Models of Human Memory and Computer Information Retrieval: Similar Approaches to Simiar Problems

by

Peter W. Foltz

Department of Psychology and
Institute of Cognitive Science
University of Colorado
Boulder, CO 80309

ICS Tech Report 91-03

Models of Human Memory and Computer Information Retrieval: Similar Approaches to Similar Problems

Peter W. Foltz[1]
Department of Psychology and Institute of Cognitive Science
University of Colorado, Boulder

---

[1] This research was supported by the Army Research Institute ARI project MDA 903-86-CO143. The author thanks Walter Kintsch, Anders Ericsson, Reid Hastie, Gerhard Fischer, Adrienne Lee, Susan Dumais, and Thomas Landauer for useful discussions and comments on earlier versions of this paper.

Abstract

Models of human memory and computer information retrieval have many similarities in the methods they use for representing information and accessing the information. This article examines the methods and representations used in both human memory modeling and computer information retrieval and discusses how they are similar and how they differ. From these similarities and differences, the features that lead to successful retrieval in both human memory and computer information retrieval domains are discussed. An analysis of these features can then help in the future design of both human and computer retrieval models.

Table of contents

Introduction

A recent estimate puts the amount of information accumulated by a normal adult over a lifetime to be on the order of 10^9 bits (Landauer 1986). Modern computers are also able to store about this same amount of information. While this great storage capacity permits humans to store many experiences and facts and computers to store many items of information, one of the keys to the flexibility of both the human brain and the computer is the ability to retrieve the relevant information quickly and accurately. Without fast access to information, both humans and computers could not make the almost instantaneous actions for which they are both known. Similarly, accessing the wrong information would lead humans and computers to either perform the wrong actions, or spend additional time trying to retrieve the correct information. Thus, a primary issue in both computer and human information retrieval is how to access the correct information in an efficient manner.

It may seem odd to discuss human and computer memories as two similar things since they differ in many ways. Physically, the two are far apart. Human memory is neurally based with rich interconnections between neurons, while computer memory is based on silicon chips with far fewer connections. Human memory has also been refined through evolutionary adaptation over many centuries, while computer memory has only existed for less than fifty years. With this long period of adaptation, the human brain has gone through many changes that have optimized it to store and retrieve information in an efficient manner. Thus, in terms of the physical design of the brain and computer memories, there are few similarities and the human memory system appears to be far superior through its complexity and efficiency.

The approaches to the study of retrieval from human and computer memory are also quite distinct. The way humans retrieve information from memory has been studied by psychologists over the past 100 years. Some of the issues studied include the representation of the information in memory, what cues can be used to retrieve information, and the length of time information can be stored in memory. One of the primary approaches in human memory research is to predict phenomena of memory retrieval through the development of psychological models of memory and to test these models via laboratory experiments. These models have been used to predict such phenomena as paired associate recall, differences in recognition and recall, the role of elaboration in recall, recall as a function of the encoding of the information, and the retention of information. While several models may account for the same phenomena, they can often differ in the way that they represent information and in their methods for accessing the information. A variety of representations and methods have been used including associative and distributed representations of information, and direct and indirect connections between cues and items in memory. The differences between models provide a means of comparison between different possible ways that we access our knowledge.

The issues in retrieval from computer memory are quite different. Rather than concerns of trying to explain how the retrieval works, the primary issues are about designing the most efficient manner to store and retrieve the information. To investigate these issues, information retrieval researchers test different methods of representing the information and different algorithms for retrieving the information from memory. These methods include such things as placing the information in hierarchical or associative networks, and using probabilistic models or statistical techniques to predict what keywords will best retrieve the desired information. Through testing these methods, researchers can determine which methods will optimize access to stored information.

Thus, an issue central to both psychologists modeling memory and information retrieval researchers designing information retrieval systems is determining how to represent the information and what methods to use to retrieve the information. While there is the added constraint for the psychologist that a model must also fit human memory phenomena, both psychologists and information retrieval researchers are concerned with designing a system that will find an optimal solution to the problem of accessing the information. To investigate this problem, both fields have used similar representations of the information and similar methods for retrieving the information. Models employing differential strength of associations between cues and information in memory and spreading activation to retrieve associated items have been used in both fields for retrieval tasks.

Given that there are similarities in the methods that are tested in these two fields, it makes sense to take a closer look at the issue of retrieval in general in order to determine what comprises successful retrieval. This issue is of interest from the point of view of both computer information and human memory retrieval. From the memory retrieval side, we can look at successful information retrieval systems and examine how human memory retrieval manages to solve analogous tasks. As there are improvements in information retrieval methods, these improvements may suggest new ways of modeling the representation and retrieval of information from memory. Conversely, from the computer information retrieval side, we can look at information retrieval as an extension of human memory. Since human memory is not optimized to accumulate and store the large amounts of information available to computers, there is a need to develop methods for humans to retrieve relevant information stored in the computer. Because human memory retrieval appears to be a very efficient system, we may then find more effective methods of retrieving information from computers through a greater understanding of how the human memory retrieval system works.

This article will examine methods and representations used in both human memory modeling and computer information retrieval and discuss how they are similar and how they differ. From these similarities and differences, the features that lead to successful retrieval in both human memory and computer information retrieval domains can be determined.

The Stages of Retrieval

The retrieval of information from both human memory and computers can be seen as occurring in three distinct stages. These stages are: generating and assembling the retrieval cues, using the cues to retrieve information, and verifying that the retrieved information is what is desired. These stages differ, though, in the type of cognitive activity used. Initially, to retrieve information, cues need to be generated. The process of generating these cues is strategic, involving controlled processing (e.g. Schneider & Shiffrin, 1977). A person develops cues that describe the information desired based on the current context. Once the cues have been generated, they are then used for the retrieval. In the case of information retrieval, the cues need to be transmitted to the computer, while in human retrieval, the generated cues are used automatically. Nevertheless, in both cases, the actual retrieval process is automatic, i.e. not under a person's strategic control. Once information has been returned from the retrieval process, it is again under the control of a strategic process which evaluates the information to determine if it is what was desired.

The retrieval of information may not be just a single cycle of these three stages, but may involve several iterations. In this case, some of the information retrieved in a previous iteration may be added as additional cues for the next retrieval attempt. While all three stages are important to retrieval, they differ greatly in the type of processes used. In many human and computer retrieval models, the automatic and the strategic components are often treated separately. For this reason, the models discussed are organized based on whether an automatic or strategic component of retrieval is being examined.

Human Memory Retrieval

Psychological Approach to Human Memory Retrieval

Human memory models are developed with the primary goal of trying to explain how our memory works. In general, the approach has tried to answer questions about how we represent information, how we store and retrieve the information, and why we may fail at retrieval. This approach is typically characterized by measurement or observation of human memory phenomena and the development of models that can explain these phenomena. In the human information processing system, there are a variety of places where information is stored and can be retrieved (e.g. sensory register, short term memory, long term memory). This article will focus on models of retrieval from long term memory since they address many of the same storage and retrieval issues as those in computer information retrieval.

Even though human memory models have succeeded in predicting many memory phenomena, the study of memory is really a study of a black box. There is no direct access to memory that would enable us to examine first hand how information is stored and processed. Therefore, we must make hypotheses about the structures and processes from testing the inputs and outputs of memory. Although the field of neuroscience has made large advances in the past 20 years, most memory retrieval research is still based on manipulating information put into memory and examining what type of information comes out. This approach does have some limits though. It is not easily possible to control all information that enters memory, how it is encoded, or when it is retrieved. For these reasons, this approach will likely never provide a full description of how memory works. It has, nevertheless, provided a wealth of information on human memory and permitted researchers to develop models of how memory may work.

Although the black box approach to memory is a limiting factor to discovering how information is processed and stored, there are still some constraints based on general observations of human memory for the development of memory models. One underlying constraint is that human memory is an efficient system. In almost all cases, we are able to retrieve the necessary information to function in the world. When driving a car, all the relevant information for operating a car, observing the rules of the road and navigating are all retrieved normally and used effectively as they are needed. We perform thousands of such task each day and seldom fail to complete them due to a failure in retrieving the relevant information.

Anderson (1990; Anderson & Milson, 1989) has used such constraints to develop a rational model of human memory. Anderson's principle of rationality is that "the cognitive system operates at all times to optimize the adaptation of the behavior of the organism". In this way, human memory has been optimized through evolution to retrieve relevant information in an efficient manner. Anderson applies this principle of rationality to develop a variety of models of cognitive processes including memory retrieval.

To apply the principle of rationality one must specify the goals of the system, the structure of the environment and some minimal computational constraints. The goals of the memory retrieval system is to retrieve relevant information while balancing the cost of searching for information. The power of the model comes from the specification of the structure of the environment. Anderson & Schooler (in press) showed that the history of usage of an item is exponentially related to the probability that it will need to be recalled. From the specification of these simple environmental constraints, Anderson then creates a probabilistic retrieval model that explains many human retrieval phenomena. While there has been some question whether the success of the model is based on additional auxiliary assumptions (Simon, in press), the model does illustrate that human memory can be modeled using very simple guidelines.

A further constraint for developing memory models is based on the richness of the information in our environment and the flexible ways we use information. Our environment is rich with information and we are able to store and retrieve a great deal of this information. This information may be stored with the relevant dimensions that serve as the context for when the information was encountered. Thus, our memory system must be sufficiently complex to encode this richness and complexity of information encountered in our environment. Upon recall, we are able to make associations between pieces of information in memory. For example, we may associate the word *dog* with both *cat* and *hot*. This suggests that there are strong interconnections between information in memory. Thus, from these observations, we can conclude that successful memory models must be efficient in storing and retrieving a wide range of interconnected information.

Memory Models

Psychological memory research is a fairly new development with almost all laboratory studies occurring over the last 100 years. The first systematic study of memory was performed by Ebbinghaus in 1885. Running the experiments primarily on himself, he used standardized conditions and procedures to study the retention of associations of nonsense syllables he had learned. Looking at the number of repetitions through a list of syllables, and the amount of time it took him to memorize a list, he was able to assess the retention of information in memory and the amount of relearning necessary after having previously studied a list. Ebbinghaus's method of presenting a subject with new information and studying how it is recalled based on various manipulations is still used widely in memory research today. A good overview of verbal learning may be found in Estes (1976).

Bartlett (1932) performed a series of experiments that differed from the methods of Ebbinghaus. In his best known experiment, subjects read an Eskimo tale titled, *The War of the Ghosts*. He then had subjects recall the text at various intervals of time. His basic findings were that over time, the recall of the story became both shorter and more normalized towards making the content more compatible with the subject's knowledge and cultural experiences. These results suggest that memory is a schematic process with new information being integrated into already established representations or schematta.

The development of memory research has progressed over the past century, going through some changes in emphasis. Up to the 1950's, the primary memory research was in the verbal learning tradition. With Miller's (1956) paper, memory research changed to emphasize structural analyses of memory. Based on data from memory span experiments, he developed the concept of the human as an information processing device with a limited capacity. This limited capacity nevertheless could be overcome to some extent through the use of strategies. This concept was then extended with many investigations of short term memory (e.g. Brown, 1958; Peterson & Peterson, 1959), sensory memory (e.g. Sperling, 1960), and long term memory (e.g. Collins & Quillian, 1969; Tulving, 1972, 1983). Overall, these studies provide models of how information is encoded, stored and retrieved from memory. The results of much of this research has then provided the bases for the development of models of memory retrieval.

While there have been many models of memory, this article will focus on three different classes of models of retrieval from long term memory. The three classes of models are characterized by compound cue, spreading activation and distributed information representations and retrieval. The models described in all three of these classes primarily focus on the automatic retrieval component of the retrieval process and models of the strategic retrieval component will be discussed later. While Tulving (1983) has espoused the distinction between retrieval of episodic and semantic memories, this article assumes the laws of retrieval are the same for both episodic and semantic memory (e.g. McKoon, Ratcliff & Dell, 1986). Similarly, no distinction is made between retrieval due to recall and recognition. While there is evidence for differences between the two types of retrieval (Kintsch, 1968; Tulving & Thompson, 1973), the research described in this article concentrates on models of recall. This permits a comparison of the models to the computer based retrieval system in which all retrieval can be characterized as being recall rather than recognition.

<u>Compound Cue Models</u>

A compound cue model of memory assumes associative connections between retrieval cues and items in memory. Examples of this model are the Search of Associative Memory (SAM) model used in Gillund & Shiffrin (1984) and Raaijmakers & Shiffrin (1981) and the compound cue priming model of Ratcliff and McKoon (1988). The model serves as a general theory of retrieval from long-term memory and has been used to predict such things as the recall of lists of memorized words with and without cues of subsets of the words and the range and decay of priming. Using cue-dependent probabilistic search of an associative memory, it incorporates both elements of probabilistic search theory (e.g., Shiffrin 1970) and associative network theory (e.g., Anderson 1972)[2].

In the SAM model, each item in memory has a probability of being sampled that is dependent on the associative strength of that element to a set of probe cues in comparison to the strengths of all other items to the same set of probe cues. While Raaijmakers and Shiffrin state that they treat memory (Long Term Store) as an interconnected network in which all elements of memory are connected to all others, in this model there are nevertheless no connections between the items in

---

2 While the SAM model is a description of an associative memory, it should not be confused with the way associative memories are described in Hinton & Anderson (1981) and Rumelhart & Norman (1988). Their descriptions of associative memory use a distributed representation of the information in memory and would fall under the distributed memory models heading in this article.

memory. Connections between memory items are only mediated through connections from two memory items to the same cue. This representation is therefore similar to a 2 layer connectionist model without connections between items on the same layer.

The retrieval of information consists of a sampling component and a recovery component. Information is first sampled by a probabilistic sampling rule as shown below.

$$P_s(I_i \mid Q_1, Q_2, \ldots Q_M) = \frac{\prod_{j=1}^{M} S_t\,(Q_j, I_i)^{w_j}}{\sum_{k=1}^{N} \prod_{j=1}^{M} S_t\,(Q_j, I_i)^{w_j}}$$

In this formula, the probability of sampling an image *I* given a set of *M* cues *Q* is based on the product of the strength of all cues to that image normalized by the sum of all the *N* item-cue strengths. The w_js serve as weights for each cue representing the relative saliency of the cue. The power of this retrieval method comes from the product of the cue strengths, S_t. An item in memory with several strong cues will have a much higher probability of being retrieved than those with fewer cues or with weaker cues. Once the cues and context have been used to sample memory, a single item is selected and then if sufficient information is available about the item, it is recovered, completing the retrieval.

The SAM model treats retrieval of information as an iterative process. In performing retrieval of information for a free recall task, the model initially uses only general context as the initial cue to retrieve an initial item. It then uses that item as a new cue to retrieve an additional item. This process of retrieving items and then using them as cues continues until a certain number of failures to retrieve any new items occurs. This iterative process takes place automatically and is thus not under conscious control. Nevertheless, Raaijmakers & Shiffrin place this retrieval process in the context of a full controlled retrieval system in which a retrieval plan is initially made, then cues are assembled in short term store (STS), then the automatic retrieval process takes place, and finally the information is evaluated under strategic control.

The learning of information takes the form of creating new associations and strengthening existing ones between items in STS and those in long term store. Since items in STS tend to be stored with many of the same contextual cues, they tend to be stored and retrieved together. This speaks to the notion of episodic memory (Tulving, 1983), in which events occurring close in time will be retrieved together since a lot of the same context cues (e.g. temporal, physical states, feelings) will be equally associated with all the information. Walker & Kintsch (1985) illustrated this effect, using SAM for modeling retrieval from script-like structures (e.g. going to the doctor). Overall, they found that they could model the automatic aspects of retrieval, but the information retrieved was very contextually based and the strategic aspects of the model were not able to account for some of the associations made by the subjects.

Thus retrieval in the SAM model is entirely dependent on the cues used. Because of this, it is somewhat limited in performing tasks in which there may be some rich associations between items unless those associations are expressed through cue-item connections. Nevertheless, the model highlights the concept that retrieval of relevant information is primarily based on the associations between information stored in memory and their cues.

Spreading Activation Models

The concept of spreading activation in a network of associated information has been widely used as a search mechanism of semantic networks. The concept of associated information can be traced back to Aristotle's associationism (Anderson & Bower, 1973), while Quillian (1969) first described how it could be implemented as a model of a memory search mechanism. Spreading activation models of retrieval assume a semantic network in which items in memory are highly interconnected. The connections between memory items are typically represented as the semantic relationships between the items with different strengths based on the semantic relatedness. As in the SAM model, initial items are activated by the cues given. The activation of these initial items then spreads along the semantic connections to activate other associated items.

Collins & Loftus (1975) and Collins & Quillian (1969) applied spreading activation to retrieval, examining semantic verification of stored facts. They found that the verification time increased as a function of the increased distance between concepts represented as a network. These increased reaction times were explained by the assumption that it took time for the activation to spread the greater distance. Further research has shown other features of how spreading activation works. The amount of spread to any item is a function of its strength of connection to the initial item (Lorch, 1982) and also a function of the relationship of the strength of that connection to the sum of the strengths of connections from the initial item (Reder & Anderson, 1980). Typically the activation will continue to spread from item to item until there are no more significant changes in the overall pattern of activation of items. The retrieval of information is then a function of how highly activated items are after the spread of activation. Like the SAM model, once the cues have been given, the spread of activation works as an automatic process and thus not under strategic control (Neely, 1977). Spreading activation models have been used as a search mechanism to explain a wide range of cognitive phenomena such as perceptual word recognition (McClelland & Rumelhart, 1981), automatic semantic priming (Neely, 1977), sentence recognition (Anderson, 1983), text comprehension (Kintsch, 1988), and planning (Mannes & Kintsch, in press).

Anderson's ACT models (Anderson 1976, 1983) use spreading activation as the retrieval mechanism for declarative information. In these models, memory is represented as two parts, a procedural production system and a declarative associative memory. For doing a sentence recognition task of studied facts such as *Napolean Bonaparte was an emperor* (Lewis & Anderson, 1976), the procedural memory consists of productions for sentence concept matching, while the declarative memory consists of a network of concepts from the studied sentences. When a probe sentence is given, words in it activate certain concepts in memory. The activation of these concepts then spreads to other memory items, with a greater spread of activation between two items that had been studied together. Activated items are then processed in a pattern matcher and tested in the productions at a rate based on how highly activated they are. One of the key notions in the ACT* model is the *fan* of any concept. The fan refers to the number of facts that are associated with any concept. The greater the number of facts associated to a concept, the slower the recognition of any of the facts. When a high-fan sentence probe is given, the fan from the element limits the amount of activation that can be spread to any memory item, since the activation must be spread among all of them. This lower activation then results in a slower matching in the procedural memory. Similarly, since the activation from multiple items converging on a single item is cumulative, the more concepts that are provided for recognition, the faster the recognition. This is again like the SAM model in that multiple strong cues provide a better retrieval performance than individual cues, or multiple weak cues, although the SAM model uses a multiplicative rule instead of a summation of activation. This means that in the SAM model, a cue with no connections to an item will have a connection strength of 0 to the item and will therefore mean that there would be a probability of 0 that the item would be retrieved.

One central issue in spreading activation models concerns the amount of spread of activation. Activation could just spread once from cued items to their associated nodes, or it could then continue to spread from those nodes on to additional nodes. The assumption that activation spreads multiple times has been widely used to explain a variety of retrieval phenomena. The Collins & Quillian (1969) model of hierarchical organization of memory assumed that the time to verify category membership was a function of the number of nodes that had to be traversed. Thus, it would take more time to verify a statement, *A Canary is an animal* than *A canary is a bird*. Anderson (1983) also uses multiple spreads of activation, however he implements a dampening function that reduces the activation of nodes to their original resting level as an exponential function of the distance from the source of the activation. While these models use the multiple step spread, DeGroot (1983) found evidence against multiple step spread in a semantic priming lexical decision study. She found that activation did spread between related words (e.g. *bull-cow*), but not between unrelated words that were both mediated by the same related word (e.g. *bull-cow-milk*). Balota & Lorch (1986) found the same results for the lexical decision task, but found evidence for mediated priming in a pronunciation task. They explained the fact that there was no mediation on the lexical decision task based on a post-access decision process that was only sensitive to relatively strong activations. Thus, while it is not clear how far activation spreads, the results of most studies indicate that the activation decays very quickly as it spreads to more distant nodes.

In terms of retrieval, a cue dependent model such as SAM and spreading activation model differ primarily only on the range of activated nodes. In SAM the only activated nodes would be those directly associated with the cues, consistent with the findings of DeGroot. In a spreading activation model with multiple spreads, retrieval is not only cue-dependent, but also dependent on the strength of connections between items in memory. While spreading activation has used priming as a key to its existence, Ratcliff and McKoon (1988) have shown how a compound cue theory using the Gillund and Shiffrin (1984) model can be used to explain priming phenomena. Their model showed that primes facilitated retrieval only when the prime and the information had been encoded simultaneously in STM. A more complete review of cue-dependence in memory models may be found in Ratcliff & McKoon (1989).

Compound cue and spreading activation nevertheless, are not always treated as separate retrieval models. Kintsch's (1989) Construction-Integration model uses both models for different parts of retrieval. In the construction phase, concepts in memory act as cues to retrieve associated information in knowledge. This process constructs a matrix of associations between related concepts. Then, given some set of information presented to the model, activation is spread in the matrix. This results in the higher activation of certain concepts in memory than others, increasing the probability that they would be incorporated into the final memory representation.

Distributed Models

Distributed models of memory differ greatly from the previous models in terms of their representation of information in memory. In the SAM model and spreading activation models such as ACT*, information is represented as discrete nodes in which a concept is stored. In distributed models of memory, the information is typically represented as a vector of features. Examples of these types of models include the CHARM model (Eich, 1982, 1985), TODAM (Murdock, 1982, 1983) and MINERVA 2 (Hintzman, 1986).

The CHARM and TODAM models are very similar in their methods of encoding and retrieval. The models have been used for explaining such phenomena as prototype extraction, depth-of-processing effects, paired associate recall, and retrieval failure. For the representation of an item, each feature in the item vector will have a certain numerical value with the absolute value of that feature representing its importance. Two items in consciousness may be associated through the process of convolution. In convolution, the sum of the products of certain features of the two item vectors is used to create a new vector. The process of convolution is similar to that of the dot product, although it results in a vector. A description of the formula for convolution may be found in Eich (1982, 1985). This resulting vector is then added to the single composite associative trace. This trace is the sum of the various associations that take place and as more associations take place, they are added to the trace. Thus, the trace serves as representation of all associations learned.

Retrieval of information takes place through the process of correlation. When a retrieval cue is given, its vector is correlated to the composite trace vector. The correlation between the two vectors is done in the same manner as in convolution, resulting in a retrieved vector. Due to the convolution and correlation process, the retrieved item will not be identical to the initially encoded item. Thus, the retrieved vector represents the original vector plus some noise from the other information that was encoded into the associative trace. When an item is retrieved, a pattern matcher is then used to match it to the closest concept. An example of this would be in retrieval from semantic memory to produce a word (Eich, 1982). The pattern recognizer would first match the retrieved vector against all items in the lexicon taking the dot product between the retrieved vector and the vectors representing each lexical item. Then based on the values of the dot products, one of the lexical items is chosen.

The MINERVA 2 model (Hintzman, 1986) shares some features with the previous models. Information is again stored as a vector of features. However, rather than storing all information as a single trace, each individual trace is stored. Hintzman uses MINERVA 2 to model the storage of episodic memory (e.g. Tulving 1983) and accounts for a variety of schema-abstraction phenomena such as differential forgetting of prototypes and category size effects. In this model, each memory trace represents a record of a particular episode or experience. Retrieval is performed through a function similar to the dot product between the vector representing the cues and each of the stored traces resulting in an activation value for each trace. The information retrieved is made up of two components, a content and an intensity. The intensity represents the total amount of activation of the traces which would be equivalent to the familiarity of the retrieved information. The content is a pattern of feature activation made up of the sums of the features of the traces adjusted for their similarity to the cues.

Overall, retrieval in distributed models differs from that of the SAM or spreading activation models. One primary difference is that the distributed models do matching based on features of the item and also on the associations laid down with similar items encoded. While the compound cue and spreading activation models have associations between similar items, they do not go down to the level of also matching based on the encoded features. Nevertheless, the models all have some basic underlying similarities for retrieval. They all encode items of information using some form of associations between the items and other information encoded with them. Because of this, the models all exhibit context dependence. They similarly all display cue-dependence in that the retrieval of information is primarily dependent on the cues given, although it can be mediated by the patterns of associations of the information in memory.

Strategic Retrieval

While the compound cue, spreading activation and distributed models address retrieval issues once the cues have been provided, they do not address many of the issues of human strategic control of retrieval. When we need to retrieve information, we must decide what are the best cues to use for the retrieval. This involves developing a context for the retrieval. Once the information is retrieved from memory, we must also evaluate it to determine if it is what we want. The process of developing the appropriate cues, and evaluating the retrieved information is under strategic control and thus can be viewed as a problem solving process. This section will examine some of the models and issues involved in the strategies of human information retrieval.

Williams (1978) performed a verbal protocol study of people recalling the names of their high school classmates from 4 to 19 years past. These protocols helped him identify some of the processes and strategies the subjects used to recall the information. Overall, he found that retrieval took the form of three phases. In the first phase, subjects would identify a context in which to search, such as a certain location or activity done in the past. Once they had identified a context, they performed a search of that context in order to retrieve the names of the people who were present in that context. Finally once the names were retrieved, they were verified as to whether they were appropriate high school classmates. The three phases were applied recursively though, so in order to help identify an appropriate context, a subject may recursively go through all three phases just to retrieve that context. The retrieval process therefore uses incomplete descriptions of the information, but refines the description as more information is retrieved (e.g. Norman & Bobrow, 1979).

From the subjects' retrieval protocols, Williams identified several strategies used for retrieval. One strategy was to use general associations between people who were related in some way. So, by identifying one person, a subject could then retrieve other people who were associated with that person. Two other strategies were to identify activities people did and locations where people stayed. This information then provided additional context to retrieve people's names. A final strategy also used was to generate common names and then test them as cues to see if there was anyone by that name that the subject could remember. Because of these strategies, retrieval of the relevant names would tend to come in episodes in which a lot of names would be recalled followed by pauses as a new context was generated. Similar results were found in the Walker & Kintsch (1985) study.

The fact that a wide variety of strategies was used shows both the flexibility in strategies of retrieval and also the intricate process used to develop the appropriate context for the retrieval. Williams' results show that search is a reconstructive retrieval process. When retrieving, we construct a partial description of the desired item and then use the description to help retrieve items. From the retrieved items, additional information can then be retrieved using the new items as context.

The Williams study highlights the need for an appropriate context for retrieval. Context effects have also been widely researched in connection with Tulving's encoding specificity principle. The principle states that "a retrieval cue can provide access to information available about an event in the memory store if and only if it has been stored as part of the specific memory trace of the event" (Tulving & Thompson, 1973, p. 359). Thus, information can only be retrieved given the appropriate context with which it was stored.

There are a variety of findings that speak to this issue. Tulving and Thompson performed a series of experiments in which subjects were given word pairs and also had to generate free associates of the second word of the word pairs. They found that the recognition of the free associates that had also been previously studied words was much lower than the recall of words if given the original associate. The results indicate that the retrieval of the words was most successful when provided with the original context in which the words had been studied.

Barclay, Bransford, Franks, McCarrell, & Nitsch (1974) found similar effects with recall of sentences. Subjects were given sentences such as *The man lifted the piano* and were later given cues like *Pianos are heavy* or *Pianos make nice sounds.* They found that subjects were better able to recall the studied sentences when the cue given established the appropriate context. The role of context is not only important for the retrieval, but also for the encoding. Bransford & Johnson (1972) had subjects read identical passages that differed only on whether or not a one sentence context was given (e.g. *This passage is about washing clothes*). The results showed that without the appropriate context, the comprehension and recall of the passages was much lower. Context effects are also evident in the research on state dependent memory (e.g. Eich, 1989). Items studied under a particular mental state tend to be recalled better if the subjects is in the same mental state. These effects have been studied using such states as intoxication (Goodwin, Powell, Bremer, Hoine & Stern 1969) and mood (Bartlett & Santrock 1970).

The role of context and strategies for using context is further evident from ecological approaches to studying memory (e.g. Neisser & Winograd, 1988). As in the Williams (1978) study, ecological approaches tend to study more naturally occurring memory phenomena as opposed to studying phenomena created in the laboratory. Barsalou (1988) studied subjects' retrieval of specific events. He found that during the retrieval, only 21% of the retrieved information described the specific events. Instead subjects tended to retrieve a lot of summarized and extended event information. When the subjects were told by the experimenter not to retrieve the summarized and extended events, the subjects found the process of retrieval much more difficult. This shows that comments and incidental events during retrieval help in retrieving the desired information. These features therefore serve as a context for the subject to retrieve the desired events and by suppressing them, retrieval becomes more difficult.

Kolodner (1983, 1984; Schank & Kolodner 1979) takes a computational approach to retrieval of autobiographical memories. Her model, CYRUS, was designed as a psychological model of retrieval of episodic memory incorporating features of both memory organization and retrieval strategies. CYRUS simulated the knowledge of Cyrus Vance during his experiences as secretary of state. The representation of episodic memory in CYRUS was a conceptual organization using Schank's (1972) conceptual dependency representation to represent events as *event-memory organization packets (E-MOPS)*. In an E-MOP, an event is organized with all of the relevant knowledge structures used to comprehend it. While this organization of information permitted retrieval of episodes organized with their associated knowledge, much of the power of CYRUS came from the retrieval strategies used. CYRUS' strategies were primarily taken from the observations of subjects' retrieval of episodic memory. As in the Williams (1978) study, the observation was that remembering required a progressive refining of a description of an event to be remembered. To retrieve, CYRUS would first choose a context with which to search. It then used such strategies as elaborating the context to fill in contextual details, and using the information retrieved to serve as alternate contexts for the search. This retrieval was used to help comprehend events. As an event was experienced, CYRUS would retrieve relevant contexts which would then provide additional information that would help CYRUS comprehend the event.

The role of proper retrieval strategies is further highlighted in studies of experts storing and retrieving large amounts of information. Rubin (1988) examined the learning of poetry languages ranging from Homeric epics to North

Carolinan ballads. In the cases examined, there were a lot of similarities in the strategies used for encoding and recalling the information. The poems learned were typically very long, on the order of 10,000 lines. Nevertheless, a trained singer could hear a poem once and produce a similar version of it with additions and elaborations if required. Rubin found that the poem learning was similar to the learning of a first language. Generative rules for the style of the poems were learned, and from this, a singer could both encode poems heard, and retrieve the poems using the rules. Thus, these generative rules serve as strategies for organizing the information and for retrieving it when needed.

Experts in a particular domain develop particularly skilled memory in order to function at their high level. Chess masters have the ability to memorize and recall the position of chess pieces on a board in a glance (Chase & Simon, 1973). However, they are only able to do so if the pieces are arranged in a meaningful way. With a random organization of chess pieces, or in a domain outside of chess, chess experts are no better at recall than novices. Nevertheless, the retrieval skill is not dependent on chess knowledge. Ericsson & Harris (1990) trained a subject to be an expert at memorizing chess boards although the subject was a novice chess player. Similarly, Chase and Ericsson (1981) studied a subject who trained himself to memorize strings of up to 80 digits presented in a sequence. Since the subject was a runner, he encoded the digits in chunks of 3-4 digits that represented running times into long term memory. This suggests that effective retrieval is not only dependent on effective retrieval strategies, but also on effective encoding strategies. Thus, specific strategies can be developed for both encoding and retrieving information, although these strategies may be highly person and domain specific.

The strategic aspects of retrieval described in this section point to the fact that retrieval is not just the process of searching a series of memory items given certain cues to see which items best match the cues. Instead, a context for the retrieval must be established through the use of strategies. Individual cues, such as the few words used in memory models for cued recall, are seldom the only items used for retrieval. For a retrieval, a person not only uses the word cue that was provided, but also a lot of the additional context of the situation when the original word was encoded for the retrieval. As information is retrieved, it too can serve as a context for retrieving additional information. Specific strategies are similarly used when initially encoding the information thereby facilitating the use of context to help retrieve it.

Summary

Human memory retrieval depends both on the strategies for retrieval and the automatic retrieval process once the cues and context have been provided. Memory research has typically concentrated on investigating one or the other of these two retrieval aspects. There are a variety of reasons why these processes have been investigated independently. When using automatic retrieval models for investigating such phenomena as cued recall, most of the cues and context are controlled in the experiment so that the only factors involved will be those provided by the experimenter. Thus, very few strategies can be used by subjects. In strategic recall, the strategies used by people may be very specific to certain situations and people. For this reason it is quite difficult to develop a single model for these retrieval strategies and thus, strategic recall is more at the descriptions stage than the modeling stage. Only a few models have incorporated some of the strategic aspects with the automatic retrieval (e.g. Kolodner, 1984; Anderson & Milson, 1989). With a better understanding of both of these aspects and how they interact, more comprehensive memory models may be developed.

Computer Information Retrieval

History and Goals of Information Retrieval Systems

In the modern age, much of the information needed by people can not be stored and maintained in memory so that it is easily accessible. For this reason, there are many ways of storing information outside of memory so that it can be easily accessed when needed. Writing, which dates back to around 6000 B.C., has long been the primary method for both communicating and storing information. While books and libraries provided people with moderately easy access to information, computers have made retrieval of information more efficient. Nevertheless, the same problem has been around as long as there has been recorded information remains. That is: how do we retrieve what we want in an efficient manner while minimizing the retrieval of undesired information? While libraries may have moved from card catalogs to computerized indexes, there still remain many difficulties in retrieving the desired information. With computers providing access to greater and greater amounts of this information, the task of retrieving the correct information becomes more complex.

Information Retrieval (IR) systems encompass a wide range of types of information retrieved and tasks they are designed for. This range includes doing such tasks as textual information retrieval, database management, management information, decision support, and question answering (Salton & McGill, 1983). Nevertheless, all of these systems share the features of having a set of items of information and a set of cues that can be given to retrieve the information. They all also share the goal of retrieving or ranking information items based on the cues provided. Again, there are the two interacting sub-processes of retrieval as discussed in the human memory section. There is a controlled process in which humans must decide what cues to provide to the information retrieval system and evaluate what is returned. Once they have decided on the cues to use, they provide the cues to the automatic process of the IR system to retrieve the desired items. This section will describe some of the models and representations used for the automatic retrieval and then some of the methods used to help users with cue selection and evaluation of retrieved items.

Retrieval Models

Automatic Retrieval

As in human memory retrieval, retrieval of an item is dependent on the similarity between that item and the cues given. The variety of IR models represent different methods of calculating and representing these similarities in order to maximize the effectiveness of the retrieval. Database management systems represents one of the simplest retrieval methods. A database consists of a set of records (e.g. *employees*), each with a set of attributes (e.g. *age, job*). For retrieval, attributes with their desired values are provided (e.g. age >30 AND occupation=welder) and all the records that match exactly those specifications are returned.

However, retrieval of textual or bibliographic information is more complex since many of the words contained in the text may match possible retrieval cues. By treating each document as a set of features, in which each feature corresponds to a term used in the document, textual retrieval can be done using inverted indexing methods (Salton & McGill, 1983; Salton, 1989). In an inverted index each term used in the database is represented as a vector with each element in the vector representing whether a particular document contains that term. Thus if a vector for a term is (1,0,1,0), it indicates that the term is contained in the first and third document. Given a query consisting of terms, the corresponding vectors can be merged by various Boolean operations.

Extensions to this model have incorporated differential term weighting instead of just indications whether or not a term exists in a certain document. Term weighting takes into account the fact that certain terms will have a greater discrimination value for a document than others. For example, a term like *computer* will typically be a less discriminating predictor in an artificial intelligence collection of documents than a term like *language* since the word *computer* would likely be used in almost all documents. The discrimination value of any particular term is highly dependent on the terms used in the stored information, how often they occur and whether they are good at predicting relevant documents. To rate the discrimination value of each term then, typically statistical or probabilistic measures are used (Salton, 1989).

A further advance in IR models is the vector space model (Salton, 1968, 1983). In the vector space model, each document is represented as a vector of weighted terms in a k dimensional space where k is the number of indexing terms in all the documents. Similarly, any query can also be represented as a vector in this space based on which terms were provided. The similarity between this query vector and all document vectors can then be calculated by the distance, or angle between two vectors. Like the distributed human memory models, this similarity is typically calculated using the sum of the products of the vectors, however the similarity value is then typically normalized so that it is between 0 and 1 using a variety of different means such as dividing by the product of the lengths of the query and document vectors (cosine measure).

Another method of retrieval used is the probabilistic retrieval model (Bookstein & Swanson, 1975; Bookstein, 1985). In this model, documents are ranked based on two primary probabilistic relations between terms and documents, the probability that a given term occurs in the relevant set of documents and the probability that a term occurs in the non-relevant set of documents. These probabilities are estimated from the collection of documents and then using Bayes theorem, the probability of the relevance of each document given the terms can be calculated.

The fact that all these methods require using the exact words used in the document to retrieve it highlights one of the deficiencies in current information retrieval techniques. People seldom know which words will describe a document and there is a great variability in the choice of words between people. For example, people choose the same single word to describe a familiar object only about 20% of the time (Furnas, Landauer, Gomez, & Dumais, 1983). Thus, direct keyword matching can fail due to *polysemy* (more than one meaning for a single word) and *synonymy* (many ways of referring to the same concept). One partial solution to this problem is *unlimited aliasing* (Furnas, Landauer, Gomez, & Dumais, 1987) in which any term may be described by a large number synonyms. Some systems have incorporated expanding words into their synonyms to avoid problems of synonymy (Salton & McGill, 1983). While this may solve problems of synonymy, it would still be difficult to distinguish between two documents if they use the same term, but are about different topics (e.g. documents on potato chips and computer chips).

The models described above are based on relations between individual documents and the cues given to retrieve them. Belkin & Croft (1987) differentiate between these *individual* based techniques and *network* based techniques. The network based techniques emphasize the representation of connections between the documents and the retrieval of documents is based on these connections to the other documents. One of the advantages of the network based approach is that the documents retrieved are then not entirely dependent on the exact terms supplied by the user, but may depend on both the terms used coupled with the relationship between the documents.

One network based approach is clustering (Salton, 1968; Voorhees, 1985). By calculating the similarity between documents, using such measures as the cosine between term vectors, documents can be clustered based on their similarity. Salton's SMART system clustered documents into large clusters and then divided those clusters into smaller clusters, and so on, resulting in a hierarchical clustering of the documents. This is then akin to a semantic network of information. Retrieval was done by comparing a query to the document clusters and the documents within the best matching cluster or clusters were then returned to the user, ranked within that cluster. Nevertheless, hierarchical clustering does not permit access of information across clusters, although items in separate clusters may be related. Another use for clustering is to allow users to browse through the clusters and connections. Because these methods lend themselves well to the problems of navigating the clusters by the users, they will be discussed below under the strategic aspects of computer information retrieval.

Connections between documents have been used with other statistical methods for evaluating the relevance between queries and documents. One such method has been to perform a factor analysis on the document by document association matrix (created either through term overlap or human similarity judgements) (Borko & Bernick, 1963; Ossorio, 1965). The associations between documents are then calculated through reducing the dimensionality to the best factors. This tends to then associate the documents based on the primary patterns of association seen in all the documents. Recently, a similar method, Latent Semantic Indexing (LSI), has been used for document retrieval (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Dumais, Furnas, Landauer, Deerwester, & Harshman, 1988). Using the term by document matrix, LSI organizes information into a semantic structure that takes advantage of some of the implicit higher-order associations of words with documents. The resulting structure reflects the major associative patterns in the documents while ignoring some of the smaller variations that may be due to idiosyncrasies in the word usage of individual documents. This permits retrieval based on the the "latent" semantic content of the documents rather than just on keyword matches. The method uses a Singular Value Decomposition to reduce the term by document matrix to represent each document as a vector of factor weights. Tests of LSI have shown that it improves retrieval effectiveness over standard vector models (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990).

A final network representation method used in IR is spreading activation. As this technique has been applied in a way that is very similar to those used in the human memory models, it will be discussed later under the hybrid models.

While there have been improvements in automatic retrieval effectiveness through the years, the overall effectiveness of systems is still not that good. In IR, the effectiveness is typically measured through *recall* and *precision*. Recall measures the proportion of relevant articles retrieved in response to a query out of all the relevant articles, while precision measures the proportion of retrieved articles that are actually relevant. Retrieval systems typically operate around 50% precision and 50% recall, and as recall goes up, precision goes down equally as much (Salton & McGill, 1983). Thus, a user of an information system will seldom receive a complete list of all documents in a database that are relevant to his query without a lot of non-relevant documents. While there remain methods that can improve the automatic retrieval, another way to improve retrieval may be through helping the user develop queries and interacting with the system to find what he or she wants. These methods constitute the strategic aspects of information retrieval.

Strategic Retrieval

In retrieval it is often not clear to the user how or what can be retrieved. Due to problems of synonymy, users may not know which terms to use to specify what they want. They also may not know what terms to use because they are not familiar with what type of information they can retrieve from the database. There are also problems with the actual interaction with the system; users may not know how to form a query or use the query language. Thus, a user interacting with a retrieval system may need to use some conscious strategies. To ease these strategic problems, there are a variety of methods used in Information Retrieval. Two primary methods are, ways of interpreting what a user wants so that the computer can retrieve it, and ways of letting the user browse through the information.

Often users are not clear about exactly what they want to retrieve. This may be due to the fact that there are many ways of specifying what they want, or that they do not have an exact idea of what they are looking for. One solution to this problem is to use expert intermediaries (Borgman, Belkin, Croft, Lesk, & Landauer, 1988). An expert intermediary is a human with knowledge of the database, the domain and the IR system. A person desiring information can make his or her queries to the intermediary. The intermediary can then re-interpret the person's queries into queries for the IR system, make the queries and return the information to the person. An advantage of this method is that people do not need to learn to use an IR system, and can just interact with another human to specify the problem for which the information is desired. Even with intermediaries to enter better queries, the information systems still may do poorly at retrieval. Blair and Maron (1985) studied paralegal intermediaries using a legal information retrieval system. They found that over all the searches, fewer than 20% of the documents returned were relevant to the queries. Nevertheless, the lawyers who had specified the original queries believed that they had retrieved about 75% of the relevant items. This result highlights the fact that retrieval systems are still not that efficient and even with intermediaries, the retrieval procedure may be doing much worse than the users think. This failure to retrieve efficiently can have dire consequences in some areas, such as failing to find certain relevant articles when writing a review, or not locating an earlier law case that would be relevant to a current one.

Natural language retrieval systems have also been used to ease the problem of specifying the queries (Warner, 1987). In a natural language system, users can specify the query in their own language and it is interpreted by the computer. This permits the user to apply the problem of retrieval and vocabulary in terms of one that is already familiar (human language) and provides a structure to the retrieval process. Thus, a user can use the same strategies he would use when describing what he wants to another person. Nevertheless, although these systems have been shown to be feasible, they have not yet developed efficient systems that are flexible enough to deal with large, diverse sets of documents and a wide range of possible words that could be used.

A second retrieval method that helps the IR system figure out what the user wants is relevance feedback (Salton & McGill, 1983; Salton & Buckley, 1990). In relevance feedback, the IR system returns a list of documents after an initial query. A user can then identify the retrieved documents as either relevant or not relevant. This relevance information is then used as a query to perform an additional retrieval. Since the new query contains more refined information about what the user considers relevant, it tends to return more relevant items and fewer non-relevant items. Salton & Buckley found about a 50% improvement in average precision on a series of standardized text collections. The relevance feedback method has several advantages. Users are no longer dependent on providing accurate terms for their queries, but can just indicate documents that appear relevant to their interests. This reduces some of the effort of having to know what is in the database in order to retrieve the information. It also allows users to have a lot more control of the search process. They can specify the direction they want their searches to go, and can iteratively narrow down their search by specifying increasingly more relevant documents.

Relevance feedback has a lot of similarities to information browsers. Information browsers use a set of rich connections between documents to allow a user to navigate through the space of information. With these connections, a user can select one item of information and after reviewing it, can select other items that are associated with it. In this manner, a user can move through a set of information, just selecting the related items that appear relevant. An early system to use browsing was ZOG (Mantei, 1982; McCracken & Akscyn, 1984). In ZOG, users navigated trees with cross connections of information through using a menu system. Nevertheless, users would often get lost in the network of information since there was no overview or map of the entire set of information. A more recent browsing retrieval system, I^3R (Intelligent Interface for Information Retrieval) incorporates more features to permit flexible navigation of information (Thompson & Croft, 1989). In I^3R users can navigate using a variety of links such as, synonym, related-to and nearest-neighbor links. The system actively updates links based on a model of the user and the goals of the current search. To avoid getting lost in the information, it also provides an overview map of where a user has been and a map of the current information in the neighborhood.

These networks of interconnected pieces of information have also been used in the form of hypertext. The original vision of hypertext came from Vannevar Bush's Memex system (Bush, 1945, 1967). In Memex, all information would be stored by association and a user could create connections between information items and navigate these "trails" of information. Hypertext refers to a wide variety of types of retrieval systems, but is typically characterized by non-linear text with many links between the text nodes. While books are typically written in a linear form, hypertext permits the expression of ideas with more associations to other ideas than could be expressed in a linear book. Thus, a short text on Paris may have connections to other texts on such topics as, France, Paris history, subway maps, and other texts about Paris. These cross references between associated items therefore let a reader choose the directions that are of primary interest at the moment rather than having the author dictate the order in which information should be read. Nevertheless, there is a certain level of overhead in performing the navigation and decisions of where to go next that is not found in reading linear text (Conklin, 1987).

A number of hypertext systems have been developed such as, the Symbolics Document Examiner (Walker, Young & Mannes, 1989), Hyperties (Schneiderman & Morariu, 1986), Superbook (Egan, Remde, Gomez, Landauer, Eberhardt, & Lochbaum, 1989) and NoteCards (Halasz, Moran, & Trigg, 1987). There has also been some research on the use of these systems indicating the strengths and weaknesses of navigating a textual information space. Overall, the studies indicate that there are some advantages in letting a user search for his or her own information in the space, but there are nevertheless, only a few systems that are superior to using the linear text in books (Nielsen, 1989).

The ability to navigate through a network of documents has been examined through research in graphical data management. Herot (1980) discusses a graphical data base in which users can traverse the space to examine information and can also zoom in on information to obtain greater detail. Fairchild, Poltrock & Furnas' SemNet (1988) used a three-dimensional graphic representation of information in which pieces of information were clustered based on similarity. Users were able to soar through the space in order to search for information. This allows users to see relationships between items in the semantic network and move to areas where the most relevant items appear to be located.

Summary

Information Retrieval tries to solve the problem of providing efficient access to the large stores of information that can be stored electronically. In this manner, IR can serve as an extension to our own memory. Like human memory, the IR retrieval process involves both an automatic retrieval component and a strategic component which can take a variety of forms depending on the task. The automatic retrieval component is based on the way the information is represented and, once a cue is given, is not under the control of the user. The strategic component of IR shows a lot of promise in improving retrieval efficiency. The automatic component can only retrieve based on exactly what is specified by the user. Since users are seldom experts in using the system or on what information is contained in the database, the strategic component permits users to provide additional information on what they want as they become clearer on the structure of the information.

The automatic retrieval models show many similarities to those used in human retrieval. The vector representations of documents are close to those used in the distributed human memory models. In IR though, the represented features are just the terms used in the text. The features used in the Latent Semantic Indexing avoid some of the problems of being direct text objects, by instead representing higher level text associations. This makes it analogous to representing higher order features of information, as is done in distributed models for schema abstraction. The strategic aspects in IR also cross into the boundaries of psychology using a lot of the same strategies that humans must use. As in the Williams (1978) and Kolodner (1984) work, in relevance feedback some information is returned and then a user evaluates that information and uses some of it as cues to help narrow down the search to the desired information.

Hybrid Models

With the many similarities between human retrieval and computer information retrieval, there are some models or systems that have incorporated features of both. These models could be called hybrid models in that they do information retrieval, but are based on applying psychological models of retrieval. These models, therefore, cross the boundaries of being just a computer information retrieval model, or just a human memory model, but have taken features from both fields in order to have an effective retrieval system. These hybrid models illustrate the complementary nature of human and computer retrieval models. As in the previous sections, these models will be examined based on their performance in automatic or strategic retrieval.

Automatic Retrieval

One of the main ideas from human memory literature that is used in information retrieval is the concept of differential associative connections between items of information. For this reason, there have been a variety of retrieval models using some form of spreading activation. One of the earliest of such models was the ACORN (Associative Content Retrieval Network) models (Giuliano, 1963). ACORN was a wired network associating index terms with documents. Strength of associations were represented by different resistors and associations were set between all terms to documents and also between all terms to other terms. Positive and negative voltages could be applied to terms that were relevant and not-relevant and the voltage for each document could then be read to determine which document received the highest activation. Because terms were connected to other associated terms, the system could retrieve documents that did not have the exact terms activated, but ones that contained terms that were highly associated to other terms. While the largest ACORN network was only 240 nouns and 240 documents, the results of the model did indicate the feasibility of such systems.

Jones (1986; Jones & Furnas, 1987) applied a spreading activation model to retrieve computer files in his "Memory Extender" (ME) personal filing system. In ME, a stored file has bi-directional associations with terms that describe the file. The strength of these links are based on how well the term describes the file. ME also uses a set of context terms that can be associated with any file that describe the context in which the file is used. When a user wants to retrieve a file, the user activates certain terms that describe the file and also terms that express the context of when the file was used. Activation spreads from these terms to files and then back from those files to other associated terms and then back to their associated files. The files can then be ordered based on their activation for retrieval by the user. Thus, in this system, the spreading activation only spreads for two steps. The two steps of spread, nevertheless, permit the activation to spread to files that may only be indirectly associated through the terms provided, thereby avoiding some of the problems of not specifying the exact words.

GRANT (Cohen & Kjeldsen, 1987) is a spreading activation model for finding sources of funding given terms describing a research proposal. GRANT treats information as a semantic memory of research concepts. It contains 4500 research concepts that are interconnected based on a set of relations such as *is-a, has-setting,* and *has-instance* links. The concepts are in turn connected to funding sources. When certain concepts are activated, activation is then spread using the constraints that activation may only spread a distance of four links and high fan concepts (e.g. *science, person*) will not spread activation.

While these systems show some promise for combining a psychological model with information retrieval, Salton and Buckley (1988) have evaluated spreading activation models against standard vector processing information retrieval models. While the spreading activation model did better than simple vector process models, they did not do as well as vector process models that normalized document length and used differential term weighting. Nevertheless, their spreading activation model was only based on term overlap for associations, and more complicated types of associative links as in the GRANT system may produce better retrieval performance. Spreading activation performance has also been questioned as to whether it will scale up to larger sets of information (Shrager, Hogg & Huberman, 1987; Huberman & Hogg, 1987). As the information sets get larger, there may be abrupt phase transitions in the performance of the spreading activation. This may affect the setting of parameters for the spread based on the size of information store.

As can be seen from the ME and the GRANT retrieval system, both use a form of constrained spreading activation. In the ME system, activation only spreads three links (term-document, document-term, term-document); while in GRANT, activation was limited to 4 links. This is still more levels of spread than found by DeGroot's (1983) findings for human priming, but may be in line with the findings of Balota and Lorch (1986) that activation does not spread a great distance. Nevertheless, for information retrieval models, it must be remembered that the key to a successful model is not psychological plausibility, but success at retrieval.

A similar approach to spreading activation models of retrieval are connectionist retrieval models. Mozer (1984) applied McClelland and Rumelhart's (1981) model of word perception to document retrieval. In his model, there were weighted bi-directional connections between terms and documents and inhibitory connections between documents. Thus terms would tend to activate other terms only when they tend to co-occur in documents and activate other related documents. Documents were then ranked based on their activation. Belew (1989; Rose & Belew, 1989) developed a connectionist information retrieval system connecting terms to documents and documents to the authors. Although weights were initially set by term frequency, it differed from Mozer's work in that it used a learning algorithm to modify weights based on a user's positive or negative feedback. This permitted the system to change its connections as it learns the relevant associations from the user. It has recently been applied to retrieving legal documents along with an interface that permits users to do relevance feedback. While the connectionist models are similar to the spreading activation models used in ME and GRANT, the connectionist models tend to use more advanced threshold or "squashing" functions for the activation of a node and let the activation spread until the network settles.

A final hybrid model differs from all others in its approach. Anderson's (1990) probabilistic model of human retrieval, as described at the start of the article combines some of the features of both information retrieval with the constraints of human memory. It serves as a model of human retrieval although it is built using some of the features of Burrel's (1980) probabilistic model of library borrowing. Incorporating history of usage and context factors, his model uses some of the environmental features that have typically not been considered in other psychological models. This is thus a different approach than the other hybrid models, which applied psychological principles to design an information retrieval model. Instead, this model applies some of the principles from information retrieval to help constrain the design of a psychological model of human retrieval.

Strategic Retrieval

It is more difficult to describe a hybrid model of strategic retrieval. Since information retrieval is partially based on evaluating users interacting with the system, most strategic aspects of information retrieval are strongly based on psychological constraints. There are, nevertheless, some strategic features and models that should be mentioned because of their strong ties to psychological theory for their development.

Since strategies play a major role in using information retrieval systems, there have been a number of studies of users of retrieval systems. These studies include studies of how people make decisions when doing a retrieval task (Blackshaw & Fischoff, 1988), analyses of query languages (Greene, Devlin, Cannata & Gomez, 1990; Davis, 1989), and the mental models and knowledge necessary for using a system (Belkin, Brooks & Daniels, 1987; Borgman, 1986). Overall, these studies permit developers to refine their systems taking into account the types of errors and misconceptions made by users. This in turn makes the actual system development guided more by the constraints of human abilities.

One system that was designed primarily based on psychological models of the users abilities and strategies was the RABBIT system (Williams, 1984). In his studies of subjects retrieving information from their own memory (Williams, 1978), Williams found that people go through the iterative process of starting off with a partial description of what they want, retrieving a general context and then retrieving the information within that context. The RABBIT system permitted users to do the same using a procedure, called *retrieval by reformulation*. Using retrieval by reformulation, a user interactively refines partial descriptions of his or her target item by criticizing successive examples. To do this, a user initially makes a query by constructing a partial description of the item in the database for which he or she is searching. The system then produces a description of an example instance from the database that matches the user's partial description. The user can then refine the query by choosing any attributes shown in the example instance and incorporating those descriptions, or variations of them, into the partial query, thereby reformulating the initial query. This process of reformulating the query continues until the desired information is retrieved.

This retrieval process permits users to use recognition memory rather than having to recall the exact description of what they are looking for. Thus, the retrieved example instances serve as a template to further aid the user in describing their target item. The advantages of such a system include that users do not need to have a complete concept of what they want to retrieve, but can interactively refine their query until they have found what they need. This solves some of the problems in database retrieval by aiding the user to learn the structure of the database and by showing how to refer to items in the database.

There have been several extensions to the RABBIT work. ARGON (Patel-Schneider, Brachman, and Levesque, 1984) extended the RABBIT work to incorporate a frame-based knowledge representation organized in a hierarchy. HELGON (Fischer & Nieper-Lemke, 1989) is an extension of ARGON incorporating a graphical display of the frame hierarchy of information. This display permits users to see how the information is organized in the database, thereby facilitating the user's navigation through the information. This, coupled with the retrieval by reformulation, procedure then helps users find what they want through navigating the information space and interactively refining their query.

RETRIEVE (Foltz & Kintsch, 1988; Fischer, Foltz, Kintsch, Nieper-Lemke & Stevens, 1989) has tried to apply the psychological constraints of both strategic and automatic recall to the development of an information retrieval system. In RETRIEVE information items were connected to terms based on their term frequency. Information items were also connected to all other information items based on their amount of term overlap. For retrieval, cues provided by the user were matched against stored information using the Raaijmakers & Shiffrin (1981) method for combining the strengths of cues. Information retrieved was displayed and users could interactively add or delete information from their queries based on what they retrieved in order to narrow down the query towards their desired information. If the exact information was not retrieved, users could use a spreading activation method to spread activation from selected items to associated items. This then permitted retrieval of information that was not directly specified with the terms provided, but closely associated. In this manner, RETRIEVE combines the psychological features of spreading activation and cue-based memory retrieval models along with providing mechanisms for users to employ reformulation strategies in an information retrieval system.

Conclusions

The purpose of this review is to point out some of the relationships between human and computer retrieval systems. From examining these relationships, we can determine what are some of the underlying features of retrieval. These features can then point out some of the implications for successful retrieval in the human and the computer domain.

Goals of retrieval

One reason to compare two fields of science is to determine whether they share similar goals. From one point of view, human memory modeling and computer information retrieval modeling do not share a common goal. Human memory models are developed with the intent of explaining how the mechanisms of memory work through modeling human memory phenomena. Computer information retrieval has the goal of building the most efficient systems to retrieve the desired information for any user. Nevertheless, the goals of these two fields do converge.

One approach is to treat human memory as an optimal (or near optimal) system as in Anderson (1990). From this point of view, we can look at whether both systems are optimized for similar tasks. In both cases, the task is to find the needed information when you need it and not at some later time (or not at all). For computer developers the motivation to have it do this is financially motivated, the better the system, the better it will sell. On the other hand, for humans, a better memory system would permit ease in competition (say for locations of food, or in the present day, remembering your presentation for a job talk). Thus, the human brain has evolved to such a level to make retrieval an efficient process. This then means that both approaches should be concerned with developing an efficient retrieval model.

The information that must be retrieved is also similar for both systems. Information retrieval systems serve primarily as extensions of our own memory. They hold a lot of the information we have not had time nor ability to study, but will need to function in our society. In ancient Greece, before information could be easily recorded and carried, mnemonic methods such as the method of loci were used to store and recall large amounts of this information. We could store all the phone numbers of our friends now, but it is far simpler to have them written somewhere within easy access. Thus, a lot of the information stored in a retrieval system is the same as what we would have to store in our head, it is just more convenient to store and access it outside of our own memory.

Given that both fields are concerned with developing an efficient retrieval model, and much of the same information needs to be stored and retrieved, there should be similarities in their approaches. These similarities reflect that although researchers in these two fields have worked independently on the problems of retrieval, they have thus far ended up with analogous approaches to the problem.

Models and representation of information

The retrieval models described in this article use a limited number of methods for creating the connections between cues and memory items, matching the information and representing the memory items. These features are outlined for some of the models in Table 1.

Retrieval Model	Example	Domain	Matching Method	Origin of connections	Representation
Compound Cue	SAM Gillund & Shiffrin (1984)	Human	Activation of items associated with cues	Cues to item strength based on STM co-occurrence	Network
Distributed	TODAM Murdock (1982)	Human	Correlation of item feature vectors to cue vector	Feature similarity	Feature
Spreading Activation	ACT* Anderson (1983)	Human	Activation of items spread from cues and other items	Based on degree of association between items	Network
Rational Memory model	Anderson (1990)	Human	Based on probabilities of history and context	Based on probabilities of history and context	Network
Inverted index	Salton & McGill (1983)	IR	Boolean operations on feature vectors	Feature similarity (term overlap)	Feature
Vector Space	SMART Salton & McGill (1983)	IR	Distance between item and cue vectors	Feature similarity (term overlap)	Feature
Probabilistic	Bookstein & Swanson (1975)	IR	Probability a cue is in a relevant item vs. prob. in a non relevant item	Feature similarity (term overlap)	Feature
Spreading Activation	Memory Extender Jones (1986)	IR	Activation of cues spreads to cues and other items	Semantic network of associations, based on feature (term) similarity	Network/ Feature
Connectionist	Rose & Belew (1989)	IR	Activation of cues spreads to other cues and items	Feature (term) similarity and user feedback	Network/ Feature

Figure 1. Features of retrieval models.

For memory items to be retrievable, there must be connections between the items and all the possible cues that could be used to retrieve it. As seen in Table 1, the origin of these connections for all of the models derive from either similarity of the features in the cues and items or are based on the co-occurence of cues and items in STM during encoding. In this manner, the connections in all models are based on temporal or featural co-occurence. Once these connections have been created, the models can use matching methods to retrieve items based on cues. In some of the models, the matching methods use the connections directly as the retrieval method (e.g. compound cue model, inverted index). However in most models, the retrieval is also based on what other information is encoded in memory. Thus in spreading activation, retrieval is based both on the cue connections and the inter-item connections. Similarly in the distributed models with a composite associative trace, retrieval is based not only on the feature similarity between a cue and items, but also on the relationship between memory items added on to the composite associative trace.

Based on the models discussed, there also are two ways of representing the memory information, as a network, or as a set of features. In a network representation, the information in memory is typically represented as discrete items of memory. These items can then connected to each other through representing activations between the items. In the feature representation, information is represented as a set of features rather than discrete elements. Thus, rather than representing item A as being strongly connected to B and weakly connected to C, it could be represented as sharing many features of B and few features of C. In this manner, connections between information in network representations are typically made through examining feature overlap. Therefore a network representation can be derived from a feature representation, as is done for spreading activation models. Overall, from Table 1, we see that there a only a few methods used for representing memory information, the connections between it and the cues, and ways of retrieving the information from memory.

A significant similarity between human memory and computer information retrieval is the strong dependence on cue-based retrieval. In order to access information, some sort of initial cue must be provided to start off the retrieval. Cues provide the means to do fast matching of the description of the desired information with all the information stored. Based on this matching, some of that stored information will be retrieved. This automatic process permits the retrieval to be narrowed down to a small set of possible pieces of information without a large load on the human system. An alternative retrieval method could work something like browsing. We could start in some random or predetermined spot of our knowledge store and move through it on associative links until we find what we want. In small knowledge stores this may be possible, but in something that has as much information as human memory, there would be too much overhead in navigating to the correct place for it to be efficient. Thus, for browsing to be useful, one would need to use cues to put the person in the right neighborhood of information before that person has a chance of finding the desired information in an efficient manner.

For cues to match items of information, they must match some features of an item. As discussed above, features play a major role in the representation of the information . These features permit information to match cues to a greater or lesser degree based on how many features match and how salient each feature is. The richer the classification of features, the better the specificity in retrieving what is desired. With just a single feature describing an item, a cue must match directly to that feature. With more features describing the item, a wider range of cues can be used and a more pronounced differentiation between items can be made. In many of the models discussed, this feature representation is done through vectors representing each item of information (e.g. distributed memory models and vector retrieval models).

Features also permit associations between items of information. Typically, information that shares more features will tend to be more associated. Associative connections have been used widely in both types of systems and provide a lot of power to retrieval. Both storage and retrieval are very much based on these associations. Information is best encoded in memory if strong associations are made along with it (e.g. Hyde & Jenkins, 1973). Most of the models described have some form of associations between items of information (as well as associations to cues). These inter-item associations mean that items that share strong associations will tend to be recalled together. The mechanisms for doing this differ somewhat. In spreading activation models, it is the spread of activation that determines which items will be recalled together, while in a vector space, it is the closeness of the two vectors that will make that determination. Nevertheless, at a higher level, these are related since the spread of activation and the placement of information in a vector space is all dependent on the similarities of features of the items.

The encoding of information and the creation of the associations may be one of the big differences between human and computer retrieval. Information retrieval uses very simple encoding schemes. Associations are typically made based on surface features such as word overlap. This does not take into account the semantic relationships that are present but not expressed in the choice of word features. Thus, a description on how to delete a file and how to erase a directory may be strongly related, but share very few surface features. For this reason, it is important to take into account some of the higher level associations that may be present. Humans are able to do just that. Encoding of information in memory is not as much feature based as meaning based. Since a human has a good understanding of the semantics of deleting and erasing, they will tend to be highly associated based on their higher level semantic features, not their surface structure. The organization of human memory will therefore have a richer and more appropriate connections than can be created by a computer.

Another key difference in encoding is the context. As discussed in the human memory section, humans use context in both encoding and retrieval of information. This context can take many forms and is typically a combination of a variety of factors. It can be the episodic information of when the information was stored, the current emotional or physical state, and other information that was stored around the same time. This provides a variety of cues and associations that are stored with any piece of information and allows many possible retrieval paths to the information. Computers, on the other hand, are much more limited in their encoding of the information. Very little context of any situation is encoded with a piece of information. This results in a much more impoverished classification of the information making it harder to retrieve. As Tulving (1983) points out, encoding specificity allows information to be retrieved only under the correct context. Without this context, computers can not use encoding specificity.

One of the features of context is that it is highly person specific. The context I use to encode a piece of information will be different from that used by another person. Although the information may be essentially the same, the context of when I store it and the associations I make to it will be quite different. For a computer retrieval system to work for more than a single person, it must make generic associations free from the person specific context. While the additional context associations for the person who made them may improve retrieval, they likely would not be helpful and may even be detrimental for another person. For these reasons, computer retrieval systems will not be able to encode the wealth of context without some specific user profiling. Some systems have used user profile information as a way of choosing the appropriate context and associations. I^3R (Thompson & Croft, 1989) keeps track of the user's sessions to build a model of the user's domain knowledge. Foltz (1989) evaluated a user profile method that kept track of user preferences for news articles to predict what type of articles they would like to see in the future. These user profiles are nevertheless still a long ways away from the level of applying all the context or creating the proper connections that humans use when encoding any piece of information.

Another difference between the two systems is decay. In humans, information that is not used often decays quickly. This means that additional effort or recoding of the information may be necessary to bring the information back. Computer systems typically do not use any form of decay. Thus, information remains stored whether it is used or not. While the vision of never forgetting anything so that everything is accessible may look enticing, there are some problems of not having decay. Luria's (1968) study of the mnemonist S, presents a good example of a person with very little decay of information. While he was able to store vast amounts of information using his highly synesthetic encoding strategies, the information stayed longer than it was needed. Because of these associations to old information, any situation he would encounter would trigger recall of the old information. This led him to being very distracted, with any event causing recall of many past events. In a computer retrieval system, the same thing can happen. As more and more information is stored, the older and unused information will become less relevant to what people want to retrieve. Since this information will still match strongly, its retrieval will interfere with the retrieval of newer, and more relevant information. Some computer systems have implemented some weighting features based on the history of usage of the information (e.g. Burrel, 1980).

Thus, computer and human models of retrieval have many similarities. They use similar representations of the information and similar methods of matching the information. There are nevertheless some large differences between the two. Computer retrieval does not have the rich encoding of information and does not use as much context for this encoding as do humans. This results in a much more impoverished representation of the information, providing fewer retrieval paths to the information. This key difference appears to be due to the fact that humans have a semantic feature based representation of information while computers use a surface feature based representation of the information. While the human's semantic representation may indeed be derived from just surface feature overlap, the number of features that can be used to encode any piece of information is many times greater for a human than a computer. Even with the ability to use all these features, a computer would need to use good encoding strategies in order to determine what features of any piece of information should be stored as context. Again, in terms of encoding the information, the human excels over the computer. These differences may account for why information retrieval systems still retrieve only small amounts of the relevant information, while human memory is very fast at retrieving relevant information.

Retrieval strategies

Humans are experts at using strategies to store and retrieve information from their own memory. They are very familiar with the structure of the information stored and the cues that can be used to retrieve it since they did the initial encoding. This is not the case in computer retrieval. Users are seldom familiar with what information is available, and how it is organized in the computer. This unfamiliarity means that they will not be as good at developing retrieval cues to give to the system. Similarly, when information is retrieved, they will not be as able to judge the relevance of the retrieved information since they do not know how much other relevant information is still not retrieved. Users are also not familiar with the ways of specifying the cues. Since many retrieval systems are term based, the exact terms must be specified to get the desired information. Without a familiarity of the information stored, it is hard to know which terms should be used.

Information retrieval systems are currently adding some of these features, such as browsing, relevance feedback and retrieval by reformulation. These methods permit users to employ familiar strategies for retrieving the information. Retrieval by reformulation allows users to interactively refine their query as they become more familiar with the structure and semantics of the stored information. Relevance feedback similarly lets users narrow down the context for where they want to search by providing additional context cues for the search. Browsing is somewhat different in that it permits users to navigate through a set of information, typically staying within the same context. In this manner, browsing can be seen as a limited form of spreading activation, in which the user controls a single pointer of activation that spreads in the direction based on which information item the user moves to. Thus, a user is in control of the search process, making the decisions on which direction to go to find a relevant piece of information.

Implications for both fields

The similarities between human and computer retrieval point out the complementary nature of these two systems. Information retrieval systems function as extensions of our own memory. For this reason, there should be some similarities in how information retrieval systems are used and designed. One such feature is to allow users to employ familiar retrieval strategies when using the system. As described above, some systems have implemented some of these strategic aids. Nevertheless, computer retrieval systems do not make the large number of associations between information items as do humans. This is due to the fact that retrieval systems have minimal encoding skills and must rely on surface feature or term based encoding as opposed to the human's semantically based encoding. The systems also do not use much context in retrieval. Typically, retrieval is performed using just a set of words provided by a user, whereas human retrieval uses many additional contextual cues, such as temporal and locational information. Thus, improvement in information retrieval models can be made through tailoring the systems to incorporate greater semantic relationships in encoding and to use greater contextual information for retrieval such as user profiles. Psychological models of memory and of retrieval strategies highlight the current abilities of the human retrieval system and can provide directions for information retrieval systems to augment the human's ability to find information.

Conversely, human memory retrieval can learn from the insights into computer information retrieval. Information retrieval researchers have worked to develop systems that retrieve information efficiently. They have taken into account such features as the frequency of occurrence of information, associative and feature based representations, and human retrieval strategies. With this effort, these models have ended up being fairly similar to those used in psychological memory modeling. This suggests that both fields are searching for analogous solutions. Developments of efficient information retrieval models may provide some guidance for psychologists as to what types of representations and algorithms can be used to model efficient human memory retrieval. The improvement that information retrieval researchers find when using relevance feedback suggests that the automatic process of retrieval need not be entirely efficient. Information can be retrieved once with a set of cues, but then with strategic intervention of providing information about the relevance of retrieved items, a query can be quickly narrowed down to finding the relevant information. This may be the case too in human memory. The actual automatic retrieval process need not be entirely efficient, since we have excellent skills at developing the context, and revising the cues to narrow down on the desired information. There are few memory models that incorporate both the automatic retrieval component and the wealth of strategies that are used to develop and revise the cues.

In information retrieval, one of the key methods to developing a retrieval system is based on the task and environmental factors. As suggested by Anderson's (1990) recent work and in the ecological approaches to human memory, human memory models in the future may examine more of what types of tasks human memory is designed to do, and what are some of the environmental constraints in which the memory must operate.

The insights from both of these fields should provide guidelines to aid in the development of better retrieval models. These models will more likely fall under the category of the hybrid models described here as they will have incorporated some of the best features and constraints of the two types of retrieval. In the long run, these insights will then provide both a better understanding of how our own retrieval system works, and how to develop better external retrieval systems.

References

Anderson, J. R. (1972). FRAN: A simulation model of free recall. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol 5). New York: Academic Press, 315-378.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. & Bower, G. H. (1973). *Human associative memory.* Washington, DC: Winston and Sons.

Anderson, J. R. & Milson, R. (1989). A rational analysis of memory. *Psychological Review.,96*(4), 703-719.

Anderson, J. R. & Schooler, L (in press).

Balota, D. A. & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciations but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*(3) 336-345.

Barclay, J. R., Bransford, J. D., Franks, J. J., McCarrell, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior, 13,* 471-481.

Barsalou, L. W. (1988). The content and organization of autobiographical memories. In U. Neisser & E. Winograd (Eds.) *Remembering reconsidered: Ecological and traditional approaches to the study of memory.* Cambridge: Cambridge University press.

Bartlett, F. C. (1932). *Remembering.* Cambridge: Cambridge University Press.

Bartlett, J. C., & Santrock, J. W. (1979). Affect-dependent episodic memory in young children. *Child Development, 50,* 513-518.

Belew, R. K. (1989). Adaptive information retrieval: Using a connectionist representation to retrieve and learn about documents. In *Proceedings of the 12th International Conference on Research and Development in Information Retrieval,* Cambridge, MA, June.

Belkin, N. J., Brooks, H. M., & Daniels, P. J. (1987). Knowledge elicitation using discourse analysis. *International Journal of Man-Machine Studies, 27,* 127-144.

Belkin, N. J., & Croft, W. B. (1987). Retrieval techniques. In M. E. Williams (Ed.) *Annual Review of Information Science and Technology,* 22, 109-145. Elsevier.

Blair, D. C., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document retrieval system. *Communications of the ACM,* 28, 289-299.

Blackshaw, L., & Fischhoff, B (1988). Decision making in online searching. *Journal of the American Society for Information Science, 39*(6), 369-389.

Bookstein, A. (1985). Probability and fuzzy-set applications to information retrieval. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology,* 20, White Plains, NY, Knowledge Industry Publications, 117-151.

Bookstein, A., & Swanson, D. R. (1975). Probabilistic models for automatic indexing. *Journal of the American Society for Information Science, 26*(1) 45-50.

Borgman, C. L. (1986). The user's mental model of an information retrieval system: an experiment on a prototype online catalog. *International Journal of Man-Machine Studies, 24,* 47-64.

Borgman, C. L., Belkin, N. J., Croft, W. B., Lesk, M. E., & Landauer, T. K. (1988). Retrieval systems for the information seeker: Can the role of the intermediary be automated? Panel Session *Human Factors in Computing Systems, CHI'88 Conference Proceedings* (Washington, D.C.), ACM, New York, May 1988, 51-53.

Borko, H., & Bernick, M. D. (1963). Automatic document classification. *Journal of the ACM, 10*(3), 151-162.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 367-374.

Brown, J. (1958). Some test of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12-21.

Burrel, Q. L. (1980). A simple stochastic model for library loans. *Journal of Documentation, 36,* 115-132.

Bush, V. (1945). As we may think. *Atlantic Monthly, 1,* 176 101-108.

Bush, V. (1967). *Science is not Enough.* Chapter V- Memex revisited. William Morrow & Co.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology,* 4, 55-81.

Cohen, P. R., & Kjeldsen, R. (1987). Information retrieval by constrained spreading activation in semantic networks. *Information Processing and Management, 23,* 4, 225-268.

Collins, A. M., & Loftus, E. F.(1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240-248.

Conklin, J. (1987). Hypertext: An introduction and survey. *Computer,* Sept. 17-41.

Davis, S. J. (1989) Usability of SQL and menus for database query. *International Journal of Man-Machine Studies, 30,* 447-455.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science. 41*(6), 391-412.

DeGroot, A. M. E. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22,* 417-436.

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., Harshman, R., (1988). Using latent semantic analysis to improve access to textual information. *Human Factors in Computing Systems, CHI'88 Conference Proceedings* (Washington, D.C.), ACM, New York, 281-285.

Ebbinghaus, H. (1885). Memory: A contribution to Experimental Psychology (translated by H. A. Ruger & C. E. Bussenues, 1913). New York: Teachers College, Columbia University.

Egan, D. E., Remde, J. R., Gomez, L. M., Landauer, T. K., Eberhardt, J., & Lochbaum, C. C. (1989). Formative design evaluation of "Superbook". *ACM Transactions on Office Information Systems*, 7(1), 30-42.

Eich, E. (1989). Theoretical issues in state dependent memory. In (H. L. Roediger & F.. I. M. Craick, Eds.) *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*. Hillsdale, NJ: Erlbaum.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89*, 627-661.

Eich, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review, 92*, 1.1-38.

Ericsson, K. A., & Harris, M. S. (1990,November). *Expert chess memory without chess knowledge: A training study*. Presented at the 31st Annual Meeting of the Psychonomic Society. New Orleans, LA.

Estes, W. (1976) *Handbook of learning and cognitive processes vol IV*. Hillsdale NJ: Erlbaum.

Fairchild, K. F., Poltrock, S. E., & Furnas, G. W. (1988). SemNet: Three-dimensional graphic representations of large knowledge bases. In R. Guindon (Ed.), *Cognitive science and its application for human-computer interaction*. Hillsdale NJ: Erlbaum.

Fischer, G., Foltz, P. W., Kintsch, W., Nieper-Lemke, H., & Stevens, C. (1989). Personal information systems and models of human memory. Department of computer science Technical report, University of Colorado, Boulder, CO

Fischer, G., & Nieper-Lemke, H. (1989) HELGON. *Proceedings of the ACM Conference on Human Computer Interaction CHI 89*. Austin, Texas, ACM.

Foltz, P. W. (1989). Using latent semantic indexing for information filtering. In R. B. Allen (Ed.) *Proceedings of the Conference on Office Information Systems (COIS)*. Cambridge, MA

Foltz, P. W., & Kintsch, W. (1988). An empirical evaluation of query by reformulation on HELGON. In A. A. Turner (Ed.), *Mental Models and User-Centered Design, Workshop Report (Breckenridge,CO)* Institute of Cognitive Science, University of Colorado, Boulder, CO, TM 88-9, 9-14.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1983). Statistical Semantics: Analysis of the potential performance of key-word information systems. *Bell Systems Technical Journal, 62*(6), 1753-1806.

Furnas, G. W., Landauer, T. K., Gomez, L. M., & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM, 30*(11), 964-971.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1-67.

Goodwin, D. W., Powell, B., Bremer, D., Hoine, H., & Stern, J. (1969). Alcohol and recall: State dependent effects in man. *Science, 163,* 1358-1360.

Greene, S. L., Devlin, D. J., Cannata, P. E., & Gomez, L. M. (1990). No IFs, ANDs, or ORs: A study of database querying. *International Journal of Man-Machine Studies, 32,* 303-326.

Guiliano, , V. W. (1963). Analog networks for word association. *IEEE Transactions on Military Electronics, 7,* 221-234.

Halasz, F. G., Moran, T. P., & Trigg, T. H. (1987). NoteCards in a Nutshell. *Proceedings of the ACM Conference on Human Factors in Computing Systems,* Toronto, Canada, April, 1987.

Herot, C. F. (1980) Spatial management of data. *ACM Transactions on Database Systems. 5*(4), 493-514.

Hinton, G. E., & Anderson, J. A. (Eds.) (1981). *Parallel models of associative memory.* Hillsdale, NJ: Erlbaum.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411-428.

Huberman, B. A., & Hogg, T. (1987). Phase transitions in artificial intelligence systems, *Artificial Intelligence, 33,* 155-171.

Hyde, T. S., & Jenkins, J. J. (1973). Recall for words as a function of semantic, graphic and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior, 12,* 471-480.

Jones, W. P. (1986). On the applied use of computer models of memory: The Memory Extender personal filing system. *International Journal of Man-Machine Studies,* 25, 191-228.

Jones, W. P., & Furnas, G. W. (1987). Pictures of relevance-- A geometric analysis of similarity measures, *Journal of the American Society for Information Science, 38,*(6), 420-442.

Kintsch, W. (1968). Recognition and free recall of organized lists. *Journal of Experimental Psychology, 78,* 481-487.

Kintsch, W. (1989). The role of knowledge in discourse comprehension: A Construction-Integration model. *Psychological Review, 95,* 163-182.

Kolodner, J. L. (1983). Reconstructive memory: A computer model. *Cognitive Science, 7,* 281-328.

Kolodner, J. L. (1984). *Retrieval and organizational strategies in conceptual memory.* Hillsdale, NJ: Erlbaum.

Landauer, T. K., Galotti, K., & Hartwell, S. (1983). Natural command names and initial learning: A study of text editing terms. *Communications of the ACM, 26*,(7), 495-503.

Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science, 10*, 477-493.

Lewis, C. H., & Anderson, J. R. (1976). Interference with real world knowledge. *Cognitive Psychology 8*, 311-355.

Lorch, R. F. (1982). Priming and search processes in semantic memory: A test of three models of spreading activation. *Journal of Verbal Learning and Verbal Behavior, 21*, 469-492.

Luria, A. R. (1968). *The mind of a mnemonist.* Cambridge, MA, Harvard University Press.

Mannes, S., & Kintsch, W. (in press). Routine computing tasks: Planning as understanding. *Cognitive Science.*

Mantei, M. M. (1982). A study of disorientation behavior in ZOG. Unpublished doctoral dissertation, Univ of Southern California.

McCracken, D. L., & Akscyn, R. M. (1984). Experience with the ZOG human-computer interface system. *International Journal of Man-Machine Studies, 21*, 293-310.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive model of context effects in letter perception, part I. An account of basic findings. *Psychological Review 88*, 375-407.

McKoon, G., Ratcliff, R. & Dell, G. S. (1986). A critical evaluation of the semantic - episodic distinction. *Journal of Experimental Psychology: Learning, Memory and Cognition. 12*,(2), 295-306.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Mozer, M. C. (1984). Inductive information retrieval using parallel distributed computation, (ICS Report 8406). San Diego: University of California, Institute for Cognitive Science,.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*, 609-626.

Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review, 90*, 316-338.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General, 106*, 226-254.

Neisser, U., & Winograd, E. (1988). *Remembering reconsidered: Ecological and traditional approaches to the study of memory.* Cambridge: Cambridge University press.

Nielsen, J. (1989) The matters that really matter for hypertext usability. *Proceedings of Hypertext 1989*. Pittsburgh, Pennsylvania, ACM press.

Norman, D. A., & Bobrow D. G. (1979). Descriptions: an intermediate stage in memory retrieval. *Cognitive Psychology, 11*, 107-123.

Ossorio, P. G. (1963). Classification space: A multivariate procedure for automatic document indexing and retrieval. *Multivariate Behavioral Research*, October, 479-524.

Patel-Schneider, P. F., Brachman, R. J., & Levesque, H. J. (1984). Argon: Knowledge representation meets information retrieval. Fairchild Technical Report 654, Schlumberger Palo Alto Research, September.

Peterson, L. R., & Peterson, M. J. (1959). Short term retention of individual items. *Journal of Experimental Psychology, 58*, 193-198.

Quillian, M. R. (1969). The teachable language comprehender. *Communications of the ACM 12*, 459-476.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93-134.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming memory. *Psychological Review. 95*, 385-408.

Ratcliff, R., & McKoon, G. (1989). Memory models, text processing, and cue-dependent retrieval. In (H. L. Roediger & F.. I. M. Craick, Eds.), *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*. Hillsdale, NJ: Erlbaum.

Reder, L. M., & Anderson, J. R. (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology, 12*, 447-472.

Rose, D. E., & Belew, R. K. (1989). Legal information retrieval: A hybrid approach. *Proceedings of the International Conference on AI & Law*, Vancouver, BC.

Rubin, D. C. (1988). Learning poetic language. In F. S. Kessell (Ed.), *The development of language and language researchers: Essays in honor of Roger Brown*. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & Norman, D. A. (1988). Representation in Memory. In R. C. Atkinson, et al. (Eds.) *Stevens' Handbook of Experimental Psychology*, Volume 2 (second edition), 511-587. New York; Wiley.

Salton, G. (1968). *Automatic Information Organization and Retrieval*. New York, NY: McGraw-Hill.

Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading, MA: Addison-Wesley.

Salton, G., & Buckley, C. (1988). On the use of spreading activation methods in automatic information retrieval. *Proceedings of the Association for Computing Machinery Special Interest Group on Information Retrieval (SIGIR)*. New York, NY: ACM Inc. 6, 147-160.

Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science. 41*, 4, 288-297.

Salton, G., & McGill, M. J. (1983). *Introduction to Modern Information Retrieval.* McGraw-Hill: New York.

Schank, R. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology.* 3, 4, 552-631.

Schank, R., & Kolodner, J. L. (1970) Retrieving information from an episodic memory. Research Report 159, Department of Computer Science, Yale University, New Haven, CT.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review, 84*, 1-66.

Schneiderman, B., & Morariu, J. (1986). The Interactive Encyclopedia System (TIES). Department of Computer Science, University of Maryland, College Park, MD.

Shiffrin, R. M. (1970). Memory Search. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 375-447.

Shrager, J., Hogg, T., & Huberman, B. A. (1987) Observations of phase transitions in spreading activation networks. *Science, 236*, 1092-1094.

Simon, H. (in press). Cognitive architectures and rational analysis: Comment. In K. Van Lehn (Ed.), *Architectures for Intelligence*. Hillsdale, NJ: Erlbaum.

Thompson, R. H., & Croft, W. B. (1989). Support for browsing in an intelligent retrieval system. *International Journal of Man-Machine Studies, 30*, 639-668.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

Tulving, E. (1983). *Elements of episodic memory*. Oxford, England: Oxford University Press.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory, *Psychological Review, 80*, 352-373.

VanDijk, T. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Voorhees, E. M. (1985). The cluster hypothesis revisited. *Proceedings of the Association for Computing Machinery Special Interest Group on Information Retrieval (SIGIR)*. New York, NY: ACM Inc. 188-196.

Warner, A. J. (1987). Natural language processing. In M. E. Williams (Ed.) *Annual Review of Information Science and Technology, 22*, 79-108. Amsterdam: Elsevier.

Walker, & Kintsch, W. (1985). Automatic and strategic aspects of knowledge retrieval. *Cognitive Science*, 9, 261-283.

Walker, J., Young, E., & Mannes, S. (In press). A case study of using a manual online. *Machine Mediated Learning*.

Williams, M. D. (1978). The process of retrieval from very long term memory. (Technical report no. 7801). San Diego: Center for Human Information Processing, University of California.

Williams, M. D. (1984). What makes RABBIT run? *International Journal of Man-Machine Studies, 21*, 333-352.