

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
24 July 2003 (24.07.2003)

PCT

(10) International Publication Number
WO 2003/060767 A3

(51) International Patent Classification⁷: **G06F 17/30**

(21) International Application Number:
PCT/US2002/041144

(22) International Filing Date:
20 December 2002 (20.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/342,956 21 December 2001 (21.12.2001) US
10/171,170 13 June 2002 (13.06.2002) US

(71) Applicant and

(72) Inventor: **DOZIER, Christopher, C.** [US/US]; 4880
Knottingham Circle, Eagan, MN 55122 (US).

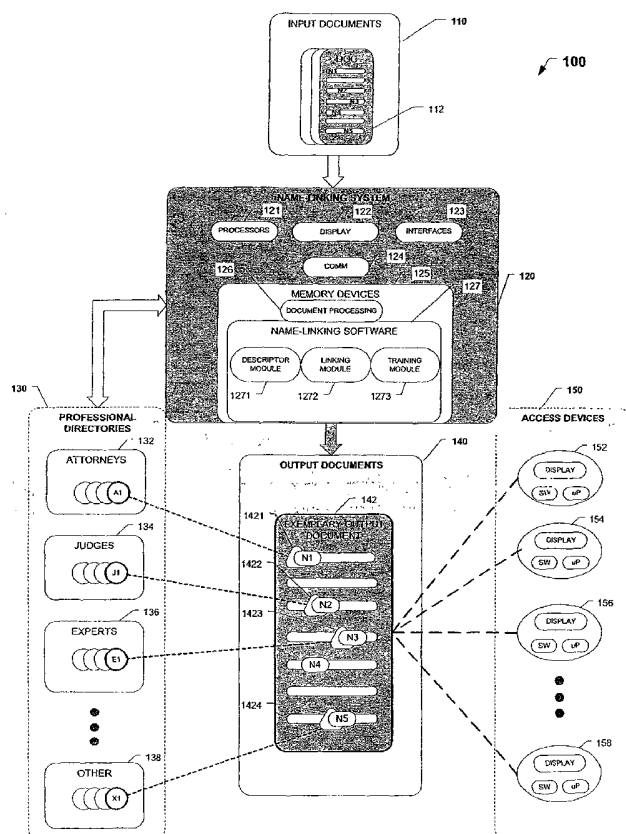
(74) Agents: **STEFFEY, Charles, E.** et al.; Schwegman, Lund-
berg, Woessner & Kluth, P.O. Box 2938, Minneapolis, MN
55402 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE,
SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,

[Continued on next page]

(54) Title: SYSTEM, METHOD AND SOFTWARE FOR AUTOMATIC HYPERLINKING OF PERSONS' NAMES IN DOCUMENTS TO PROFESSIONAL DIRECTORIES



(57) Abstract: Hyperlinking or associating documents to other documents based on the names of people in the documents has become more desirable. Although there is an automated system for installing such hyperlinks into judicial opinions, the system is not generally applicable to other types of names and documents, nor well suited to determine hyperlinks for names that might refer to two or more similarly named persons. Accordingly, the inventor devised systems, methods, and software that facilitate hyperlinking names in documents, regardless of type. One exemplary system includes a descriptor module and a linking module. The descriptor module develops descriptive patterns for selecting co-occurrent document information that is useful in recognizing associations between names and professional classes. The linking module tags names in an input document, extracts co-occurrent information using the descriptive patterns, and uses a Bayesian inference network that processes a (non-inverse-document-frequency) name-rarity score for each name along with the name and selected co-occurrent document information to determine appropriate hyperlinks to other documents, such as entries in professional directories.



ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK,
TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report:
26 August 2004

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference 962.017WO1	FOR FURTHER ACTION <small>see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.</small>	
International application No. PCT/US 02/41144	International filing date (day/month/year) 20/12/2002	(Earliest) Priority Date (day/month/year) 21/12/2001
Applicant WEST PUBLISHING COMPANY, DBA WEST GROUP		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 6 sheets.
☐ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the report

- a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.
- ☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).
- b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :
- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. ☒ **Certain claims were found unsearchable** (See Box I).

3. ☒ **Unity of invention is lacking** (see Box II).

4. With regard to the title,

- ☐ the text is approved as submitted by the applicant.
- ☒ the text has been established by this Authority to read as follows:

SYSTEM, METHOD AND SOFTWARE FOR AUTOMATIC HYPERLINKING OF PERSONS' NAMES IN DOCUMENTS TO PROFESSIONAL DIRECTORIES

5. With regard to the abstract,

- ☒ the text is approved as submitted by the applicant.
- ☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No.

- ☒ as suggested by the applicant.
- ☐ because the applicant failed to suggest a figure.
- ☐ because this figure better characterizes the invention.

1
☐ None of the figures.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 02/41144

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G06F17/30

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)

IPC 7 G06F

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

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

EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DOZIER C., HASCHART R.: "Automatic extraction and linking of person names in legal text"</p> <p>PROCEEDINGS OF RIAO 2000, [Online] April 2000 (2000-04), pages 1305-1321, XP002271447</p> <p>Retrieved from the Internet: URL: http://citeseer.nj.nec.com/408698.html > [retrieved on 2004-02-09] cited in the application the whole document</p> <p style="text-align: center;">----- -/--</p>	1-16, 34-39



Further documents are listed in the continuation of box C.



Patent family members are listed in 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 document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) 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

"T" later 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

"X" 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.

"&" document member of the same patent family

Date of the actual completion of the international search

2 March 2004

Date of mailing of the international search report

29. 06. 2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Denoual, M

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 02/41144

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>BORGMAN C L ET AL: "GETTY'S SYNONAME AND ITS COUSINS: A SURVEY OF APPLICATIONS OF PERSONAL NAME-MATCHING ALGORITHMS" JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE, AMERICAN SOCIETY FOR INFORMATION. WASHINGTON, US, vol. 43, no. 7, August 1992 (1992-08), pages 459-476, XP001176656 ISSN: 0002-8231 the whole document</p> <p>-----</p>	1-16, 34-39
X	<p>WINKLER ET AL: "Machine Learning, Information Retrieval, and Record Linkage" NISS WORKSHOP ON DATA QUALITY, XX, XX, 30 November 2000 (2000-11-30), XP002253864 the whole document</p> <p>-----</p>	1-16, 34-39
X	<p>NEWCOMBE H. B.: "Handbook of Record Linkage" 1988, OXFORD UNIVERSITY PRESS, OXFORD, XP002272144 pages 4-33</p> <p>-----</p>	1-6, 8-16, 34-39
A	<p>SMITH D A ET AL: "Disambiguating geographic names in a historical digital library" RESEARCH AND ADVANCED TECHNOLOGY FOR DIGITAL LIBRARIES. 5TH EUROPEAN CONFERENCE, ECDL 2001. PROCEEDINGS (LECTURE NOTES IN COMPUTER SCIENCE VOL.2163), RESEARCH AND ADVANCED TECHNOLOGY FOR DIGITAL LIBRARIES. 5TH EUROPEAN CONFERENCE, ECDL 2001. PROCEEDI, [Online] pages 127-136, XP002272130 2001, Berlin, Germany, Springer-Verlag, Germany ISBN: 3-540-42537-3 Retrieved from the Internet: URL:www.perseus.tufts.edu/Articles/geodl01.pdf> [retrieved on 2004-02-12] the whole document</p> <p>-----</p>	1-16, 34-39

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 02/41144

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 25-27
because they relate to subject matter not required to be searched by this Authority, namely:
Claims 25 to 27 are considered as claiming a mere presentation of information (Rule 39.1(v) PCT).
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-16, 34-39

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Claims Nos.: 25-27

Claims 25 to 27 are considered as claiming a mere presentation of information (Rule 39.1(v) PCT).

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-16,34-39

A method for automatically generating hyperlinks

2. claims: 17-24

A person name categorization method

3. claims: 28-33

A query disambiguation method
