Abstract: A document search and retrieval system and method stores documents in groups based on content. The documents are self-organized into a hierarchy of conceptual clusters, and branches of the hierarchy are stored separately in distinct physical stores, each having an index. In response to a query, the system finds the concepts (clusters) that best match the search criteria and returns the documents from those content categories. The indexing, clustering, and searching are performed using document themes and/or summaries. Themes are automatically developed by stemming and scoring phrases from the sentences in each document, and clustering the sentences containing the highest-scoring stems. A set of phrases (themes) is taken from each cluster. Document summaries are taken from text segments for each cluster of sentences within a document, then strung together to create a summary.
INTERNATIONAL SEARCH REPORT

International application No
PCT/US06/11931

A CLASSIFICATION OF SUBJECT MATTER

IPC G06F 7/00 (2006 G1), 7/1330 (2006 01)

USPC 707/4

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)

US PLUS 2 1 1 715/53 1, 707/5, 707/3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Calishain, "Google Hacks 100 Industrial-Strength Tips & Tools", O'Reilly, 28 Feb 2003, First Ed

Electronic database consulted during the international search (name of data base and, where practicable, search terms used)

Please See Continuation Sheet

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>Calishain et al., &quot;Google Hacks 100 Industrial-Strength Tips &amp; Tools&quot;, O'Reilly, 28 Feb 2003, First Ed</td>
<td>1-5, 8-23, 33-36</td>
</tr>
<tr>
<td>Y</td>
<td>US 2001/0056350 A1 (Calderone et al) 27 Dec 2001 (27 12 2001), Paragraphs [0212, 0056], Fig. 6</td>
<td>6</td>
</tr>
<tr>
<td>Y</td>
<td>US 2002/018861 1A1 (Smalley et al) 12 Dec 2002 (12 12 2002), Paragraphs [0033, 0013], Fig. 11, items 70, 74, 78</td>
<td>7</td>
</tr>
<tr>
<td>Y</td>
<td>US 4,474,454 A (Sigmra) 2 Oct 1984 (02 10 1984), Col. 16, Lines 5-20, Col. 2, Lines 1-5</td>
<td>26</td>
</tr>
</tbody>
</table>

X Further documents are listed in the continuation of Box C

T See patent family annex

* Special categories of cited documents

A' document defining the general state of the art which is not considered to be of particular relevance

E' earlier application or patent published on or after the international filing date

L' document which may throw doubts on prior art claims or which is cited to establish the publication date of another citation or other special reason (as specified)

O' document referring to an oral disclosure, use, exhibition or other means

P' document published prior to the international filing date but later than the priority date claimed

X' document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

Y' document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y' document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

A' document member of the same patent family

Date of the actual completion of the international search 10 May 2007 (10 05 2007)

Date of mailing of the international search report 01 Aug 2007

Name and mailing address of the ISA/US

Autho KasNOficer

Joseph D Wong

Telephone No (571) 272-3608

Facsimile No (571) 273-3201

Form PCT/ISA/210 (second sheet) (April 2005)
<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>Y</td>
<td>US 6,741,959 B1 (Kaiser) 25 May 2004 (25.05.2004), Col. 10, Lines 1-10; abstract</td>
<td>28</td>
</tr>
<tr>
<td>Y</td>
<td>US 5,740,456 (Harei et al) 14 Apr 1998 (14.04.1998), Col. 4, Lines 45-55; Col. 5, Lines 43-55; Col. 9, Lines 66-67; Col. 10, Lines 1-15, Fig. 3, item 120, abstract.</td>
<td>29</td>
</tr>
</tbody>
</table>
Continuation of B. FIELDS SEARCHED Item 3:
Dialog EEDD, WIPO, Cirilab Instigator web site, ACM Digital Library, IEEE Explore, DTIC, ResearchDisclosure.com, CiteSeer, EBSCO Computer Science Index, Inventor Name, Assignee, ESP@CENET, IBM Research Reports, HP Labs Research Reports, amazon.com