Understanding natural climatic variability has gained new urgency as we ponder the direction and magnitude of future global climatic change. As Schindler *et al.* (*24*) have shown, lake surface temperatures are likely to vary as a result of global warming. Evidence of past changes in lake temperature, such as shown here, is important for understanding the response of lakes to future anthropogenic climatic change.

REFERENCES AND NOTES

1. W. A. Watts, in *Studies in the Lateglacial of North-West Europe*, J. J. Lowe, J. M. Gray, J. E. Robinson, Eds. (Pergamon, Oxford, 1980), pp. 1–21.
2. R. J. Mott, *Syllogeus* **55**, 281 (1985).
3. _____, D. R. Grant, R. Stea, S. Occhietti, *Nature* **323**, 247 (1986).
4. R. R. Stea and R. J. Mott, *Boreas* **18**, 169 (1989).
5. H. E. Wright, Jr., *Quat. Sci. Rev.* **8**, 295 (1989).
6. D. M. Peteet, in *Abrupt Climatic Change—Evidence and Implications*, W. H. Berger and L. D. Labeyrie, Eds. (Reidel, Dordrecht, 1987), pp. 185–193.
7. D. M. Peteet *et al.*, *Quat. Res.* **33**, 219 (1990).
8. W. S. Broecker *et al.*, *ibid.* **30**, 1 (1988).
9. D. R. Engstrom, B. C. S. Hansen, H. E. Wright, Jr., *Science* **250**, 1383 (1990).
10. I. R. Walker, J. P. Smol, D. R. Engstrom, H. J. B. Birks, *Can. J. Fish. Aquat. Sci.* **48**, 975 (1991).
11. R. J. Mott, *Can. J. Earth Sci.* **12**, 273 (1975).
12. For midge analysis, 1 to 8 ml of sediment were deflocculated in warm 5% KOH and sieved on a 95-μm mesh. Head capsules retained on the sieve were hand sorted from the sediment at ×50 magnification from a Bogorov counting tray. Relative abundances of midge taxa are based on a minimum of 50 chironomid head capsules. For a more detailed summary of methods, see (*10*).
13. I. R. Walker and R. W. Mathewes, *J. Paleolimnol.* **2**, 61 (1989).
14. I. R. Walker, *Hydrobiologia*, in press.
15. The core chronology is based on ^{14}C dates reported (*3*) for core MS 68-27 (GSC-1643, 9.460 ± 0.220 ka, 5 cm of sediment collected 28 to 33 cm above the Younger Dryas clay; GSC-3862, 10.100 ± 0.130 ka, 5 cm of sediment collected immediately above the Younger Dryas clay; GSC-1645, 11.300 ± 0.180 ka, 5 cm of sediment collected immediately below the Younger Dryas clay; and GSC-1067, 12.600 ± 0.270 ka, 5 cm of gyttja deposited immediately above the basal clay), and stratigraphic correlation of lithologies between cores. Comparable deposits, distributed throughout Atlantic Canada, yield similar dates (*3*).
16. The weighted-averaging transfer function was developed from a temperature-constrained canonical correspondence analysis of midge distributions among 26 Labrador lakes. We implemented this analysis using the computer program CANOCO, Version 3.0. The Labrador lakes were distributed among arctic tundra, forest-tundra, and forest sites, with summer surface-water temperatures ranging from 7.6° to 18°C. For a more detailed summary of the temperature inference procedure, see (*10*).
17. J. M. Line and H. J. B. Birks, *J. Paleolimnol.* **3**, 170 (1990).
18. C. J. F. ter Braak and H. van Dam, *Hydrobiologia* **178**, 209 (1989).
19. T. C. Atkinson *et al.*, *Nature* **325**, 587 (1987).
20. W. S. Broecker *et al.*, *ibid.* **315**, 21 (1985).
21. V. N. Rampton, R. C. Gauthier, J. Thibault, A. A. Seaman, *Geol. Surv. Can. Mem.* **416**, 1 (1984).
22. G. R. Coope and W. Pennington, *Philos. Trans. R. Soc. London Ser. B* **280**, 337 (1977).
23. B. Ammann, *Eclogae Geol. Helv.* **82**, 183 (1989).
24. D. W. Schindler *et al.*, *Science* **250**, 967 (1990).

25. This work was funded by the Natural Sciences and Engineering Research Council of Canada, the Geological Survey of Canada, and the Atmospheric Environment Service of Environment Canada. Geological Survey of Canada Contribution 44290.

20 February 1991; accepted 4 June 1991

Global Text Matching for Information Retrieval

GERARD SALTON* AND CHRIS BUCKLEY

An approach is outlined for the retrieval of natural language texts in response to available search requests and for the recognition of content similarities between text excerpts. The proposed retrieval process is based on flexible text matching procedures carried out in a number of different text environments and is applicable to large text collections covering unrestricted subject matter. For unrestricted text environments this system appears to outperform other currently available methods.

THE PROBLEM OF TEXT RETRIEVAL from large heterogeneous text databases, in which the vocabulary varies widely and the subject matter is unrestricted, becomes increasingly important every year. These databases include newspaper articles, newswire dispatches, textbooks, dictionaries and encyclopedias, manuals, magazine articles, and so on. The normal text analysis and text indexing approaches that are based on the use of available thesauruses and other vocabulary control devices are difficult to apply in unrestricted text environments, because the word meanings are not stable in such circumstances and the interpretation varies depending on context. The applicability of more complex text analysis systems that are based on the construction of knowledge bases covering the detailed structure of particular subject areas, together with inference rules designed to derive relationships between the relevant concepts, is even more questionable in such cases. Complete theories of knowledge representation do not exist, and it is unclear what concepts, concept relationships, and inference rules may be needed to understand particular texts (*1*).

We take advice from Wittgenstein and others who suggest that text understanding must be based on a study of how text words are used in the language (*2*). In so doing, we are not primarily interested in deriving detailed descriptions of text content, but instead we want to recognize text portions within which the text meanings are homogeneous. An identification of semantically homogeneous text excerpts makes it possible to supply text links between related text portions, leading to the generation of structured text representations that lend themselves to selective reading and perusal of large texts. In addition, the recognition of related text portions leads to the retrieval of relevant texts in answer to available search requests.

In the identification of related text portions in unrestricted text environments, the texts themselves must necessarily form the basis for the text analysis. The following text comparison process is used. Each text in a collection is broken down into individual text units—for example, sections, paragraphs, and sentences. A standard automatic indexing procedure is utilized to assign to each text unit (or each available search request) a set of weighted content identifiers, or terms, collectively used to represent text content. If t terms in all are available, a given text item $\mathbf{D}_i$ may then be represented by a term vector as $\mathbf{D}_i = (w_{i1}, w_{i2} \ldots, w_{it})$, where w_{ik} is the weight of term T_k assigned to document $\mathbf{D}_i$. A weight of zero is used for terms that are absent from a particular document and a positive weight for terms actually assigned. Similarities between particular text items (or between text items and information requests) are detected by comparison of the term vectors representing the texts at various levels of detail. For example, two texts may be assumed to be related when a sufficient global similarity between them is accompanied by local similarities between included text paragraphs or text sentences.

The similarity measurement between two texts must depend on the types and weights of the coinciding terms in the respective term vectors. In retrieval, the most valuable terms for document content representation are those best able to distinguish particular texts from the remainder of the collection. Thus, the term weighting system should assign low weights to high-frequency terms that occur in many documents of a collection and high weights to terms that are important in particular documents but unimportant in the remainder of the collection. The weight of terms that occur rarely in a collection is of no consequence, because such terms contribute little to the text similarity.

A well-known term weighting system following that prescription assigns weight w_{ik} to term T_k in document $\mathbf{D}_i$ in proportion to the frequency of occurrence of a term in $\mathbf{D}_i$ and in inverse proportion to the number of documents to which the term is assigned (*3*). When the texts are represented by term vectors, the similarity (*sim*) between two items $\mathbf{D}_i$ and $\mathbf{D}_j$ is conveniently obtained as the inner product between corresponding vectors, or $sim(\mathbf{D}_i, \mathbf{D}_j) = \sum_{k=1}^{t} w_{ik}\, w_{jk}$. Thus, the similarity between two texts depends on the weights of coinciding terms

Department of Computer Science, Cornell University, Ithaca, NY 14853.

*To whom correspondence should be addressed.

in the two term vectors.

In practice the document length, and hence the number of non-zero term weights assigned to a document, varies widely. The same is true of the values of the occurrence frequencies, f_{ik}, of terms T_k in $\mathbf{D}_i$. To give each item an equal chance of being retrieved, it is convenient to use an enhanced term weighting formula by normalizing for document length. A high-quality term weight for terms assigned to paragraphs or larger size texts is

$$w_{ik} = \frac{\left(0.5 + 0.5\frac{f_{ik}}{\max f_{ip}}\right)\left(\log\frac{N}{n_p}\right)}{\sqrt{\sum_{k=1}^{t}\left(0.5 + 0.5\frac{f_{ik}}{\max f_{ip}}\right)^2\left(\log\frac{N}{n_p}\right)^2}} \quad (1)$$

where w_{ik} is the weight of term T_k in document $\mathbf{D}_i$, f_{ik} is the occurrence frequency of T_k in $\mathbf{D}_i$, N represents the number of documents in the collection, n_k is the number of documents with term T_k, and the maximization operation is extended over all terms in a given vector. The weights of Eq. 1 range from 0 to 1 and include an enhanced term frequency factor ($0.5 + 0.5 f_{ik}/\max f_{ip}$) that varies only between 0.5 for terms with zero frequency and 1 for the most frequent terms. In addition, an inverse collection frequency factor, $\log(N/n_k)$, is used, which is large for terms that occur rarely in the collection and small for terms assigned to many documents in a collection. When the terms are weighted in accordance with Eq. 1, the values of the corresponding inner product similarity function between text items depends on the proportion of matching terms in the two vectors (*1*).

For short text excerpts, such as text sentences, a text similarity measure that depends on the proportion of matching terms is not indicative of coincidence in text meaning. For example, fragments such as "consider figure x" or "equation y is an example" may be repeated many times in ordinary texts, but the vocabulary coincidences in such fragments are not meaningful. Short texts are therefore compared by means of an unnormalized term weight, equivalent to the numerator of Eq. 1 without the length normalization of the denominator. The unnormalized term weight ranges in size from zero to log N, and the corresponding inner product text similarity depends on the number (rather than the proportion) of matching terms. The experimental output included in this report was generated by means of the normalized similarity measure of Eq. 1 for comparisons involving full documents or document paragraphs and unnormalized measures for sentence comparisons.

Consider a brief example of the text linking system for the companion article "Developments in automatic text retrieval," obtained by text paragraph and text sentence linking (*1*). The paragraphs in that article are numbered from 1 to 45, and three paragraphs (numbers 5, 32, and 43) are used in Table 1 as initial query texts to be compared with all other text paragraphs for text linking purposes. The topic "knowledge-based retrieval" is reasonably self-contained in the original article, and Table 1 shows that paragraph 32 is linked only to adjacent text paragraphs (these paragraphs exceed inner product similarity threshold 0.120). When paragraphs 5 and 43 are used for text searching, several links again exist to adjacent paragraphs. However, paragraph 5 on "Boolean retrieval" is also linked to a large section dealing with file technologies (paragraphs 38 to 41), and paragraph 43 is

Table 1. Paragraph linking in sample text. Numbered paragraphs are taken from the sample text, which is the companion article (*1*).

Query paragraph	Retrieved items (>0.120)	Sentence links
¶5 (second under **Conventional Retrieval Methods**) "Boolean retrieval"	¶4, 6 adjacent	10 to ¶4 5 to ¶6
	¶38, 39, 40, 41 (first four ¶s under **Retrieval Strategies**) "File technologies"	3 to ¶38 4 to ¶39 5 to ¶40 2 to ¶41
¶32 (sixth ¶ under **Linguistics- and Knowledge-Based Approaches**) "Knowledge-based retrieval"	¶33, 34, 35, 36 (All adjacent ¶'s)	Multiple
¶43 (third to last ¶ under **Retrieval Strategies**) "On-line search, relevance feedback"	¶38 (first ¶ under **Retrieval Strategies**)	3 to ¶38 6 to ¶41
	¶41, 44 adjacent	4 to ¶44
	¶10 (third ¶ under **Alternative Retrieval**) "Cosine similarity"	None to ¶10 (short paragraph)

Table 2. Multistage and single-stage encyclopedia search [global document comparison: (R) relevant, (N) nonrelevant].

Initial query	Retrieved items stage one (threshold 0.20)	Retrieved items stage two (threshold 0.25)	Retrieved items stage three (threshold 0.30)
18 Abandonment	7041 Desertion (R) (0.3651)	11690 Husband and Wife (R) (0.3137)	5994 Conjugal Rights (N) (0.3463)
	7024 Derelict (R) (0.3646)		
	20688 Separation (R) (0.3298)	17527 Parent and Child (R) (0.2559)	
	20122 Salvage (R) (0.2210)		
	4517 Caroline of Brunswick (N) (0.1652)		
	5994 Conjugal Rights (N) (0.1581)		
	5886 Community Property (N) (0.1552)		

Table 3. Retrieval evaluation for electronic mail messages (1984 messages, 180 queries; global text comparison).

Recall	Average precision	
	Messages with quotations	Messages without quotations
0.1	0.9480	0.6724
0.3	0.9279	0.6402
0.5	0.9016	0.5806
0.7	0.8097	0.4348
0.9	0.7188	0.3644

linked to paragraph 10, dealing with the cosine similarity. Although the global similarity between paragraphs 43 and 10 exceeds the threshold of 0.120 (in part because paragraph 10 is quite short), no sentence links are found for these two paragraphs. In a practical situation, paragraphs 10 and 43 may therefore be considered to be unrelated.

A typical search conducted in an automated encyclopedia is illustrated in Table 2. In this case a specific encyclopedia article is used as a search request, and an ordered list directing the user to additional encyclopedia articles in the same subject area is wanted (*4*). Article 18 (Abandonment) is used as a query in Table 2. The initial search produces an ordered list of articles in decreasing order of the inner product similarity. A similarity threshold of 0.20 is used initially, and the newly retrieved items above the threshold are used in turn as new search queries to retrieve additional articles. The output of Table 2 shows three search stages, the retrieval threshold being increased from 0.20 in stage one to 0.25 and 0.30 in subsequent stages. Query 7041 (Desertion) produces one additional item (11690, Husband and Wife), and query 20688 (Separation) produces article 17527 (Parent and Child). Finally, query 11690 (Husband and Wife) retrieves one more item with a similarity greater than 0.30 (5994, Conjugal Rights).

The text retrieval operations can in principle be evaluated by experts to assess the relevance or nonrelevance of each retrieved item with respect to the corresponding query, followed by the computation of formal recall and precision measures (*5*). In practice, the use of objective relevance assessments between queries and stored documents is preferable. For the encyclopedia searches, the published cross-references between articles can be used as indicators of relevance. For the sample search of Table 2, six of the seven articles retrieved by the multistage search are properly referenced in article 18; hence, the retrieval precision is 6/7 (0.86). A seventh relevant article (7287 Divorce) is not retrieved by the global text comparison because of its extreme length. (Long articles must be subdivided into sections to obtain proper retrieval performance when normalized term weight similarities are used.) The search recall for the sample search is then also 6/7.

The multistage search of Table 2 may be compared with a single-stage search in which all articles are retrieved by a single search that uses the initial query only. The second column of Table 2 shows that when seven articles are retrieved in stage one, the three documents obtained in addition to the original four all have a low similarity with the query, and all of them are nonrelevant. The precision and recall of the single level search is therefore 4/7 = 0.57.

Table 3 contains an evaluation of retrieval operations conducted with a collection of electronic news messages. Specifically, 180 news messages were used as queries and processed against a message collection of 1984 messages. Electronic messages are characterized by the fact that reply messages often contain quotations from the original query message. Furthermore, message headers are used that contain a subject line for each message and, optionally, cross-references to other messages. For evaluation purposes, message B may be termed relevant to query message A whenever A and B contain an identical subject description or when B contains a reference to A.

The evaluation output of Table 3 shows nearly perfect results for the message collection that includes the quoted portions. When the recall is 0.50 (after retrieval of 50% of the relevant items for each of the 180 queries), the precision is 90% (90% of the retrieved items are relevant). When the messages are processed without quotations, nearly 60% of the retrieved items are relevant at a recall level of 0.50 (*6*). The output of Table 3 appears especially favorable when one considers that the objective relevance assessments actually used are substantially flawed. In practice, reply messages often contain subject descriptions that are slightly different from those used in the original primary messages; these differences in the subject descriptions lead to a false assessment of nonrelevance for the reply messages, even though their early retrieval is perfectly in order. The result is a depressed search precision. The recall is similarly affected unfavorably by chains of messages in which message B replies to A, C replies to B, D replies to C, and so on. In all of these cases, the original subject descriptions may be maintained, even though the actual message content may shift over time. Hence, the rejection of the later replies (such as C and D) in response to query A may be justified even though the system treats all of these messages as relevant in response to A.

The output of Table 3 was obtained by global text matching alone. The performance of the additional sentence comparisons within globally matching texts is illustrated in Table 4. Here a semifixed retrieval threshold is used for each query, consisting of either the top 20 items or all items retrieved up to the rank of the last relevant item for that query, whichever comes earlier. Because many items must be retrieved for each query, even though most queries exhibit one or two relevant documents only, the initial precision (0.1787) is relatively low. The lower part of Table 4 shows that the local text match at the sentence level acts to promote precision. As the number of required sentence matches and matching terms in these sentences increase, fewer documents are retrieved in response to the query messages, but the number of relevant

Table 4. Retrieval evaluation for electronic mail collection with sentence pair comparisons (1984 messages, 180 queries).

Type of comparison	Retrieved items (number)	Retrieved relevant items (number)	Retrieval precision
No sentence matching	2071	370	0.1787
One matching sentence pair required			
At least 2 term matches	1420	333	0.2345
At least 3 term matches	599	217	0.3623
At least 4 term matches	280	128	0.4571
At least 5 term matches	150	75	0.5000
At least 6 term matches	94	51	0.5426
Two matching sentence pairs required			
At least 2 term matches	888	281	0.3164
At least 3 term matches	296	135	0.4561
At least 4 term matches	109	66	0.6055
At least 5 term matches	49	31	0.6327
At least 6 term matches	27	18	0.6667

retrieved items grows sharply.

The text linking and retrieval system described in this report is applicable to text collection in arbitrary subject areas covering texts that vary widely in scope and length. Its operations depend on either the use of discursive English language queries that provide good descriptions of the wanted subjects or the availability of relevant text excerpts that can serve as initial queries. When a sufficient global text similarity exists between the available query vectors and the vectors representing the stored text items and local similarities are detected between certain paragraphs and sentences included in the sample texts, the conclusion follows that the texts are closely related. No other text search and retrieval approach currently contemplated appears to offer equal promise for unrestricted text environments and arbitrary subject matter.

REFERENCES AND NOTES

1. G. Salton, *Science* **253**, 974 (1991).
2. L. Wittgenstein, *Philosophical Investigations* (Basil Blackwell, Oxford, 1953).
3. G. Salton, *Automatic Text Processing—The Transformation, Analysis and Retrieval of Information by Computer* (Addison-Wesley, Reading, MA, 1989); _____ and C. Buckley, *Inf. Process. Manage.* **24**, 513 (1988).
4. The 29-volume Funk and Wagnalls encyclopedia being searched consists of approximately 24,900 articles numbered in alphabetical order from 1 to 24,900. The article length varies from a few words (for cross-references) to many pages of printed text. The authors are grateful to the Microsoft Corporation for making available a machine-readable version of the encyclopedia.
5. Formally, evaluation parameters such as recall and precision are often used, representing the proportion of relevant items retrieved and the proportion of retrieval items that are relevant, respectively. Recall is the number of retrieved and relevant items divided by the total number of relevant items in the collection; precision is the number of retrieved and relevant items divided by the total number retrieved.
6. G. Salton and C. Buckley, "Flexible Text Matching for Information Retrieval," *Technical Report TR 90-1158* (Computer Science Department, Cornell University, Ithaca, NY, 1990).

7. This study was supported in part by the NSF under grant IRI 89-15847.

13 March 1991; accepted 5 June 1991

Wright's Shifting Balance Theory: An Experimental Study

MICHAEL J. WADE AND CHARLES J. GOODNIGHT

Experimental confirmation of Wright's shifting balance theory of evolution, one of the most comprehensive theories of adaptive evolution, is presented. The theory is regarded by many as a cornerstone of modern evolutionary thought, but there has been little direct empirical evidence supporting it. Some of its underlying assumptions are viewed as contradictory, and the existence and efficacy of the theory's fundamental adaptive process, interdemic selection, is the focus of controversy. Interdemic selection was imposed on large arrays of laboratory populations of the flour beetle *Tribolium castaneum* in the manner described by Wright: the differential dispersion of individuals from demes of high fitness into demes of low fitness. A significant increase in average fitness was observed in the experimental arrays when compared to control populations with equivalent but random migration rates. The response was not proportional to the selection differential: The largest response occurred with interdemic selection every two generations rather than every generation or every three generations. The results indicate that the interdemic phase of Wright's shifting balance theory can increase average fitness and suggest that gene interactions are involved in the observed response.

WE REPORT THE RESULTS OF A 4-year experimental investigation of Wright's shifting balance theory of adaptive evolution (*1–4*). Wright's theory, proposed 60 years ago (*1, 3*), is one of the most widely known and comprehensive theories of adaptative evolution (*5*), regarded by many as "a cornerstone of modern evolutionary thought" (*6*, p. 265) and as "the dominant theory of evolution in the 20th century" (7, p. 625). However, there are several aspects of the theory that to date "have never been analyzed in detail" (*8*) and are unsupported by empirical data (*9*), either at the population or molecular level. Recent mathematical investigations of Wright's theory (*10*) illustrate that, under certain conditions, interdemic selection through differential dispersion, the central adaptive mechanism of Wright's theory, can cause genetic change. However, it is not known whether the conditions necessary in the model are met by populations in nature. Our experiments with the flour beetle *Tribolium castaneum* demonstrate the efficacy of interdemic selection by differential dispersion for causing genetic change.

M. J. Wade, Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637.
C. J. Goodnight, Department of Zoology, University of Vermont, Burlington, VT 05405.

Wright (*3*) identified three phases important to his theory, all of which are acting simultaneously: (i) random genetic drift when "the set of gene frequencies drifts at random in a multidimensional stochastic distribution about the equilibrium set characteristic of a particular fitness peak" (*3*, p. 455); (ii) mass selection when the set of gene frequencies drifts far enough within one deme to cross over into the domain of attraction of a different adaptive peak—that is, "There ensues a period of relatively rapid change in this deme, dominated by selection among individuals (or families) . . ." (*3*, p. 455); and (iii) interdemic selection when a deme ". . . by excess dispersion, systematically shifts the position of equilibrium [of other demes] toward its own position" (*3*, p. 455).

Some phases of the shifting balance process appear to be in conflict with one another. In order for phase (i) to operate efficiently, small numbers of breeding adults and little migration are required, but these conditions make phase (ii), mass selection, inefficient unless selection is much stronger than random genetic drift. We do not know how much heritable variation among demes in local mean fitness can result from the combined action of random genetic drift and mass selection: "The relative importance of natural selection and random genetic drift . . . remains the most important unsolved problem in our understanding of the mechanisms that bring about biological evolution" (*11*, p. 164). We know little of the existence or the density of "local adaptive peaks" or the evolutionary role of epistasis for fitness (*8, 12–13*), but recent theory (*14*) indicates its potential importance. Little is known about the rate of origination of the genetic and phenotypic variation in average fitness among demes which are necessary for operation of the third phase: Just as individual (mass) selection requires genetic variation among individuals, interdemic selection requires genetic variation among demes (*15*). Last, the export of gene combinations in phase (iii) may require the dispersion of large numbers of adults from one deme to another and interfere with the first phase. In summary, quantitative information from the laboratory or field is lacking for each of the three phases.

Local breeding numbers, N_e, and migration among demes, m, affect the rate of genetic differentiation of demes for fitness, and small amounts of differential migration