	Signals: Flagged	Signals: Nonflagged
Relevant	d^A	$1 - d^A$
Nonrelevant	f^A	$1 - f^A$

where $d^A = R^A$ and $f^A = \dfrac{R^A\cdot G - R^A\cdot G\cdot P^A}{(1-G)\cdot P^A}$

IS B Events	Signals: Flagged	Signals: Nonflagged
Relevant	d^B	$1 - d^B$
Nonrelevant	f^B	$1 - f^B$

where $d^B = R^B$ and $f^B = \dfrac{R^B\cdot G - R^B\cdot G\cdot P^B}{(1-G)\cdot P^B}$

Given the two systems there can be defined two contingency tables to show how the systems signal (Flagged or Nonflagged). One table is needed for each event. Thus, one contingency table is for the hypothesis "Relevant document," and the other for the hypothesis "Nonrelevant document." If the performance of the individual systems described by the ISs is known, it is possible to compute those tables, but each table still contains one free parameter that represents the similarities in the way the two systems perform for relevant (respectively, not-relevant) documents.

The contingency table for hypothesis "Relevant document" is

Hypothesis Relevant document		System B signals: Flagged B	System B signals: Nonflagged B	
System A signals	Flagged A	x	$d^A - x$	d^A
	Nonflagged A	$d^B - x$	$1 - d^A - d^B + x$	$1 - d^A$
		d^B	$1 - d^B$	1

where x is the probability that both flag simultaneously, given that the document is relevant.

The contingency table for hypothesis "Nonrelevant document" is

Hypothesis Nonrelevant document		System B signals: Flagged B	System B signals: Nonflagged B	
System A signals	Flagged A	y	$f^A - y$	f^A
	Nonflagged A	$f^B - y$	$1 - f^A - f^B + y$	$1 - f^A$
		f^B	$1 - f^B$	1

where y is the probability that both flag simultaneously given a not-relevant document (false alarm).

The orthogonal product of A@B when A and B are dependent and controlled by x and y is

DOI: 10.1002/asi

$$A@B = \left[\begin{array}{l|cccc} & \text{Flagged A \& B} & \text{Flagged A \& Nonflagged B} & \text{Nonflagged A \& Flagged B} & \text{Nonflagged A \& B} \\ \hline \text{Relevant} & x & d^A - x & d^B - x & 1 - d^A - d^B + x \\ \text{Nonrelevant} & y & f^A - y & f^B - y & 1 - f^A - f^B + y \end{array}\right]$$

The "dependent" orthogonal product should be used in Step 2 of Phase 1 if the two systems are dependent. To do this, in Step 1 of Phase 1, the values of *x* and *y* must be evaluated based on the test collection.

The Effect of User Preferences on the Combined Output

In a previous section, we saw how to compute the optimal combination strategy during the training phase. The optimal combination strategy is applied during the online phase to compute an optimal combination output for the user. The optimal combination strategy depends on the system's performance parameters, which may be expressed in terms of precision and recall, on the generality of the collection (G), and on the user payoff matrix (U). In this section the effect of the user payoff matrix U on the precision and recall of the combined output is explored analytically.

Consider the two systems A and B described in the Decision Theoretic Approach section, and assume two different payoff matrices U1 and U2 for two users. Let us suppose that U1 assigns high penalty ($U1_{2,1} = -20$) for disregarding relevant documents and low penalty for reading nonrelevant documents ($U1_{1,2} = -2$). This assignment means that the user prefers high recall. On the other hand, U2 assigns high penalty ($U2_{1,2} = -20$) for reading nonrelevant documents and low penalty for disregarding relevant documents ($U1_{2,1} = -5$), reflecting this user's preference for high precision.

U1 Actions	Events: Relevant	Events: Nonrelevant
Read	20	−2
Disregard	−20	0

U2 Actions	Events: Relevant	Events: Nonrelevant
Read	20	−20
Disregard	−5	0

The optimal combination strategy for user U1 is computed by applying Step 3 and Step 4 of Phase 1:

Step 3 of Phase 1:

$$EP = tr((A@B)DU\Pi) = tr\left[\begin{bmatrix} 0.72 & 0.18 & 0.08 & 0.02 \\ 0.0028125 & 0.0534375 & 0.471875 & 0.896563 \end{bmatrix} \cdot \begin{bmatrix} d11 & d12 \\ d21 & d22 \\ d31 & d32 \\ d41 & d42 \end{bmatrix} \cdot \begin{bmatrix} 20 & -2 \\ -20 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0.2 & 0 \\ 0 & 0.8 \end{bmatrix}\right]$$

Step 4 of Phase 1:

The optimal combination rule that maximizes expected payoff for user U1 is D1:

$$D1 = \left[\begin{array}{l|cc} & \text{Read} & \text{Disregard} \\ \hline \text{Flagged A \& Flagged B} & 1 & 0 \\ \text{Flagged A \& Nonflagged B} & 1 & 0 \\ \text{Nonflagged A \& Flagged B} & 1 & 0 \\ \text{Nonflagged A \& Nonflagged B} & 0 & 1 \end{array}\right]$$

The optimal combination rule of user of type U1 can be interpreted as a simple rule: The user should read a document when either system suggests reading the document (logical OR). Based on the IS $A@B$ and the optimal combination rule D1, the expected precision and recall of the combined output is R = 0.98, P = 0.7. Thus, in the proposed combination framework, a user preferring higher recall will, in fact obtain higher recall for the combined output.

In a similar way, the optimal combination rule for a user of type U2 is computed:

$$D3 = \left[\begin{array}{l|cc} & \text{Read} & \text{Disregard} \\ \hline \text{Flagged A \& Flagged B} & 1 & 0 \\ \text{Flagged A \& Nonflagged B} & 0 & 1 \\ \text{Nonflagged A \& Flagged B} & 0 & 1 \\ \text{Nonflagged A \& Nonflagged B} & 0 & 1 \end{array}\right]$$

The optimal combination rule of user of type U2 can be interpreted as: The user should read a document only when both systems suggest reading the document (logical AND).

Based on the IS $A@B$ and the optimal combination rule D3 the expected precision and recall of the combined output is R = 0.72, P = 0.98. Thus, in the proposed combination framework, the user preferring higher precision obtains higher precision for the combined output.

Empirical Evaluation of the New Combination Framework

An empirical evaluation of the new combination framework was performed. Different IF systems processed the same queries on the same test collection (TREC-6). For the different systems the Lemur 2.1 toolkit (http://www-2.cs.cmu.edu/~lemur/) implementing various IR algorithms was used, and each algorithm was considered as a different IR system. The relevance threshold was set for each system, so that documents

DOI: 10.1002/asi

ranked above the threshold were flagged and those ranked below the threshold were not flagged. Thus, the systems were actually turned into filtering systems (giving Boolean decisions, relevant or nonrelevant) rather than providing rankings. A software tool was developed to perform experimental runs. The software receives as input the parameters for a specific run and performs all the required computations, determining the optimal combination rule given the actual performance of the component systems, and the user's preferences.

Data Set: TREC-6 test collection (over 600,000 documents)
Topics: 32 TREC-6 ad hoc topics that have more than 50 relevant results
Queries: The topics titles were used as queries.
System A: Lemur 2.1, IR algorithm—simple tf-idf with feedback
System B: Lemur 2.1, IR algorithm—okapi with feedback
Generality: *G* was set to 0.00005 based on the number of relevant documents per topic and the number of documents in the collection.
User payoff function: Two different payoff matrices were considered. U1 represents a user preferring high recall while U2 represents users preferring higher precision.

U1	Events	
Actions	Relevant	Nonrelevant
Read	20	−2
Disregard	−20	0

U2	Events	
Actions	Relevant	Nonrelevant
Read	20	−20
Disregard	−5	0

For the precision and recall computation, the first 1000 documents returned from each system were considered.

Our assumption was that the recall and precision of the combined output would improve according to user's preferences, i.e., for U1 the combined output would obtain higher recall than the recall of any of the individual systems, while for U2 the combined output should reveal higher precision than either of the individual systems. We also expected that the expected payoff from the combined output for both users would be higher than the expected payoff from each of the systems. The assumption was based on the analytical analysis presented in the above section.

The evaluation consisted of steps in accordance to the phases of the framework (presented in the Decision Theoretic Approach section). Steps 1–7 refer to the training phase (phase 1) and step 8 refers to the online phase (phase 2).

1. Select 32 topics from the TREC-6 collection with a minimum of 50 relevant documents in the collection.
2. Set a threshold value on the number of documents to be considered relevant (one out of 10, 20, 30, 40, 50)
3. Randomly select 10 topics of the 32 (from Step 1).
4. Run the 10 queries selected in Step 3 using each of the systems A and B. We used the titles of the topics as queries.
5. Compute Precision and Recall (at document points) for each query for each system, using the threshold value as the document point. Also, compute the average precision and recall for each system.
6. Compute optimal decision rule for the combined system *A@B* based on the precision and recall computed at Step 5. We assumed no dependency between the two systems.
7. Run the remaining 22 queries to obtain the set of retrieved documents for each system (22 out of the 32 queries from Step 1, excluding the 10 queries selected at Step 3).
8. Generate combined output by applying the optimal combination strategy (computed in Step 6) on the document sets obtained from A and B (performed during Step 7).
9. Compute average precision and recall (over queries) for each of the systems and for the combined output.
10. Repeat steps 2–9 ten times and average the results, to assure stability.
11. Repeat steps 2–10 for different threshold values.

The averaged results of the evaluation runs for users of type U1 and U2 are shown in Tables 2 and 3, respectively.

In Table 2, as expected, the recall of the combined output is higher than the recall of either of the individual systems for all thresholds, while the combined precision is usually lower than the highest precision obtained by both individual systems. This result suits the preferences of user U1 as reflected by the payoff matrix. We emphasize that this happy outcome is not guaranteed by the dominance theorem, because (a) it is not known that the systems are independent, and (b) it is not known that the optimal combination rule can be determined on one set of topics and validated on another. This positive result suggests that there is a relation between the systems, independent of the topic.

TABLE 2. Results for user profile U1.

	System A		System B		Combined output	
Threshold	Recall	Precision	Recall	Precision	Recall	Precision
10	0.064	0.477	0.066	0.495	0.087	0.475
20	0.113	0.411	0.110	0.425	0.143	0.389
30	0.156	0.383	0.148	0.390	0.190	0.360
40	0.178	0.343	0.172	0.348	0.218	0.315
50	0.195	0.313	0.198	0.328	0.243	0.286

TABLE 3. Results for user profile U2.

	System A		System B		Combined output	
Threshold	Recall	Precision	Recall	Precision	Recall	Precision
10	0.064	0.477	0.066	0.495	0.066	0.495
20	0.113	0.411	0.110	0.425	0.110	0.425
30	0.156	0.383	0.148	0.390	0.114	0.454
40	0.178	0.343	0.172	0.348	0.132	0.430
50	0.195	0.313	0.198	0.328	0.151	0.419

Also for U2 (see Table 3), results match the user's preference. The precision of the combined output is equal or higher than the precision of either individual system for every threshold, while recall is usually lower than the highest recall obtained by the individual systems. This also provides confirmation for the belief that the relation between systems transcends the specific topics considered, at least within the "small world" of TREC topics.

The graph in Figure 2 plots the precision, recall, and odds of relevance as a function of a single parameter *c* that reflects

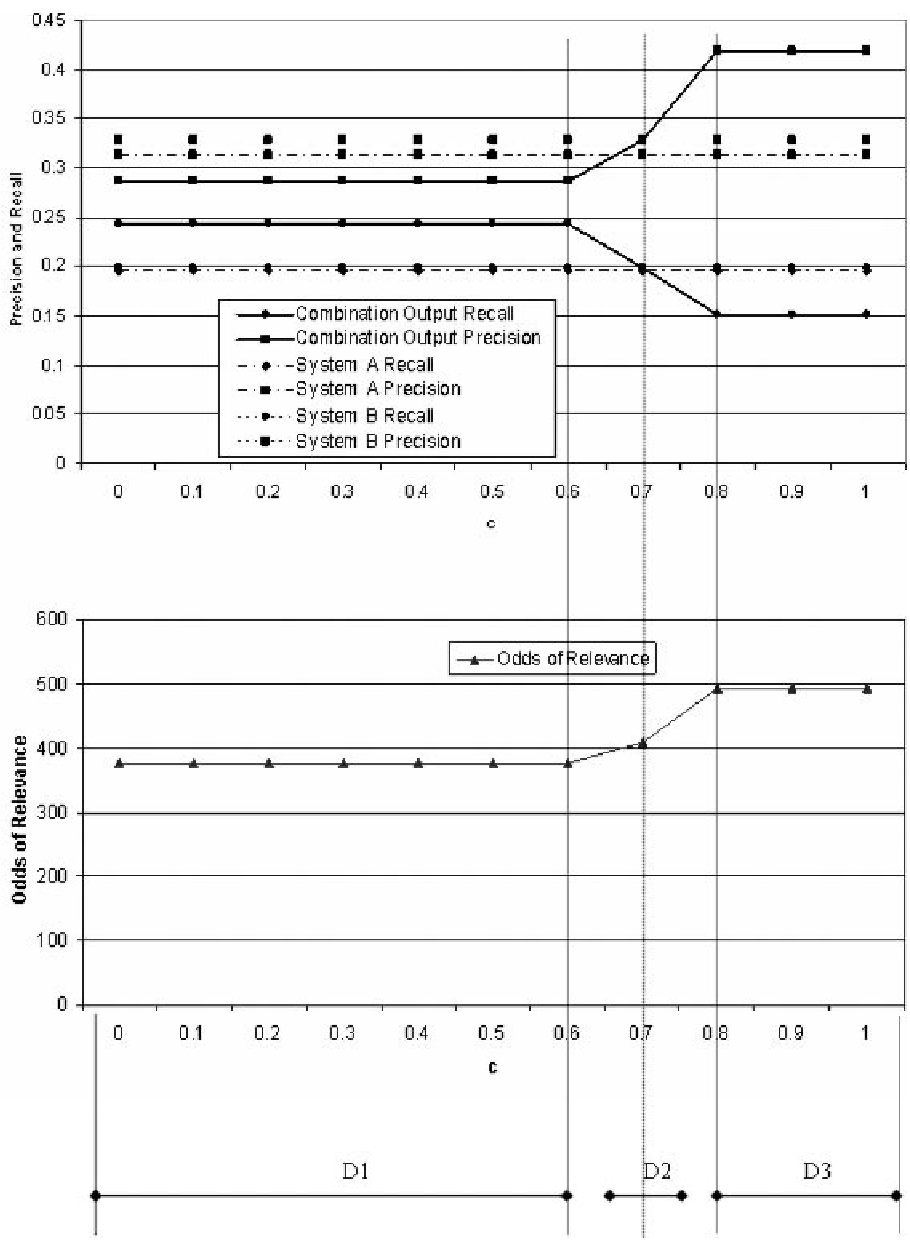

FIG. 2. Precision and recall of optimal combination of Lemur tf-Idf System A and Lemur Okapi System B on TREC-6.

DOI: 10.1002/asi

possible preferences regarding precision and recall. Assume a payoff matrix $U(c)$ whose values depend on a parameter c ranging between 0 and 1. For lower values of c, $U(c)$ represents a user who prefers higher recall while higher values of c represents users who prefers higher precision (as for example, high values for c would result in high penalty for the situation described in $U(c)_{1,2}$, i.e., reading nonrelevant documents, which means high penalty for low precision).

$U(c)$	Events	
Actions	Relevant	Nonrelevant
Read	1	$-c$
Disregard	$c - 1$	0

The top graph in Figure 2 shows the precision and recall of the combined output as a function of c. The graph below plots the odds of relevance (defined as: *Pr*(*Relevant*/*Flagged*/*Pr*(*Relevant*/*Nonflagged*) as a function of c. (The Appendix describes the derivation of the odds of relevance from given P, R, and G).

As observed from the graphs in Figure 2, the value of c affects the values of the odds of relevance, and the precision and recall. This is because c modifies the payoff matrix, which, in turn, directly affects the optimal combination strategy. The optimal combination rule for $c = 0.1, 0, 2, 0, 3, 0, 4, 05, 0, 6$ is D1, for $c = 0.7$ is D2, and for $c = 0.8, 0.9, 1.0$ is D3 where D1, D2, and D3 are:

$D1 =$

	Read	Disregard
Flagged A & Flagged B	1	0
Flagged A & Nonflagged B	1	0
Nonflagged A & Flagged B	1	0
Nonflagged A & Nonflagged B	0	1

$D2 =$

	Read	Disregard
Flagged A & Flagged B	1	0
Flagged A & Nonflagged B	0	1
Nonflagged A & Flagged B	1	0
Nonflagged A & Nonflagged B	0	1

$D3 =$

	Read	Disregard
Flagged A & Flagged B	1	0
Flagged A & Nonflagged B	0	1
Nonflagged A & Flagged B	0	1
Nonflagged A & Nonflagged B	0	1

$D1$ is a combination strategy for users preferring higher recall since it accepts every document that is flagged by any of the systems (OR operation). When $D1$ is the optimal combination strategy the recall of the combined output is higher than the recall of the individual systems (see Figure 2). $D3$ is a combination strategy for users preferring high precision, as only documents that are flagged by both systems are included in the combined set of documents (AND operation). When $D3$ is the optimal combination strategy the precision of the combined output is higher than the precision of the individual systems (see Figure 2). $D2$ represents an intermediate situation in which the user ignores the information from A.

Discussion and Future Issues

The IS-based combination framework presented in this article is related to other frameworks or combination algorithms in the sense that it enables users to best use each of the combined algorithms to obtain improved output. However, the IS-based framework is distinct in that it allows users to define their preferences with regard to precision and recall. It then improves the performance of the combination according to the user's specific preferences yielding different combination rules for different users. We have shown analytically and empirically that for each user preference the combined output provides higher expected payoff. It is important to mention that the interface with multiple systems is unseen for the user because the output that is sent to the user is already the result of the application of the optimal combination rule. The output consists of an optimal set of relevant and nonrelevant documents, which is the standard form of filtering systems' output. The model is designed to combine N systems. However, to simplify the presentation we described here we refer to the combination of two systems. The generalization is straightforward for all the steps of the combination algorithm, the only step that requires some explanation for the generalization to N systems is the application of the orthogonal parameter @ to N. The @ is a binary operator and is applied to n-ary arguments by computing it in steps, i.e., applying it to combine two systems, and then, applying the operator on the result of the combined two systems as one argument, and the third system as the second argument, and so on. Thus, for example, for three systems A, B, C the operator is applied as follows: (A@B)(@c. Because the @ operator is associative: (A@B) @C = A@ (B@C), there is no problem in extending to more systems.

We emphasize that this work has four aspects. The theoretical aspect is really a demonstration that, with the appropriate translation, all of the machinery of information structures can be transferred to the filtering problem. Second, we have built a system that accepts user preferences and carries out all of the needed optimizations, given the matrix describing the combined system. Third, we have shown that, even though there is every reason to expect that systems are correlated, we can treat them as being stochastically independent, computing the performance matrix from precision and recall, and using the orthogonal product. Fourth, we have shown that for the TREC collection, one may train an optimal fusion or combination rule on some topics, and find it to be valid for others as well. Here, validity does not mean that it is the best possible rule for the other topics, but that it provides improvements in precision or recall of exactly the type that the user desires.

One important direction for future research is to consider the effect of the threshold used for each system on the retrieval results. In the current experiment, the same threshold

DOI: 10.1002/asi

was used for both systems. Other possible choices are an optimal threshold for each of the combined systems, or an optimal threshold for systems which we know will be combined. Cherikh (1989) has explored this extremely complex problem in the context of missile detection. We plan to compare the performance of information filtering systems when different settings of the threshold are employed.

Acknowledgments

The authors thanks several agencies for their support (PBK) under Contracts DARPA: N66001-97-C-8537; ARDA: 2002-H-790400-000; ARDA: NBCHC040028; NSF: IIS98-12086; and ITR: 0205178. We acknowledge many helpful conversations with colleagues and students, particularly N. Dalkey, R. Blankenbecler, M. Cherikh, K.B. Ng, and D. Fradkin, and a reviewer for providing a number of helpful comments and questions.

References

Ahituv, N., & Ronen, B. (1988). Orthogonal information structures—A model to evaluate the information provided by second opinion. Decision Sciences, 19, 255–268.

Bartell, B., Cottrell, G., & Belew, R. (1994). Automatic combination of multiple ranked systems. In W.B. Croft & C.J. Van Rijsbergen (Eds.), the Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 173–181). New York: ACM.

Belkin N.J., & Croft, W.B. (1992). Information filtering and information retrieval two sides of the same coin? Communications of the ACM, 35(12), 29–38.

Belkin, N.J., Cool, C., Croft, W.B., & Callan, J.P. (1993). The effect multiple query representations on information retrieval system performance. In R. Korfhage, E. Rasmussen, & P. Willett (Eds.), Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 339–346). New York: ACM.

Blackwell, D. (1951). Comparison of experiments. In J. Neyman (Ed.), the Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (pp. 93–102). Berkeley: University of California Press.

Blackwell, D., & Girshick, M.A. (Ed.) (1979). Theory of games and statistical decisions. New York: Dover. (Original work published 1954)

Callan, J.P. (1994). Passage-level evidence in document retrieval. In W.B. Croft & C.J. Van Rijsbergen (Eds.), Proceedings of ACM-SIGIR94 (pp. 302–309).

Cleverdon, C. (1967). The Cranfield tests on index language devices. ASLIB Proceedings, 19, 173–192.

Cherikh, M. (1989). Optimal decision and detection in the decentralized case. Unpublished doctoral dissertation, Case Western Reserve University, Cleveland, Ohio.

Croft, W., & Harper, D.J. (1979). Using probabilistic models of document retrieval without relevance information. Journal of Documentation 35, 285–295.

Croft, W.B., & Turtle, H. (1989). A retrieval model for incorporating hypertext links. In F. Halasz, & N. Meyrowitz (Eds.), the Proceedings of the ACM Conference on Hypertext (pp. 213–224). New York: ACM.

Croft, W.B., & Turtle, H.R. (1992, March). Retrieval of complex objects. Paper presented at the 3rd International Conference on Extending Database Technology, Vienna, Austria.

Croft, W.B. (Ed.). (2000). Combining approaches to information retrieval. In W.B. Corft (Ed.), Advances in information retrieval (pp. 1–36). Dordrecht: Kluwer.

Elovici, Y., Shapira, B., & Kantor, P. (2003). Using the information structure model to compare profile-based information filtering systems. Information Retrieval 6(1), 75–98.

Fagin, R. (1996). Combining fuzzy information from multiple systems. In R. Hull (Ed.), Proceedings of Fifteenth ACM Symposium on Principles of Database Systems (pp. 216–226). New York: ACM.

Fisher, H.L., & Elchesen, D.R. (1972). Effectiveness of combining title words and index terms in machine retrieval searches. Nature, 238, 109–110.

Fox, E., & Shaw, J. (1994). Combination of multiple searches. In D.K. Harman (Ed.), Proceedings of the Second Text REtrieval Conference (pp. 243–252). Gaithersburg, MD: National Institute of Standards and Technology.

Fox, E.A., Nunn, G.L., & Lee, W. C. (1988). Coefficients for combining concept classes in a collection. In Y. Chiaramella (Ed.), the Proceedings of the Eleventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 291–307). New York: ACM.

Frisse, M.E., & Cousins, S.B. (1989). Information retrieval from hypertext: Update on the dynamic medical handbook project. In F. Halasz, & N. Meyrowitz (Ed.), the Proceedings of Hypertext'89 (pp. 199–212). New York: ACM.

Fuhr, N., & Buckley, C. (1991). A probabilistic learning approach for document indexing. ACM Transactions on Information Systems, 9(3), 223–248.

Hanani, U., Shapira, B., & Shoval, P. (2001). Information filtering: Overview of issues, research and systems. User Modeling and User-Adapted Interaction 11, 203–259.

Harman, D. (1992). The DAPRA TIPSTER project. ACM SIGIR Forum, 26(2), 26–28.

Harmandas, V., Sanderson, M., & Dunlop, M.D. (1997). Image retrieval by hypertext links. In N.J. Belkin, A.P. Narasimhalu, & P. Willett (Eds.), the Proceedings of the Twentieth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 27–31). New York: ACM.

Hilton, R.W. (1990). Failure of Blackwell's theorem under Machina's generalization of expected utility analysis without the independence axiom. Journal of Economics Behavior & Organization 13(2), 233–244.

Hull, D., Pedersen J., & Schutze, H. (1996). Method combination for document filtering. In A. Hopper (Ed.), the Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 279–287). New York: ACM.

Kantor, P.B. (1994). Decision level data fusion for routing of documents in the TREC3 context: A base case analysis of worst case results. In the Proceedings of the Third Text REtrieval Conference (pp. 319–332). Gaithersburg, MD: National Institute of Standards and Technology.

Kaszkiel, M., & Zobel, J. (1997). Passage retrieval revisited. In N.J. Belkin, A.P. Narasimhalu, & P. Willett (Eds.), the Proceedings of the Twentieth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 178–185). New York: ACM.

Knaus, D., Mittendorf, E., & Schauble, P. (1995). Improving a basic retrieval method by links and passage level evidence. Proceedings of the Third Text REtrieval Conference (pp. 247–256). Gaithersburg, MD: National Institute of Standards and Technology.

Kraft, D.H., & Bookstein, A. (1978). Evaluation of information retrieval systems: A decision theory approach. Journal of the American Society of Information Science 29, 31–40.

Lee. J.H. (1995). Combining multiple evidence form different properties of weighting schemes. In A.E. Fox, P. Ingwersen, & R. Fidel (Eds.), the Proceedings of Proceedings of the Eighteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 180–188). New York: ACM.

Lee, J.H. (1997). Analyses of multiple evidence combination. In N.J. Belkin, A.P. Narasimhalu, & P. Willett (Eds.), Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 267–276). New York: ACM.

Marschak, J. (1971). Economics of information systems. Journal of the American Statistical Association, 66, 192–219.

McGill, M., Koll, M., & Norreault, T. (1979, October). An evaluation of factors affecting document ranking by information retrieval systems

(p. 126). Unpublished manuscript, Syracuse University School of Information Studies, Syracuse, NY.

McGuire, C.B., & Radner, R. (Eds.). (1986). Decision and organization. Amsterdam: Elsevier-North Holland.

Ng, K.B., & Kantor, P.B. (2000). Predicting the effectiveness of naïve data fusion on the basis of system characteristics. Journal of the American Society of Information Science and Technology, 51(13), 1177–1189.

Oard W.D. (1997). The state of the art in text filtering. User Modeling and User Adapted Interaction, 7(3), 141–178.

Rajashekar, T.B., & Croft, W.B. (1995). Combining automatic and manual index representations in probabilistic retrieval. Journal of the American Society of Information Science, 46(4), 272–283.

Ronen, B., & Spector, Y. (1995). Evaluating sampling strategy under two criteria. European Journal of Operational Research, 80, 59–67.

Salton, G.J., & Buckley, A.C. (1993). Approaches to passage retrieval in full text information systems. In R. Korfhage, E. Rasmussen, & P. Willett (Eds.), Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 49–58). New York: ACM.

Saracevic, T., & Kantor, P. (1988). A study of information seeking and retrieving. III. Searchers, searches and overlap. Journal of the American Society of Information Science, 39(3), 197–216.

Thompson, P. (1993). Description of the PRC CEO algorithm for TREC-2. In D.K. Harman (Ed.), the Proceedings of the Second Text REtrieval Conference (pp. 271–274). Gaithersburg, MD: National Institute of Standards and Technology.

Tumer, K., & Ghosh, J. (1999). Linear and order statistics combiners for pattern classification. In A.J.C. Sharkey (Ed.), Combining artificial neural nets (pp. 127–161). London: Springer-Verlag.

Turtle, H., & Croft, W.B. (1991). Evaluation of an inference network-based retrieval model. ACM Transactions on Information Systems, 9, 187–222.

Van Rijsbergen, C. (Ed). (1979). Information retrieval (2nd ed.). London: Butterworth.

Vogt, C., & Cottrell, G. (1998). Predicting the performance of linearly combined IR systems. In W.B. Croft, C.J. Van Rijsbergen, R. Wilkinson, & J. Zobel (Eds.), Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 172–179). New York: ACM.

Vogt, C.C., & Cottrell, G.W. (1999). Fusion via a linear combination of scores. Information Retrieval, 1(3), 151–173.

Voorhees, E., Gupta, N.K., & Laird, B. (1995). Learning collection fusion strategies. In A.E. Fox, P. Ingwersen, & R. Fidel (Eds.), the Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 172–179). New York: ACM.

Von Neumann, J., & Morgenstern, O. (Ed.). (1944). Theory of games and economic behavior. Princeton, NJ: Princeton University Press.

DOI: 10.1002/asi

Appendix

Odds of Relevance as Function of P, R, and G

The odds of relevance (OOR) as function of precision P, recall R, and density G is derived below:

$$OOR = \frac{Pr(Relevant/Flagged)}{Pr(Relevant/Nonflagged)}$$

$$Pr(Relevant/Flagged) = \frac{Pr(Flagged/Relevant) \cdot Pr(Relevant)}{Pr(Flagged/Relevant) \cdot Pr(Relevant) + Pr(Flagged/Nonrelevant) \cdot Pr(Nonrelevant)} = P$$

$$P = \frac{R \cdot G}{R \cdot G + Pr(Flagged/Nonrelevant) \cdot (1 - G)}$$

$$\Downarrow$$

$$Pr(Flagged/Nonrelevant) = \frac{R \cdot G - R \cdot G \cdot P}{(1 - G) \cdot P}$$

$$Pr(Nonflagged/Nonrelevant) = 1 - Pr(Flagged/Nonrelevant) = 1 - \frac{R \cdot G - R \cdot G \cdot P}{(1 - G) \cdot P}$$

$$Pr(Relevant/Nonflagged) = \frac{Pr(Nonflagged/Relevant) \cdot Pr(Relevant)}{Pr(Nonflagged/Relevant) \cdot Pr(Relevant) + Pr(Nonflagged/Nonrelevant) \cdot Pr(Nonrelevant)}$$

$$= \frac{(1 - R) \cdot G}{(1 - R) \cdot G + \left(1 - \frac{R \cdot G - R \cdot G \cdot P}{(1 - G) \cdot P}\right) \cdot (1 - G)}$$

$$= \frac{(1 - R) \cdot G \cdot P}{(1 - R) \cdot G \cdot P + ((1 - G) \cdot P - R \cdot G - R \cdot G \cdot P) \cdot (1 - G)}$$

$$\frac{Pr(Relevant/Flagged)}{Pr(Relevant/Nonflagged)} = \frac{P}{\frac{(1 - R) \cdot G \cdot P}{(1 - R) \cdot G \cdot P + ((1 - G) \cdot P - R \cdot G - R \cdot G \cdot P) \cdot (1 - G)}}$$

$$= \frac{(1 - R) \cdot G \cdot P + ((1 - G) \cdot P - R \cdot G \cdot - R \cdot G \cdot P) \cdot (1 - G)}{(1 - R) \cdot G}$$

$$\frac{Pr(Relevant/Flagged)}{Pr(Relevant/Nonflagged)} = \frac{P}{\frac{(1 - R) \cdot G \cdot P}{(1 - R) \cdot G \cdot P + ((1 - G) \cdot P - R \cdot G - R \cdot G \cdot P) \cdot (1 - G)}}$$

$$= \frac{(1 - R) \cdot G \cdot P + ((1 - G) \cdot P - R \cdot G \cdot - R \cdot G \cdot P) \cdot (1 - G)}{(1 - R) \cdot G}$$

DOI: 10.1002/asi