# DEFENSIVE PUBLICATION 

## UNITED STATES PATENT OFFICE

Published at the request of the applicant or owner in accordance with the Notice of Dec. 16, 1969, 869 0.G. 687. The abstracts of Defensive Publication applications are identified by distinctly numbered series and are arranged chronologically. The heading of each abstract indicates the number of pages of specification, including claims and sheets of drawings contained in the application as originally filed. The files of these applications are available to the public for inspection and reproduction may be purchased for 30 cents a sheet.

Defensive Publication applications have not been examined as to the merits of alleged invention. The Patent Office makes no assertion as to the novelty of the disclosed subject matter.

## PUBLISHED JULY 18, 1972

900 O.G. 817

T900,006
INFORMATION STORAGE AND RETRIEVAL SYSTEM AND METHOD
Matthews P. Perriens, Rockville, Md., and John H. Williams, Jr., Annandale, Va., assignors to International Business Machines Corporation, Armonk, N.Y. Continuation of application Ser. No. 736,837, June 13, 1968. This application Apr. 19, 1971, Ser. No. 135,467 Int. Cl. G06f $1 / 00,7 / 00,15 / 00$
U.S. Cl. 340-172.5

3 Sheets Drawing. 21 Pages Specification


An information storage and retrieval system and method capable of handling queries in sentence form and presenting responses either as identification numbers of potentially pertinent documents or as documents in full text form. Stored documents containing so-called search symbols are known to the system. A probability calculator determines the likelihood of the occurrence of a search symbol at least once in a given stored document in the system's data base. The system and method causes the query text to be scanned so as to determine which search symbols are contained therein. The term overlap is used to designate a search symbol in the query and also in a given stored document. The particular document in the system's data base having the smallest joint probability of occurrence of overlap search symbols is designated as having the highest relevance potential within the data base to a given query. The stored document having the next larger joint probability of occurrence of overlap search symbols, has the next highest relevance potential. In this manner, any select number of relevant stored documents may be outputted by the system and method, in the order of relevance, as either identification numbers of potentially pertinent documents or as the documents per se in full text form.

July 18, 1972
INPORMATION STORAGE AND RETRIEVAL SYSTEM AND METHOD


INFORMATION STORAGE AND RETRIEVAL SYSTEM AND METHOD

Original Filed June 13, 1968


CONCORDANCE

| SEARCH SYMBOLS | DOCUMENT | NUMBERS |  |  |  |
| :---: | :---: | :---: | :---: | :--- | :--- |
| $S S_{1}$ | $D O C_{1}$ | $D O C_{17}$ | $D O C_{23}$ | $\cdots$ | $D O C_{m}$ |
| $S S_{2}$ | $D O C_{3}$ | $D O C_{12}$ | $D O C_{17}$ | $\cdots$ | $D O C_{m-x}$ |
| $S S_{3}$ | $D O C_{3}$ | $D O C_{23}$ | $D O C_{50}$ | $\cdots$ | $D O C_{m-x}$ |
| $S S_{4}$ | $D O C_{2}$ | $D O C_{5}$ |  |  |  |
| $\vdots$ | $\vdots$ |  |  |  |  |
| $S S_{n}$ | $D O C_{7}$ | $D O C_{18}$ | $D O C_{19}$ | $D O C_{20}$ | $D O C_{m-x}$ |

DOCUMENTS

FIG. 4

| $\begin{aligned} & \text { s } \\ & \text { 翤 } \\ & \text { in } \end{aligned}$ | SS 1 | DOC | $\mathrm{DOC}_{2}$ | $D 0 C_{5}$ | ${ }^{\text {DOC }} 6$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | $S_{2}$ | $\mathrm{DOC}_{4}$ | ${ }^{D} 0 C_{6}$ |  |  |
|  | $S_{3}$ | DOC1 | $\mathrm{DOC}_{3}$ | $\mathrm{DOC}_{5}$ |  |
| $\begin{aligned} & \text { I } \\ & \text { w } \\ & \text { 岕 } \end{aligned}$ | $S_{4}$ | $\mathrm{DOC}_{2}$ |  |  |  |
|  | $S_{5}$ | $\mathrm{DOC}_{5}$ | $\mathrm{DOC}_{8}$ |  |  |
|  | SS 6 | $\mathrm{DOC}_{4}$ | ${ }^{D} 0 C_{6}$ |  |  |

INFORMAIION STORAGE AND RETRIEVAL SYSTEM AND METHOD
FIG.5A

|  | $S S_{1}$ | $S S_{2}$ | $S S_{3}$ | $S S_{4}$ | $S S_{5}$ | $S S_{6}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $D O C_{1}$ | $x$ |  |  |  |  |  |
| $D O C_{2}$ | $x$ |  |  |  |  |  |
| $D O C_{5}$ | $x$ |  |  |  |  |  |
| $D O C_{6}$ | $x$ |  |  |  |  |  |
| ITERATION \#I |  |  |  |  |  |  |

FIG.5B

|  | $S S_{1}$ | $S S_{2}$ | $S S_{3}$ | $S S_{4}$ | $S S_{5}$ | $S S_{6}$ |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| $D O C_{1}$ | $x$ |  |  |  |  |  |
| $D O C_{2}$ | $x$ |  |  |  |  |  |
| $D O C_{5}$ | $x$ |  |  |  |  |  |
| $D O C_{6}$ | $x$ | $x$ |  |  |  |  |
| $D O C_{4}$ |  | $x$ |  |  |  |  |
| ITERATION \#2 |  |  |  |  |  |  |

FIG.5C
SEARCH SYMBOLS

|  | $S S_{1}$ | $S S_{2}$ | $S S_{3}$ | $S S_{4}$ | $S S_{5}$ | $S S_{6}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $D O C_{1}$ | $x$ |  | $x$ |  |  |  |
| $D O C_{2}$ | $x$ |  |  | $x$ |  |  |
| $\sum_{8}^{=}$ | $D O C_{5}$ | $x$ |  | $x$ |  | $x$ |
|  | $D O C_{6}$ | $x$ | $x$ |  |  |  |
| $D O C_{4}$ |  | $x$ |  |  | $x$ |  |
| $D O C_{3}$ |  |  | $x$ |  |  |  |
| $D O C_{8}$ |  |  |  |  | $x$ |  |

FIG. 6

| DOCUMENT <br> NUMBER | JOINT PROBABILITY |
| :--- | :--- |
| $D O C_{1}$ | $P\left(S S_{1}\right) \times P\left(S S_{3}\right)=P_{1}$ |
| $D O C_{2}$ | $P\left(S S_{1}\right) \times P\left(S S_{4}\right)=P_{2}$ |
| $D O C_{5}$ | $P\left(S S_{1}\right) \times P\left(S S_{3}\right) \times P\left(S S_{5}\right)=P_{5}$ |
| $D O C_{6}$ | $P\left(S S_{1}\right) \times P\left(S S_{2}\right) \times P\left(S S_{6}\right)=P_{6}$ |
| $D O C_{4}$ | $P\left(S S_{2}\right) \times P\left(S S_{6}\right)=P_{4}$ |
| $D O C_{3}$ | $P\left(S S_{3}\right)=P_{3}$ |
| $D O C_{8}$ | $P\left(S S_{5}\right)=P_{8}$ |

FIG. 7

| DOCUMENT <br> NUMBER | JOINT PROBABILITY |
| :--- | :---: |
| $D O C_{5}$ | $P_{5}$ |
| $D O C_{6}$ | $P_{6}$ |
| $D O C_{4}$ | $P_{4}$ |
| $D O C_{2}$ | $P_{2}$ |
| $D O C_{1}$ | $P_{1}$ |
| $D O C_{3}$ | $P_{3}$ |
| $D O C_{8}$ | $P_{8}$ |
| GIVEN |  |
| $P_{5}<P_{6}<P_{4}<P_{2}<P_{1}<P_{3}<P_{8}$ |  |

