Relevance and Utility:
A Comparison Using the Retrieval Situation Model

Working Paper No. 346

William C. Sasso

The University of Michigan

# Relevance and Utility:

# A Comparison Using the Retrieval Situation Model

William C. Sasso

Graduate School of Business Administration
The University of Michigan
Ann Arbor, MI 48109

## Abstract

Two concepts, relevance and utility, have been proposed as basic criteria for the evaluation of document retrieval system effectiveness. To compare their relative merits, a model of the document retrieval process is developed. This model, called the Retrieval Situation Model (RSM), is an expansion of Bookstein and Cooper's model of a document retrieval situation [1]. The concept of utility and several definitions of relevance are compared using the RSM. The paper concludes that utility is the most practical basis for actual system evaluation, and questions the value of such traditionally accepted measures of retrieval system effectiveness as recall and precision.

## Introduction

A document retrieval system is intended to facilitate the location of material <u>relevant</u> to an inquirer's information need. Relevance, then, is a central concept in the comparison and evaluation of these systems [2, p. 6]. Cleverdon, Mills, and Keen [3] proposed six criteria for the evaluation of document retrieval systems. The two criteria which gauge the effectiveness of the system, precision and recall[1], can be assigned values only if we are able to distinguish relevant documents from those not relevant. To determine a precision value, we need only make this distinction within the set of documents retrieved by the system. In order to establish a value for recall, however, we must be able to partition the entire document collection into sets of relevant and not-relevant documents. To use precision and recall as criteria for document retrieval system evaluation, we must define relevance and operationalize its definition.

As Wilson [4, p. 457] has noted, " relevance is a highly general and vague notion that can be made specific and precise in a large number of ways." It is not surprising, therefore, that the literature shows a surplus, rather than a shortage, of definitions of relevance. The historical

---

[1]Precision is the ratio between the number of relevant documents retrieved and the number of documents retrieved; Recall is the ratio between the number of relevant documents retrieved and the number of relevant documents in the collection. An ideal document retrieval system would have both precision and recall equal to one.

development of the concept is well described by Saracevic [5], whose paper goes into far greater detail on the subject than is possible here.

This paper will review and expand a mathematical model of a document retrieval system proposed by Bookstein and Cooper [1]. The extended model will be used as a vehicle for the comparison of several important definitions of relevance, including Cooper's logical relevance [6], Wilson's situational relevance [4], and Bookstein's relevance [7]. Moreover, because Cooper [6,8,9] has articulately championed the use of utility, rather than relevance, as the basis for the evaluation of retrieval system effectiveness, utility will also be related to the extended model and compared to the three definitions of relevance.

This paper, then, is organized into three major parts. In the first part, the Bookstein-Cooper model is introduced and expanded. In the second, the three definitions of relevance are presented and related to the expanded model. The competitor concept, utility, is similarly treated. In the final part, the results of part two are compared, and some general conclusions are presented.

## The Retrieval Situation Model

Bookstein and Cooper [1] have proposed a mathematical model of a document retrieval system. The first section of this part presents their model. In the second section we expand it both backward, so that it includes the processes by

which the index record and the user's request are generated, and forward, so that it includes the interaction of the user with the retrieved document(s). The expanded model, of which Bookstein and Cooper's model is a component, will be referred to as the Retrieval Situation Model, or RSM.

We note here that many of the desirable mathematical qualities which hold for the Bookstein-Cooper model do not hold for the RSM. This is due primarily to the vagaries of human action, which the Bookstein-Cooper model excludes. While perhaps unfortunate, we do not consider this lack of mathematical rigor debilitating with respect to the purpose of this paper.

## Bookstein and Cooper's Model

Bookstein and Cooper define a document retrieval system, $S_{bc}$[2], as a quadruple

$$S_{bc} = \{I, R, V, T\}, \text{ where}$$

$I$ is the set of index records or document representations used by the document retrieval system for manipulation. These representations can take such forms as a set of index terms, a binary vector representing a uniterm classification, the document's abstract, or the document itself. Whatever the form of representation, the set $I$ contains one such representation for each document in the sys-

[2]A list identifying all symbolic representations used in this paper will be found in Appendix A.

tem.

$R$ is the set of user requests or system-manipulable query formulations. The user of the system has expressed his information need in a natural language query and this query has been transformed into a system-manipulable request. The set $R$ consists of a all conceivable system-manipulable requests.

$V$ is the set of retrieval status values which the system can return for a given document in response to a particular request. In the simplest case, $V$ will consist of only two values, retrieved and not retrieved. A more sophisticated system may return a value between 0 and 1, reflecting an attempt to predict the probability that the user will find a document relevant to his request. Alternatively, this value can be considered an assessment of the degree to which the document is relevant to the request.

$T$ is the retrieval function. Given elements $r_m$ in $R$ and $i_n$ in $I$, $T$ maps each pair $(r_m, i_n)$ to some value $v$ in $V$. For a given request $(r_m)$ and a particular index record $(i_n)$, the function $T$ determines a unique value from $V$, "indicating the degree to which it is predicted that the document represented by the index record will be found relevant to the request by the patron" [1, p. 155].

Bookstein and Cooper summarize the operation of their model as follows:

> Given a request, $r$, the function $T_r$ uniquely breaks the set of index records, $I$, into a set of subsets and induces a simple ordering on these subsets. This process can be used in a search of

the collection by means of the following idealized
search procedure. To search the collection, a user
begins by examining the "first" subset of docu-
ments; if he wishes, he continues by searching the
"next" subset of documents, etc. Each subset is
searched randomly. The system operates by reducing
a random search of the full collection to random
searches of the much smaller subsets defined
above; these subsets are presumed to be enriched
with relevant documents. [1, pp. 156-157]

Our representation of the Bookstein-Cooper model is
shown in Figure I.

## Figure I.

### The Bookstein-Cooper model

$R$          $I$

$T$

$V$
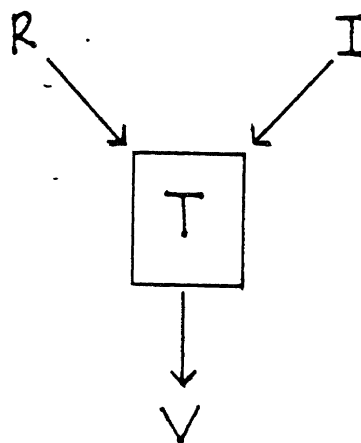
Although the Bookstein-Cooper model takes $I$ as a
given, the process of creating an index record is dis-
cussed. The index record is derived from its source docu-
ment, and the derivation process entails a significant in-
formation loss. This loss is necessary in order to reduce
the index record to a system-manipulable size and for-
mat. The reduction is usually accomplished via an indexing
procedure, with syntactic, structural, and semantic com-
ponents.

The syntax establishes the set of legitimate index terms which, whether natural language or abstract, may be used to build the document's index record. The procedure's structure defines the relationships between different terms; it may be either explicit or implicit. An explicit structure is overtly defined in the system's documentation, while an implicit one is determined by the de facto assignment of terms to documents. Similarly, the structure may be hierarchical, like the Dewey Decimal System, or uniterm. The semantics of the indexing language are the rules suggesting how each index term is to be applied. The assignment of index terms to a document determines the retrievability of that document, so the process of assigning a term to a document should be based on the expected utility of the document for inquiries under the term.

Bookstein and Cooper note that a similar transformation is generally performed on the user's query to formulate the system-manipulable request.

## Extensions of the Bookstein-Cooper Model

In this section we extend the Bookstein-Cooper model by adding to it descriptions of the processes by which

    (1) the document was created,

    (2) the document was processed to generate an index record,

    (3) the user's information need was created and expressed as a natural language query,

    (4) the natural language query was formulated into a system-manipulable request, and

(5) the output of the retrieval system was used to
modify the user's state of knowledge and (possi-
bly) the information need representation in the
form of the request.

We assume that a document can be described as follows:
a document is created by an author, working in a particular
environment, in order to express an identifiable idea. These
three variables, author, environment, and idea, determine a
document.

Environment is a complex variable, and includes the
physical, emotional, and intellectual situation in which the
author works. Thus a change in the environment may cause a
given author to express the same idea in a significantly
different document. Consider, for example, the change in
tone of White House statements on the Watergate break-in
over the period June, 1972 to June, 1974. Although the
author, Ron Ziegler, and the idea, the denial of high-level
White House involvement, remained the same, the tone of the
communications became far more defensive and uncertain.

Similarly, the author is a complex factor. In an in-
variant environment, different authors will create different
documents in order to express the same idea, even in a con-
stant environment. Furthermore, the same author has the
ability to produce different expressions of the same idea.
As evidence supporting this assertion we offer the first
draft of any important document, complete with alterations,
additions, and deletions. This factor we call the author in-
cludes his deductive and inductive reasoning powers, his in-

tellectual and experiential knowledge, his current emotional and physical state, his ability to express his ideas clearly, and his concern to do so in the particular case at hand.

The document is an expression of an idea or of a solution to a particular problem. Where does this idea (or solution) come from? We assume that the idea is the product of an ongoing interaction between the author and his environment. The author and environment continually interact with each other, in a fashion similar to that of an axe being sharpened on a grindstone. The idea is analogous to the spark generated by the sharpening process.

Using a notation similar to that of Bookstein and Cooper, we define a document creation system, $S_{dc}$, as a sextuple $\{A, E, Id, T_i, D, T_e\}$ where

> $A$ is the set of all potential authors, individuals capable of creating a document,
>
> $E$ is the set of all possible environments in which an author can work.
>
> $Id$ is the set of all possible ideas.
>
> $T_i$ is a relation taking a point in $A \times E$ to one or more points in $Id$; $T_i$ is the "interaction" relation $T_i(a,e) \dashrightarrow id$ in $Id$.
>
> $D$ is the set of all possible documents.
>
> $T_e$ is a relation taking a point in $(A \times E \times Id)$ to one or more points in $D$; it is the "expression" relation $T_e(a,e,id) \dashrightarrow d$ in $D$).

A representation of this system is shown in Figure II.

$$A \longleftrightarrow T_i \longleftrightarrow E$$
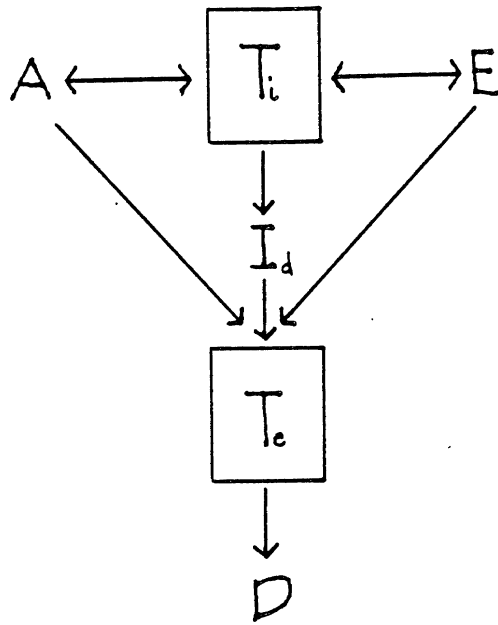
$$I_d$$

$$T_c$$

$$D$$

**Figure II.**

**The Document Creation System** $(S_{dc})$

The document so derived must now be indexed; it must be transformed into a system-manipulable description, the index record. We assume that a classifier, working in a specific organizational context, with a particular indexing procedure, accepts the document and creates an index record for it. Thus, four factors determine the index record: the document, the classifier, the organizational context, and the indexing procedure.

Like the author and the environment in the document creation system, the classifier and the organizational context are complex variables capable of producing unpredictable variations in the index records produced. The most significant new question predicted by the index creation system is that of how the indexing procedure should be con-

ceptualized. Is it another input to the process which creates the input record, or is it the process itself? For simplicity's sake, we assume that the indexing procedure is the index record creation transformation $T_{ic}$. We define the index record creation system, $S_{ic}$, as a sextuple $\{D,\ C,\ O,\ T_i,\ I,\ T_{ic}\}$ where

> $D$ is the set of all documents to be represented in the system, and is the output of $S_{dc}$ above.

> $C$ is the set of all possible classifiers who generate index records for use in system $S_{bc}$.

> $O$ is the set of all possible organizational contexts, in which classifiers operate on documents to create index records.
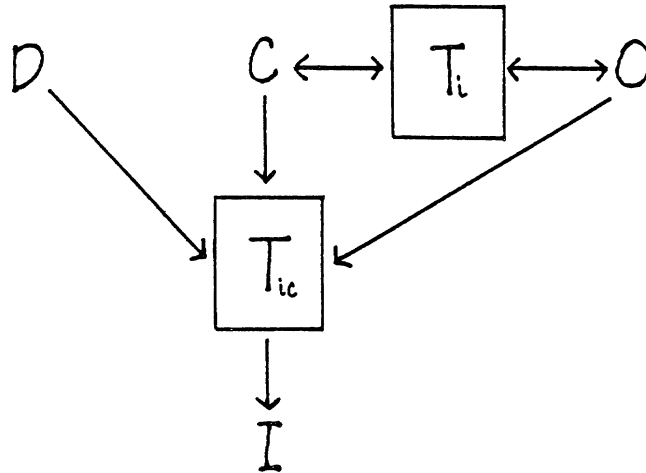
> $T_i$ is the interaction relation as in $S_{dc}$ above.

> $I$ is the set of all index records present in $S_{bc}$.

A representation of the index record creation system is presented in Figure III.

We now wish to represent the generation of the user's information need and its evolution into a system-manipulable request, and include this representation as part of our model. Fortunately, it seems reasonable to assume that there is a strong analogy between the process of generating and expressing ideas and that of generating and expressing information needs. We substitute the user for the author, the need for the idea, and the query for the document. Now we are ready to define a system in which a query is generated, $S_{qg}$, as a sextuple $\{U,\ E,\ N,\ T_i,\ Q,\ T_{qg}\}$ where

## Figure III.

### The Index Record Creation System $S_{ic}$



$U$ is the set of all possible users,

$E$ is the set of all possible environments in which a users can work,

$N$ is the set of all possible information needs,

$T_i$ is the interaction relation, defined as in $S_{dc}$ above,

$Q$ is the set of all possible natural language queries, which express an information need $n$ in $N$, and

$T_{qg}$ is a relation taking points in $U \times E \times N$ to one or more points in $Q$; is is the "query generation" relation $T_{qg}(u,e,n) \text{---}>q$ in $Q$.

This system is depicted in Figure IV.

In similar fashion, the system which formulates requests from natural language queries is directly analogous to the index record creation system. Analogous to document is query, to classifier is formulator, and to index record

## Figure IV.

### The Query Generation System $S_{qg}$

$$U \longleftrightarrow \boxed{T_L} \longleftrightarrow E$$

$$\downarrow$$

$$N$$

$$\downarrow$$

$$\boxed{T_{qg}}$$

$$\downarrow$$

$$Q$$
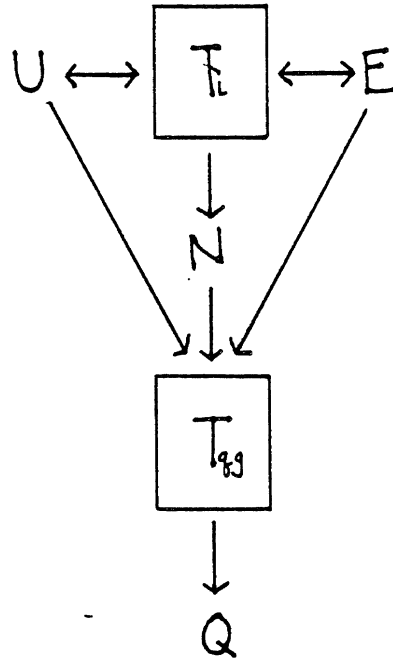
is request. This enables us to define the request formulation system, $S_{rf}$, as the sextuple $\{Q, F, O, T_i, R, T_{rf}\}$ where

$Q$ is the set of all possible queries produced by $S_{qg}$ above,

$F$ is the set of all people who formulate requests for $S_{bc}$,

$O$ is the set of all possible organizational contexts in which a formulator transforms a query into a request,

$T_i$ is the interaction relation defined in $S_{dc}$ above,

$R$ is the set of all possible system-manipulable requests in $S_{bc}$, and
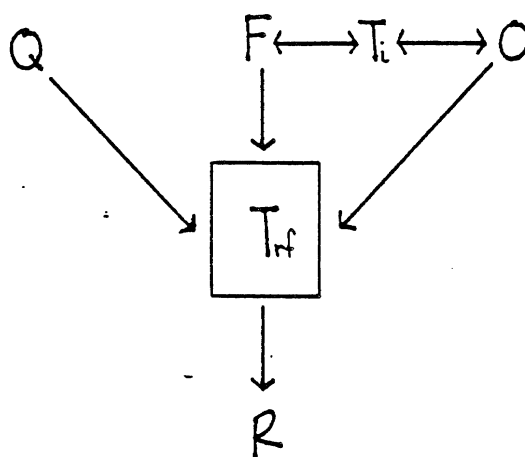
$T_{rf}$ is a relation taking a point in $Q \ X \ F \ X \ O$ to one or more points in $r$; it is the "request formulation"

relation $T_{rf}(q,f,o) \text{--->} r$ in $R$.

This system is depicted in Figure V.

## Figure V.

### The Request Formulation System $S_{rf}$



As has been noted earlier, the retrieval system $S_{bc}$, produces output in the form of ordered subsets of index records. We assume that these subsets are related back to their source documents, which are then presented to the user. These documents become part of his environment, and their content may become part of his intellectual knowledge. Thus, we define the information use system, $S_{iu}$, as the sextuple $\{V, D^*, T_{iu}, U, E, T_i\}$ where

$V$ is the set of retrieval status values returned by $S_{bc}$,

$D^*$ is the set of retrieved documents presented to the user,

$T_{iu}$ is a relation taking a value $v$ in $V$ to a document

$d^*$ in $D^*$; it is the information use relation

$T_{iu}(v) \text{---}> d^*$ in $D^*$,
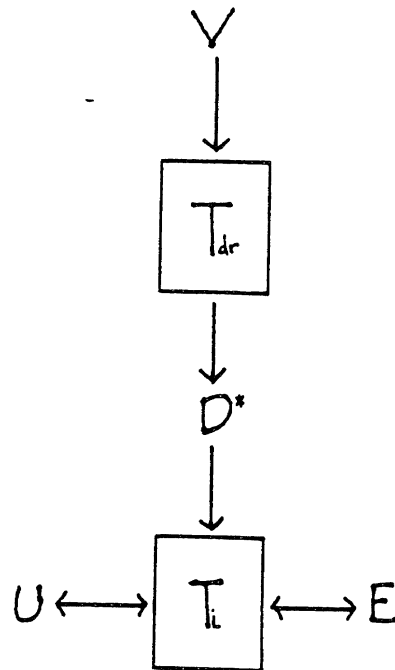
$U$ is the user,

$E$ is the user's environment, and

$T_i$ is the "interaction" relation, defined in $S_{dc}$ above.

The information use system $S_{iu}$ is presented in Figure

VI.

## Figure VI.

### The Information Use System $S_{iu}$



The Retrieval Situation Model is presented in its en-
tirety in Figure VII. We note here that all transformation
(denoted $T_x$ or $T_{xx}$) are defined as relations because they do
not determine a unique point in their range for any given
point in their domain. There is, for example, no unique $id$

in *Id* to which constant values $a_c$ in *A* and $e_c$ in *E* are mapped by $T_i$. Similarly, there is no unique *d* in *D* to which constant values $a_c$ in *A*, $e_c$ in *E*, and $id_c$ in *Id* are mapped by $T_e$. This is perhaps unfortunate, but it makes the model far more representative of the human world, and matters little for the purposes of this paper.

Swanson [10] has discussed the trial and error nature of the document retrieval process. The RSM can very easily be modified to represent this feature of the retrieval process. If we regard the model presented in Figure VII as one cycle of an ongoing process, where the user-environment interaction becomes a node connecting both component $S_{iu}$ of the previous cycle with component $S_{qg}$ of the current cycle, and make a similar connection between $S_{iu}$ of the current cycle and $S_{qg}$ of the next cycle, then Figure VII becomes one link in a chain of RSMs, depicting an iterative retrieval process. In fact, the user-environment interaction in $S_{iu}$ can overlap the author-environment interaction in $S_{dc}$ as well. That is to say, there is no reason why those who generate queries cannot generate documents (and vice versa). This we will regard as a special case; it therefore is depicted in Figure VIII as a dashed line, while the user-environment interaction will be shown as a solid line.

## Relevance and Utility in the RSM

In this part of the paper, several important definitions of relevance and the concept of utility are reviewed
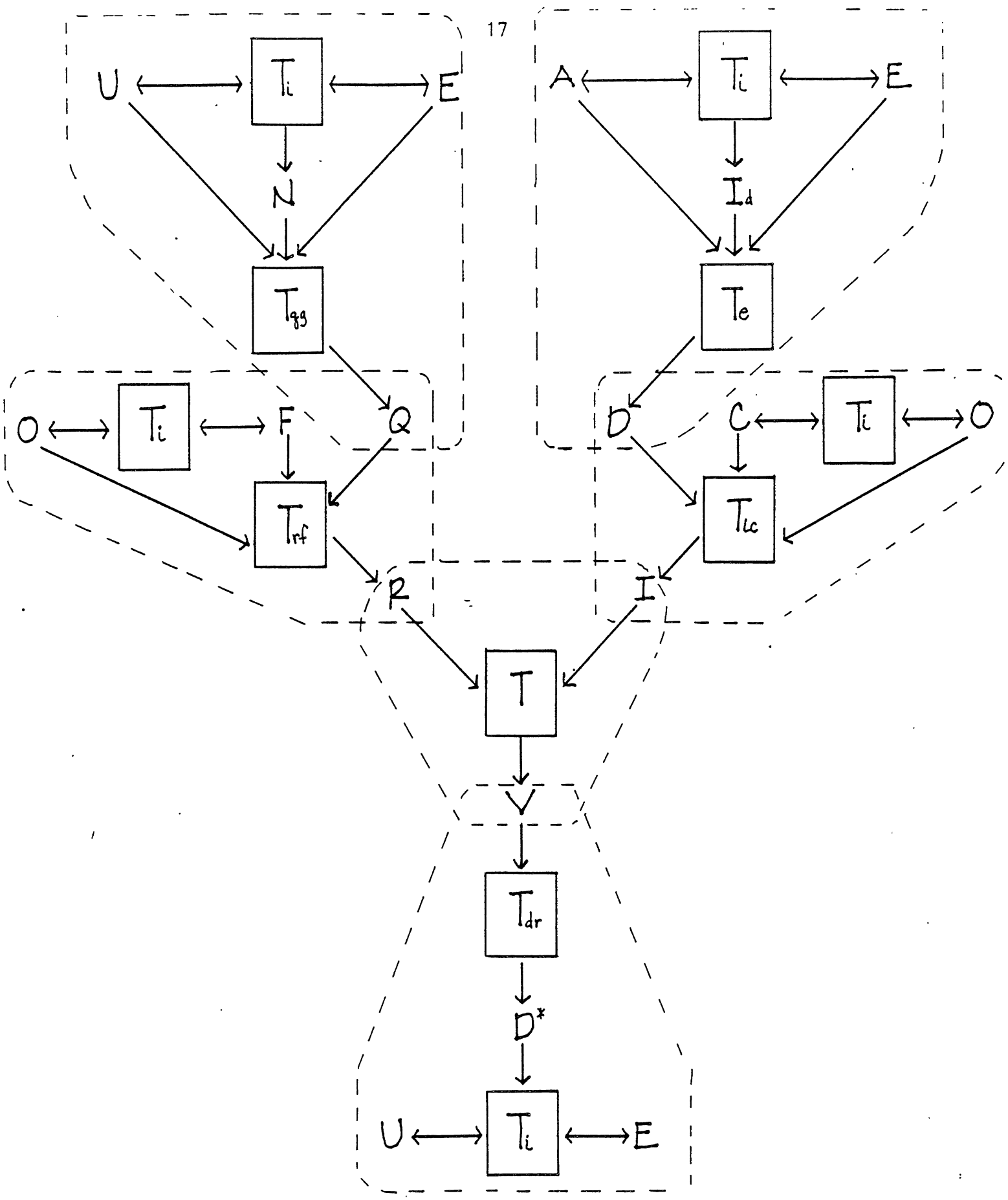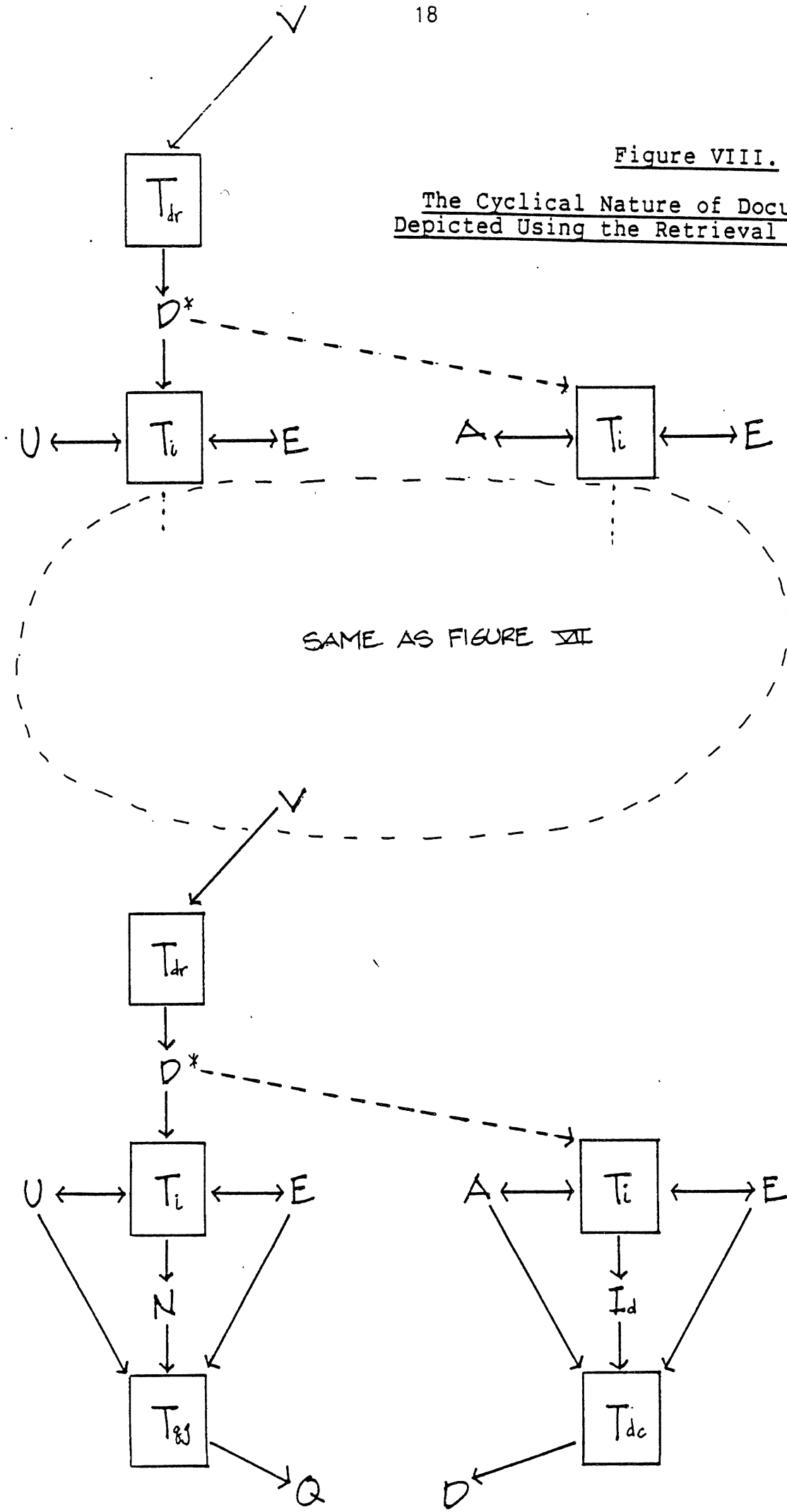
Figure VII.

The Retrieval Situation Model (RSM)

Figure VIII.

The Cyclical Nature of Document Retrieval
Depicted Using the Retrieval Situation Model

and related to the Retrieval Situation Model. This part of the paper consists of four sections. The first deals with Cooper's logical relevance, while the second considers Wilson's situational relevance. Bookstein's relevance is treated in the third section, and the final one examines Cooper's concept of utility.

## Cooper's Logical Relevance

Cooper [6] proposes a definition of logical relevance, based on the well-investigated concept of logical consequence, as a response to the plethora of existing definitions. These he considers unsatisfactory because they define relevance using " ... terms ... no less mysterious than the term being defined. " [6, p. 20]. To avoid this, Cooper bases his concept of logical relevance on the bedrock of logical consequence.

Beginning with the assumption that relevance is a type of relation between some piece of stored information and an information need within the mind of the user, Cooper proposes a series of stages through which the information need evolves during the retrieval process. The information need itself is a psychological state; it is unobservable, for it exists only in the user's mind. This information need may be described in words, but the words, however well they may describe the need, are not the same thing as the need. The words form a query, a user's description of his need in natural language. This query, in turn, is trans-

formed into a request, a system-manipulable statement intended to represent the original information need.[3] Note, however, that neither the query nor the request is necessarily a complete and accurate representation of the information need. Cooper therefore defines the information need representation as a complete, accurate, and concrete (preferably linguistic) expression of the information need.

On the one hand, then, we have the user's information need perfectly expressed by the information need representation. On the other, we have a set of stored pieces of information, e.g., a set of documents containing facts. Our definition of relevance, then, should enable us to select those pieces of information which answer (wholly or partially) the information need.

In order to employ the concept of logical relevance in a rigorous fashion, Cooper is forced to make the three simplifying assumptions which follow.

> Restriction 1: The search query is essentially a yes-or-no type question, or what amounts to the same thing, a true-or-false question.

> Restriction 2: The data stored in the system are stored in the form of well-formed sentences of one of the well-formed formalized languages such as the classical first-order predicate calculus.

> Restriction 3: The retrieval system is an inferential one, in the sense that it deduces a direct answer to input questions. In the case of yes-or-no questions it gives the answers "Yes", "No", or "Don't Know" ... [6, p. 23]

Under these conditions, Cooper proposes a restricted

---

[3]The RSM terminology is intended to conform to Cooper's terminology and scheme of information need manifestations.

definition of logical relevance which we paraphrase as follows.[4]   A stored piece of information is logically relevant to a (completely and accurately expressed) information need if and only if it is a member of the smallest possible set of stored pieces of information which satisfy the information need. Extrapolating from pieces of information to documents, we note that a document is logically relevant if and only if it is a   member of the smallest possible set of documents which satisfies the information need. This smallest possible set need not be unique; there may, for example, be two (or more) sets of $n$ pieces of information which enable us to answer the question "Is a senior faculty member teaching course X next term?"

The restricted definition of logical relevance, then, is a relationship (a particular relationship, that of logical consequence) between an information need and a set of stored pieces of information. Figure IX. depicts this relationship. Further, because of Cooper's assumption that an information need can be expressed perfectly and accurately, we can consider logical relevance to be a relationship between either the query or the request and the set of stored pieces of information. These relationships are also shown in Figure IX.

Cooper notes that the relaxation of the three restrictions causes some problems. Introduction of the use of

---

[4]A paraphrase is presented because Cooper's definition is phrased in the terminology of formal logic. The interested reader is encouraged to consult the original [6].

# Figure IX.

## Restricted Logical Relevance in the RSM
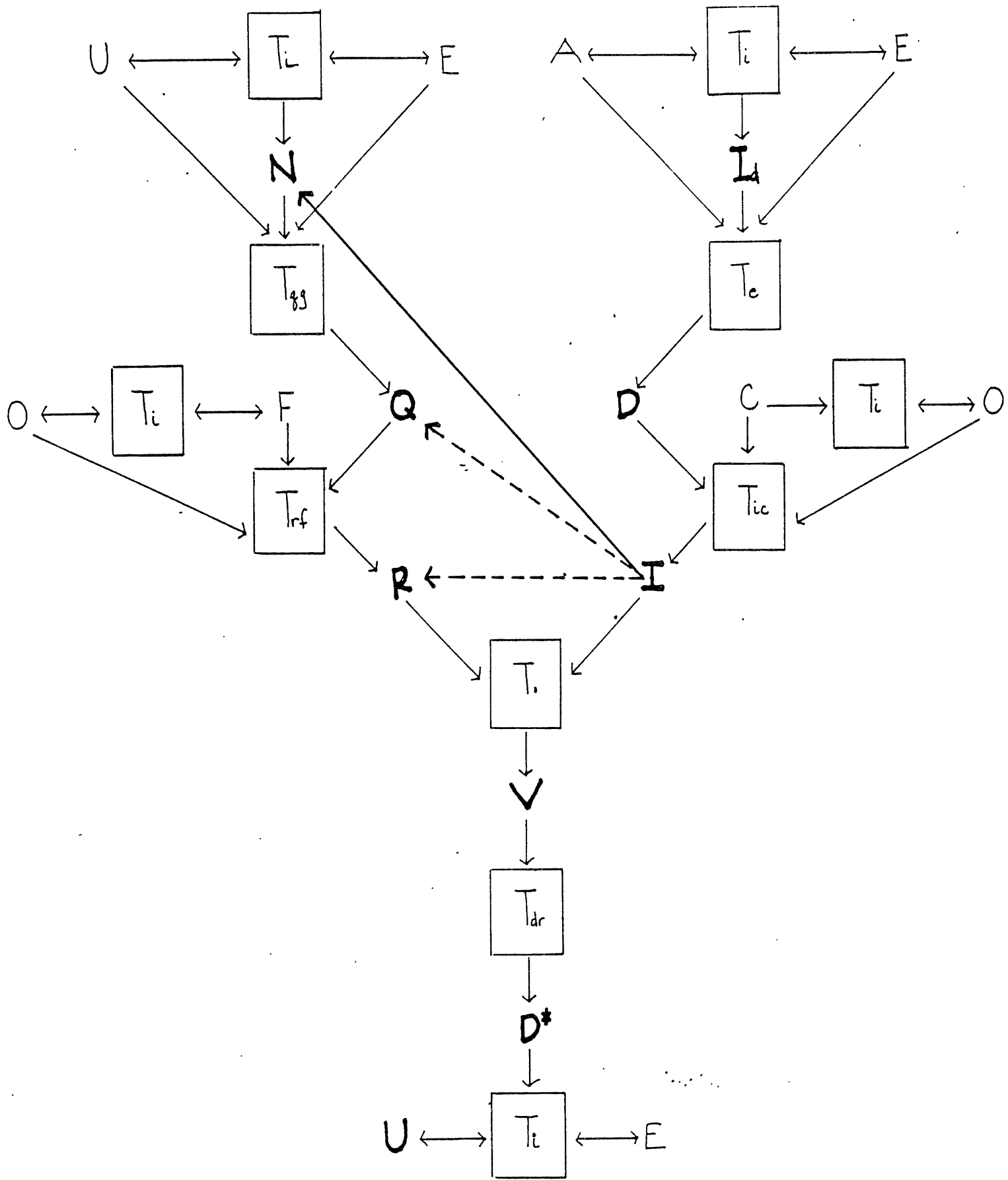
natural language removes the rigor of the logical conse-
quence concept, because the notion of logical consequence is
not defined in a natural language context. The relaxation of
restriction three allows the inclusion of non-inferential
systems, which Cooper refers to as "referential" systems.
He characterizes the differences between inferential and
referential systems as follows.

> In the inferential system, the only information
> used in deducing an answer is information stored
> within the system's memory, and all of the logical
> deductions are also carried out by the system it-
> self. In a reference retrieval system, the user's
> own information in his personal memory is also
> brought into play, and the user contributes his
> own powers of reasoning to the deductive process
> as well. Thus the stored information available for
> satisfying an information need with the help of a
> reference retrieval system must be viewed as the
> user's own knowledge plus the system's stored
> data, and the deductive apparatus is the user's
> own deductive power plus whatever deductive power
> has been programmed into the system. [6, p. 29]
> (emphasis in the original)

Several points here are worthy of note. First, this is
the most general retrieval situation, the one most frequent-
ly encountered in the course of human events. The general
definition is the one which, if practical, will enable us to
evaluate the effectiveness of actual retrieval sys-
tems. Second, we note that this definition describes a
relationship (but no longer the specific relationship of
logical consequence) between the union of information pos-
sessed by the system and the information possessed by the
user and the information need. This relationship is shown,
using the RSM, in Figure X.

Figure X.

General Logical Relevance in the RSM
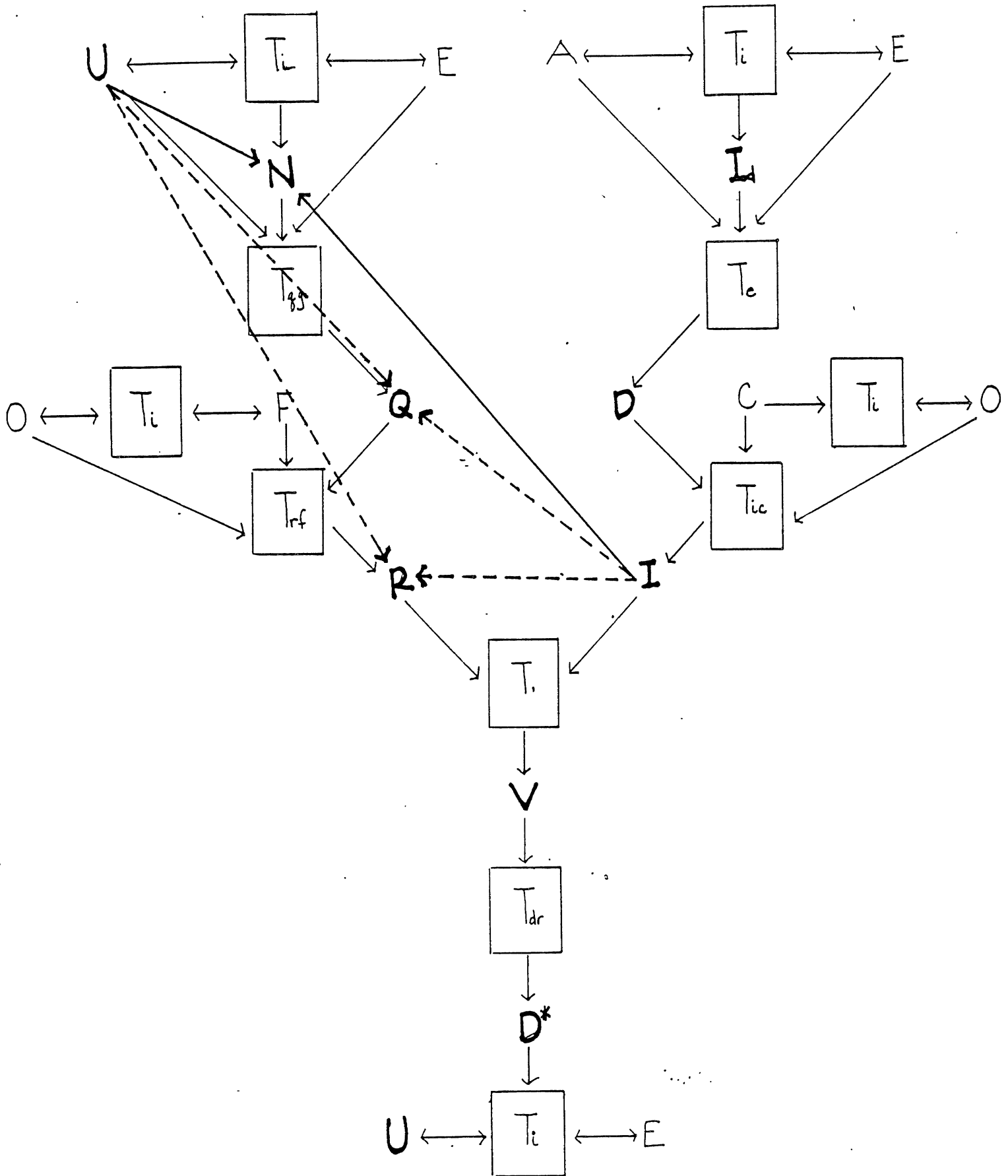
*Wilson's Situational Relevance*

In developing his concept of situational relevance, Wilson [4] begins with Cooper's definition of logical relevance. Noting that Cooper has based logical relevance on the theory of deductive logic, Wilson proposes a complementary type of relevance based on inductive logic. He suggests that information may have relevance without being logically relevant if it strengthens or weakens the user's belief in a given conclusion. This information is evidence used in the inductive reasoning process. Wilson calls this type of relevance evidential relevance, and explains it as follows:

> ... an item of information $I_j$ is [evidentially] relevant to a conclusion $h$ on premisses $e$ if the degree of confirmation, or probability, of $h$ on evidence $e$ and $I_j$ is greater or less than the degree of probability of $h$ on $e$ alone. [4, p. 460]

Information is evidentially relevant if it changes one's belief or state of mind, if it strengthens or weakens the case for a particular conclusion. Wilson considers this point of great importance because the use of inductive logic is far more common than that of deductive logic. In the real world, an important question will often generate a number of convincing arguments rather than a single conclusive argument. Having introduced the concept of evidential relevance, Wilson goes on to describe situational relevance.

First of all, situational relevance relates to a particular individual's situation as he himself sees it. Further, it is limited to those aspects of his situation which

<u>concern</u> him. Concern is described as follows.

> An aspect or feature of a situation will be said to be of concern to a person if the feature can exhibit any one of several different specific states or conditions, and if the individual cares which state or condition is the current one. [4, p. 461]

Wilson contrasts concern with interest. A person may be interested in a subject without having a preference function on the set of its possible states. One might be interested in quasars, but one is concerned about one's health. Situational relevance ignores those aspects of the situation merely of interest, and focuses on the features of concern. Wilson explains situational relevance as follows:

> Let us use the symbol $I^*$ to stand for a person's whole stock of information ... at a given time. Then an item from that stock, $I_j$, is situationally relevant if it, with the other members of the whole stock $I^*$, is logically or evidentially relevant to some question of concern. [4, p. 462]

> We can say briefly: items of information are situationally relevant if they answer, or help to answer, some question of concern. [4, p. 463]

Wilson proceeds to elaborate on the characteristics of situational relevance. It is a subjective concept; it can only be determined by the user. It is a dynamic concept, changing in response to changes in the user's information state, in his preference order for states of the world, and in the membership of his set of questions of concern. The credibility of information items is a determinant of their situation relevance--"... as long as I do not think them true, they are not situationally relevant." [4, p. 463]

Wilson proceeds to address the question of sig-

nificance. An item of information need not be significant because it is situationally relevant. A report that my house is still standing is not, under most circumstances, a significant piece of information, even though the state of my house is a question of concern. The house was standing when I left it, and I know (from my general stock of knowledge) that a standing house will generally remain standing, at least in the absence of earthquakes, explosions, wrecking crews, and a limited number of similar phenomena. After an earthquake, however, a report that my house is still standing may be very significant, because knowledge that an earthquake has occurred has thrown serious doubt upon the applicability of my general knowledge of the fact that houses tend to remain standing. Information is significant, then, when it reports a change in the status of a factor of concern, or when it reports that an expected or possible change has failed to occur.

But significance, Wilson suggests, should go one step further. Not only must the item of information report a state other than that previously expected but it must report a condition either "... higher or lower in preference than the condition previously thought to exist ... or ... correlated with an expectation of change for better or worse on the part of the recipient" [4, p. 467]. At this point we have arrived at Wilson's objective, an expression of "what we would like to be able to expect from information systems" [4, p. 470]. The ideal information system will provide us

with significant, situationally relevant information. In order to do so, it must be enormously powerful. Wilson sketches such as system as follows:

> To be successful, such a system would have to do the equivalent of deciding, for each piece of information in its supply, (a) whether it was directly situationally relevant for the particular person concerned; (b) whether, if not directly relevant, it was indirectly relevant on the basis of other elements of the person's view of the world, and (c) whether, if directly or indirectly relevant, it was significant. To do this it would require the equivalent of a complete representation of the person's view of the world and of his concerns, as well as deductive and inductive logical capacities. If it was to work in "tutorial" mode, it would have to decide, for each item it proposed to deliver, whether it would be accepted if offered, and, if not, whether there was some other sequence of items that would be accepted if offered with the result that the original item would consequently be accepted. This would require the equivalent of a complete theory of learning and understanding for the person concerned. [4, p. 470]

Situational relevance, then, is a relationship between an individual and a stored item of information. It causes the user to change his state of belief in what the current state of a feature of concern is, as can do so via either an inductive or a deductive logical process. In the Retrieval Situation Model, situational relevance is a relationship between a user's internal state of mind and the system 's internally stored information, as shown in Figure XI.

Wilson imposes a number of screens or filters on the flow of relevant information to the user. Credibility is one of these screens; it decreases the stream of situationally relevant information. The concept of significance is another filter, with two screens. Significance filters out the

potentially infinite number of reports which indicate that the status quo remains the status quo in its first screen, and removes those items reporting trivial differences in factors of concern in its second screen. The distinction between features of concern and features of interest is another screen; the flow of information is decreased by removing the features of interest. In this fashion, Wilson's ideal information system protects its user from information overload. This filtration process is shown in Figure XII.

Figure XII

Information Screening in Wilson's Ideal System
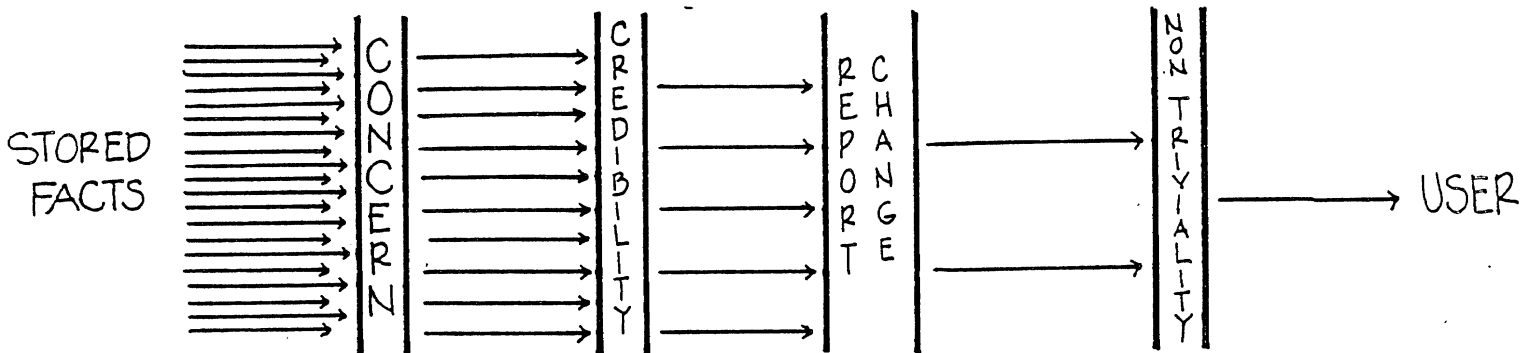


*Bookstein's Relevance*

Based on the assumption that the definition of relevance to be used in document retrieval system evaluation should be one consistent with the purpose of the system, Bookstein [7] proposes a definition he considers "operational." The system's purpose, in his view, is

> ... to satisfy its patrons, and not to match subjects to requests, no matter how well it succeeds at this. Matching topics might be a useful means to this end, but we should distinguish a means from its end [7, p. 270]

Figure XI.

Situational Relevance in the RSM
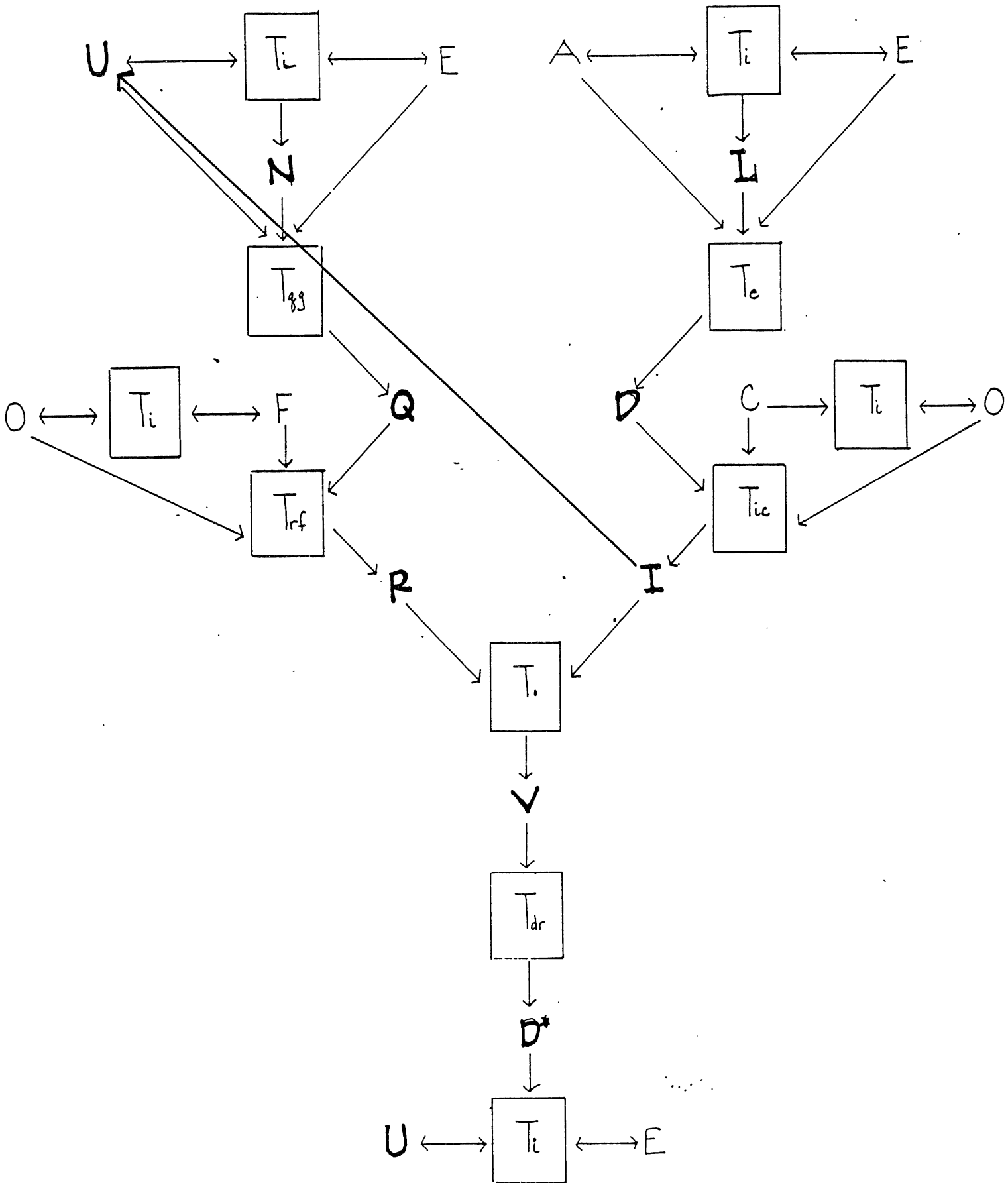
Bookstein is aware that the system cannot perfectly predict the user's evaluation of a document. What it can do is compare representations of the user's information need and of the document's content, and attempt to assess the degree to which they match. This assessment is expressed as a retrieval status value; it is a judgement made by the document retrieval system and guarantees little regarding the satisfaction of the patron. Bookstein contrasts retrieval status value with relevance, which he defines as follows:

> We suggest that relevance be defined as a relation between an individual, at the time he senses a need for information, and a document. We shall say that the document is relevant to the person if he feels the need that brought him to examine the document are [sic] satisfied, at least in part; we shall concede the patron as the final arbiter regarding the relevance of the documents given him ... [7, p. 269]

Relevance, then, is an unspecified relationship between a person and a document. It is distinguished from topicality, a relationship between a representation of the user's information need and a representation of the document's content. A document retrieval system assesses the topicality of a document's representation to a request, and this assessment, the retrieval status value, is used as a predictor of the document's relevance to the user's information need. Relevance can only be manifest as a subjective judgement made by the user, reporting whether or not a document satisfies his information need. Given a set of users cooperative enough to divide a set of retrieved documents

into those relevant and those not relevant, the definition approaches operationality; we can almost begin to evaluate the degree to which different systems are effective in using topicality as a predictor of relevance.

The universe, unfortunately, does not stand still while the retrieval process takes place. But if Bookstein's definition is to be taken literally, this must transpire, for it defines relevance as a relation between a document and the user "at the time he senses a need for information". Since the user generally senses the need for information before going to the trouble of retrieving it, this phrase is rather problematic. Operationally speaking, it will be far easier to change the definition than to change the laws of space and time, so we propose that the offending phrase be replaced by "at the time he receives a document in response to a previously expressed information need." This modification has the additional advantage that it can include the effect of presentation order of retrieved documents. For the sake of comparison, Bookstein's definition is shown, both as proposed and as amended, in Figure XIII.

A second objection to either version of Bookstein's relevance is that it cannot be used to determine recall, since it cannot be used to determine how many relevant documents exist in the collection which were not retrieved. Similarly, the presentation order of the set of retrieved documents will influence the calculation of precision values. This, however, seems less serious, as its effect may

# Figure XIII.

## Bookstein's Relevance in the RSM
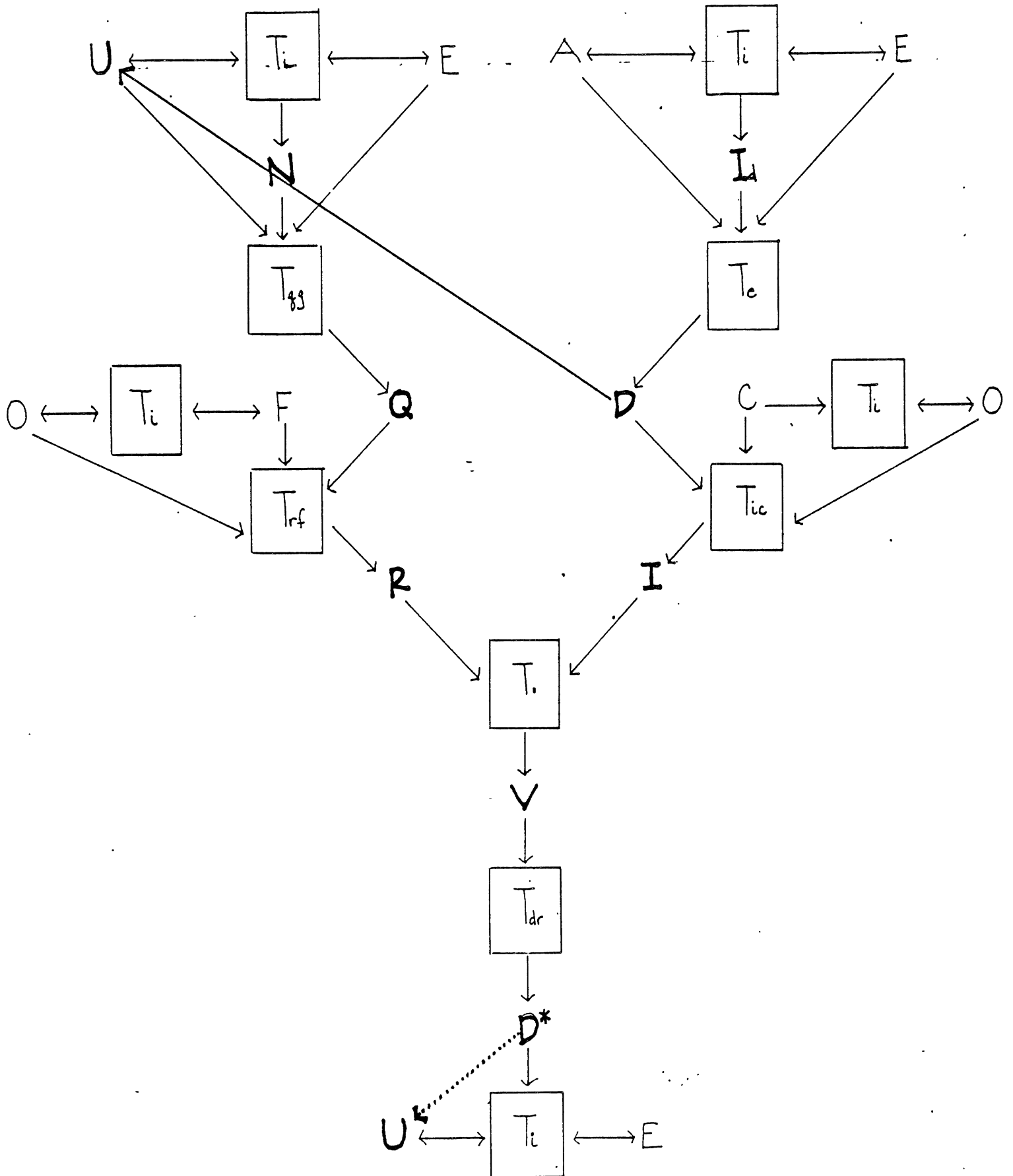
be lessened by the use of large samples and expected values.

## Cooper's Utility

In his exposition of logical relevance, Cooper [6] juxtaposes to it the concept of utility. This latter, he feels, is the ultimate measure of a document retrieval system's effectiveness. Logical relevance is related to utility; the logically relevant document will, in general, have a greater utility than a randomly selected one. But logical relevance is neither necessary nor sufficient for utility—there are at least three other factors involved.

The first of these is the comprehensibility of the document to the specific user. If the logical relevance of the document cannot be discerned by the user, the document has little or no utility to him. If I cannot read Russian, a Russian document has has no utility to me, regardless of the logical relevance of its content to my information need. The user must be able to understand the document in order to for it to be useful.

The second factor determining utility is the importance of the information it presents. Important information is generally more useful than unimportant information. Unfortunately, Cooper does not present any examples of this distinction, but its meaning is intuitively clear.

The final factor Cooper mentions as a determinant of utility is credibility.

A sentence or document may be logically relevant to an information need but useless nonetheless, simply

because the user of the system has no faith in its
accuracy. [6, p. 36]

In other words, a logically relevant document may have
no utility because it is not comprehensible, not important,
not credible, or because of a combination of these fac-
tors. Thus utility, like situational relevance, is subject
to an information filtration process. The utility
relationship, between a user and a document, is shown in the
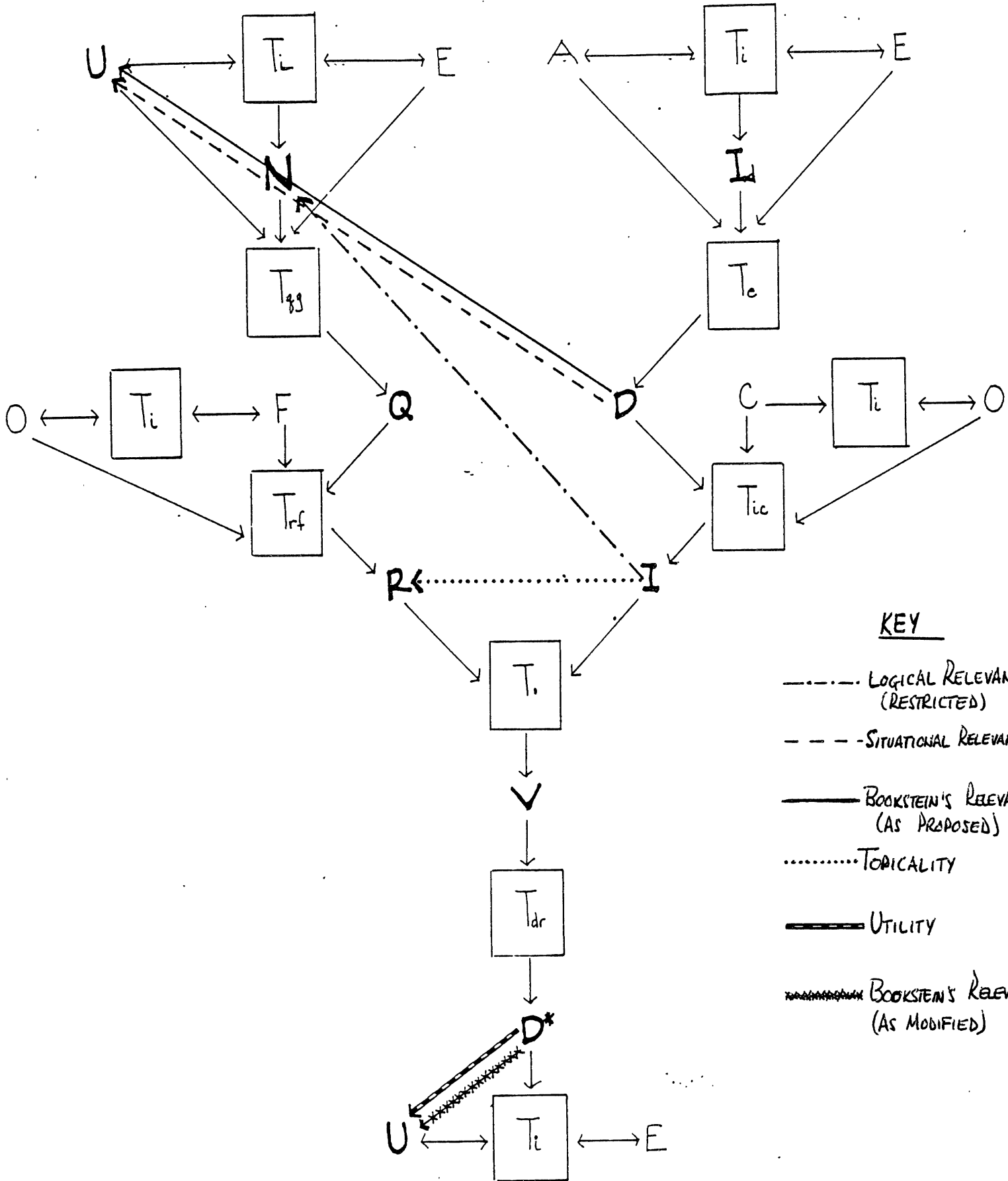RSM in Figure XIV.

## Discussion and Summary

Having developed the Retrieval Situation Model and re-
lated the concepts of relevance and utility to it, we now
wish to examine the results of our efforts and determine
their importance. In the discussion section, we will attempt
to evaluate the usefulness of the RSM, to compare the
several definitions of relevance with utility and with each
other, to draw some general conclusions about the evaluation
of retrieval effectiveness, and to discuss the implications
of our study for the traditional measures of effectiveness,
recall and precision. In the summary, we review the major
developments and conclusions of the paper.

### *Discussion*

First of all, we wish to determine the utility of the
Retrieval Situation Model. We believe that clear evidence
for the utility of the model has been presented. The RSM
clearly describes the six subsystems of the document

# Figure XIV.

## Utility in the RSM

Figure XV.

Relevance and Utility
Compared Using the RSM



KEY

— . — . — . Logical Relevance
(Restricted)

— — — — Situational Relevance

———— Bookstein's Relevance
(As Proposed)

............ Topicality

══════ Utility

xxxxxxxxxx Bookstein's Relevance
(As Modified)

retrieval process, showing their components, inter-relationships, and the external factors affecting them. Moreover, it can depict the trial and error nature of the document retrieval process. Finally, it has successfully accommodated the several definitions proposed, facilitating their comparison within a consistent framework.

What generalizations concerning the evaluation of document retrieval system does the RSM suggest? Its major contribution here is to point out that the evaluation of the effectiveness of a document retrieval system (i.e., the Bookstein-Cooper model) is likely to lead to suboptimization, because the document retrieval system is only one of six subsystems involved. The effectiveness of the overall process is what we really want to measure (and improve). This will require the inclusion of such factors outside the document retrieval system as users, classifiers, and formulators. In an unstable organizational context, the context itself must be considered. In exceptional circumstances, it will be necessary to factor in effects caused by variations in the environment. Use of the RSM implies a more complex process of retrieval system evaluation.

What insight has use of the RSM given us about the three definitions of relevance? First, it has enabled us to make statements concerning the operationality of each definition. Beyond this, it can help us select a definition for use as a basis of retrieval system evaluation.

We have seen that logical relevance is operational only

under the assumption that the system-manipulable request accurately and completely describes the user's information need. Because this assumption is patently unrealistic, we conclude that logical relevance is not an operational concept.

It is not entirely clear whether or not Wilson intends situational relevance to be an operational concept. The RSM, in any case, has shown that it is not.

We have seen that Bookstein's relevance is not an operational definition, as proposed, although it is easily modified to become so. Similarly, utility is an operational concept. Unfortunately, neither of these is susceptible to any simple and elegant measurement procedures, but rather both require cumbersome, subjective, and inexact techniques at present.

Topicality, the relationship between the request and the index record, has been shown to be an operational concept. It is not, however, a meaningful basis for the evaluation of retrieval effectiveness. By definition, that which the system retrieves is topical; this will not help us discriminate between systems. Topicality may, however, be a valuable basis for evaluation of the index record creation system $(S_{ic})$ and the request formulation system $(S_{rf})$.

Does the RSM help us select one of these concepts as an operational basis for the evaluation of retrieval effectiveness? As is noted above, only the modified version of Bookstein's relevance and utility are operational concepts

with meaningful discrimination between different sys-
tems. The RSM thus narrows the field from five candidates to
two. Moreover, if we examine the RSM representations of
these two candidates (cf. Figure XV.), we see that
Bookstein's relevance (as modified) is essentially a simpler
version of utility, simpler in that it lacks the filters
Cooper discusses as determinants of utility. This suggests
that, of the concepts discussed in this paper, utility is
the most appropriate basis for evaluation of retrieval ef-
fectiveness.

Finally, we wish to consider what all of this means in
terms of the use of recall and precision as measures of
retrieval effectiveness.

From an operational point of view, we must rule out
the use of recall, because it cannot be assessed using
either utility or Bookstein's relevance (as modified). The
utility of a retrieved document can be approximated (see
Cooper [8]), but no utility approximation can be made for a
document not retrieved. According to the definition of
recall, such a judgement must be made for each document in
the collection. Recall cannot be used as a criterion for
evaluation of retrieval effectiveness in a utility-based
evaluation scheme.

On the other hand, estimation of a precision value
based on utility is not out of the question. Two considera-
tions, however, make the value of such a estimate rather un-
certain.

The more effective the retrieval function is, the less likely it is that the user will examine all the documents that are retrieved. Suppose that ten documents are retrieved, and that after examining three, the user declares his information need satisfied. He stops examining the set of retrieved documents, stating that the first and third documents were useful. Should precision be valued at .2 (2/10) or at .667 (2/3)?

The second objection follows from Cooper's suggestion [6] that utility is a continuous-valued variable. It seems highly unlikely that users will be able to assign fractional utility values to documents in any consistent fashion. The alternative is to credit each document of any perceived utility as being useful, but doing so seems equally un-desirable. This, for example, would credit a system more for retrieving two marginally useful documents than for retrieving a single document which fully satisfied the user's information need!

Our RSM-based analysis suggests that utility should be selected as the basis for evaluation of retrieval effective-ness. This, in turn, rules out the use of recall, and strongly discourages the use of precision.

## Summary

This paper began by reviewing Bookstein and Cooper's model of a document retrieval system. This model was ex-panded into the Retrieval Situation Model, which represents

the entire retrieval process. The Bookstein-Cooper model is one of six subsystems of the RSM. The others deal with the processes by which documents are created and indexed, queries are generated and transformed into system-manipulable requests, and the set of retrieved documents is used. The components of each of these subsystems are outlined. Their interrelationships and the influence of factors outside the RSM are shown. The ability of the RSM to depict the cyclical, trial and error nature of the retrieval process is demonstrated.

After building the Retrieval Situation Model, we use it to compare several definitions of relevance. Our major conclusions are presented below.

(1) The RSM, though in some senses a primitive and inexact model, is a useful vehicle for the examination of the retrieval process and for the comparison of alternate versions of such concepts as utility and relevance.

(2) Neither Cooper's logical relevance nor Wilson's situational relevance is a practical basis for the evaluation of the document retrieval function.

(3) Bookstein's relevance, though not operational as defined, can be modified to become an operational concept which can be used for retrieval evaluation. This modification, however, removes the significant differences between it and utility.

(4) Utility is a feasible, though arduous, basis for the evaluation of retrieval function evaluation.

(5) The traditionally accepted measures of retrieval effectiveness, recall and precision, have been shown to be inappropriate if utility is used as the basis for evaluation.

As always, much more remains to be done than has yet been done. For example, the elaboration of the RSM's com-

ponent systems and of their principles of interaction should be undertaken. These and similar efforts can be expected to give us important insights into the nature of the document retrieval process.

## Appendix A: List of Symbols

| Symbol | Meaning | Found in Systems ... |
|--------|---------|----------------------|
| $A$ | Author | $S_{dc}$, $S_{iu}$ |
| $C$ | Classifier | $S_{rf}$ |
| $D$ | Document | $S_{dc}$, $S_{ic}$ |
| $D^*$ | Documents retrieved | $S_{iu}$ |
| $E$ | Environment | $S_{dc}$, $S_{qg}$, $S_{iu}$ |
| $F$ | Formulator | $S_{rf}$ |
| $I$ | Index record | $S_{ic}$, $S_{bc}$ |
| $Id$ | Idea | $S_{dc}$ |
| $N$ | Need of information | $S_{qg}$ |
| $O$ | Organizational context | $S_{rf}$, $S_{ic}$ |
| $Q$ | Query (natural language) | $S_{qg}$, $S_{rf}$ |
| $R$ | Request (system-manipulable) | $S_{rf}$, $S_{bc}$ |
| $T$ | Retrieval function | $S_{bc}$ |
| $T_e$ | Expression process | $S_{dc}$ |
| $T_i$ | Interaction process | $S_{qg}$, $S_{dc}$, $S_{rf}$, $S_{ic}$, $S_{iu}$ |
| $T_{ic}$ | Index creation process | $S_{ic}$ |
| $T_{iu}$ | Information use process | $S_{iu}$ |
| $T_{qg}$ | Query generation process | $S_{qg}$ |
| $U$ | User | $S_{qg}$, $S_{iu}$ |
| $V$ | Value of retrieval status | $S_{bc}$, $S_{iu}$ |

# References

1. Bookstein, A., and W.S. Cooper. "A General Mathematical Model for Information Retrieval Systems". *Library Quarterly.* 46(2):153-167; 1976.

2. van Rijsbergen, C.J. *Information Retrieval* (2nd edition). London, UK: Butterworth & Co; 1979.

3. Cleverdon, C.W., J. Mills, and M. Keen. *Factors Determining the Performance of Indexing Systems,* Volume I - *Design,* Volume II - *Test Results.* Cranfield, UK: ASLIB Cranfield Project; 1966.

4. Wilson, P. "Situational Relevance. *Information Storage and Retrieval.* 9:457-471; 1973.

5. Saracevic, T. "Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science". *Journal of the American Society for Information Science.* 26:321-343; 1975.

6. Cooper, W.S. "A Definition of Relevance for Information Retrieval". *Information Storage and Retrieval.* 7:19-37; 1971.

7. Bookstein, A. "Relevance". *Journal of the American Society for Information Science.* 30(5):269-273; 1979.

7. Cooper, W.S. "On Selecting a Measure of Retrieval Effectiveness, Part I., 'The "Subjective" Philosophy of Evaluation', and Part II., 'Implementation of the Philosophy'". *Journal of the American Society for Information Science.* 24:87-100 and 413-424; 1974.

9. Cooper, W.S. "Perspectives on the Measurement of Retrieval Effectiveness". *Drexel Library Quarterly.* 14(2):25-39; 1979.

10. Swanson, D.R. "Information Retrieval as a Trial-and-Error Process". *Library Quarterly.* 47(2):128-148; 1977.