Abstract: A method for identifying clusters of similar documents from among a set of documents is described. A particular document is selected from among available documents of the set of documents, and a probe is generated based on the particular document. The probe comprises one or more features. Documents are found that satisfy a similarity condition using the probe from among the available documents. Some or all of the documents that satisfy the similarity condition are associated with a particular cluster of documents. The process can be repeated to generate further clusters. The method can be implemented with a computer, and associated programming instructions can be contained within a computer readable carrier.
Published:
— with international search report
A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 15/00 (2007.01)
USPC - 715/514
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
715/514

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 715/500; 707/1; 3, 5 (view text search terms below)

Electronic database consulted during the international search (name of database and, where practicable, search terms used)
PubWEST(PGPB,USPT,EPAB,JPAP; PLUR-YES; DialogPRO(Engineering); Google Scholar; Text search terms: identifying clusters similar documents, probe generated, satisfy similarity condition, among available documents, cluster documents

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category*</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>US 2003/0167163 A1 (GLOVER et al.) 04 September 2003 (04.09.2003) entire document, especially Abstract, FIG.7 and para [0015]-[0017], [0036]-[0038], [0042], [0044], [0051], [0061]-[0062], [0085] and [0087]</td>
<td>1-11</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C.

Date of the actual completion of the international search
11 August 2007 (11.08.2007)

Date of mailing of the international search report
15 FEB 2008

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-3201

Authorized officer:
Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)