



US007826665B2

(12) **United States Patent**
Bressan et al.

(10) **Patent No.:** **US 7,826,665 B2**
(45) **Date of Patent:** **Nov. 2, 2010**

(54) **PERSONAL INFORMATION RETRIEVAL
USING KNOWLEDGE BASES FOR OPTICAL
CHARACTER RECOGNITION CORRECTION**

(75) Inventors: **Marco Bressan**, La Tronche (FR);
Hervé Dejean, Grenoble (FR);
Christopher R. Dance, Meylan (FR)

2002/0131636 A1* 9/2002 Hou 382/181
2002/0191847 A1 12/2002 Newman et al.
2003/0069877 A1 4/2003 Grefenstette et al.
2003/0086615 A1 5/2003 Dance et al.
2005/0086205 A1 4/2005 Franciosa et al.
2005/0086224 A1 4/2005 Franciosa et al.

FOREIGN PATENT DOCUMENTS

(73) Assignee: **Xerox Corporation**, Norwalk, CT (US)

DE 10104270 8/2002
GB 2392290 2/2004

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1254 days.

OTHER PUBLICATIONS

Kilgariff et al., "Introduction to the Special Issue on the Web as Corpus," Computational Linguistics, vol. 29, No. 3, pp. 333-347, (2003).
Handley et al., "Document Understanding System Using Stochastic Context-Free Grammars," pp. 5.

(21) Appl. No.: **11/299,453**

(22) Filed: **Dec. 12, 2005**

Prior Publication Data

US 2007/0133874 A1 Jun. 14, 2007

(51) **Int. Cl.**
G06K 9/00 (2006.01)

(52) **U.S. Cl.** **382/181; 382/186; 382/187**

(58) **Field of Classification Search** 382/181,
382/231

See application file for complete search history.

References Cited

U.S. PATENT DOCUMENTS

5,483,052 A 1/1996 Smith, III et al.
5,604,640 A 2/1997 Zipf et al.
5,754,671 A* 5/1998 Higgins et al. 382/101
5,850,480 A 12/1998 Scanlon
6,783,060 B2 8/2004 Marappan
6,823,084 B2 11/2004 Myers et al.
7,120,302 B1* 10/2006 Billester 382/229
2001/0044324 A1 11/2001 Carayiannis et al.

(Continued)

Primary Examiner—Samir A. Ahmed

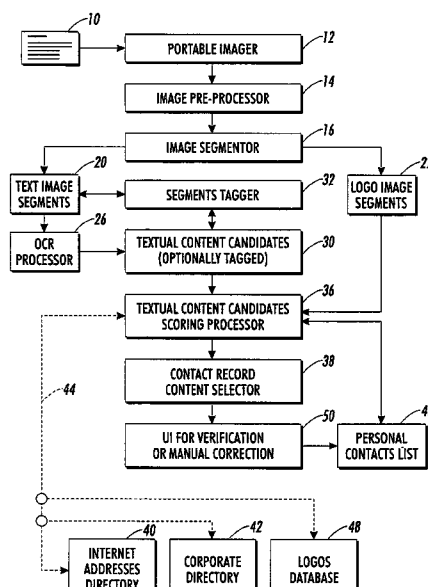
Assistant Examiner—Edward Park

(74) *Attorney, Agent, or Firm*—Fay Sharpe LLP

(57) **ABSTRACT**

In a system for updating a contacts database (42, 46), a portable imager (12) acquires a digital business card image (10). An image segmenter (16) extracts text image segments from the digital business card image. An optical character recognizer (OCR) (26) generates one or more textual content candidates for each text image segment. A scoring processor (36) scores each textual content candidate based on results of database queries respective to the textual content candidates. A content selector (38) selects a textual content candidate for each text image segment based at least on the assigned scores. An interface (50) is configured to update the contacts list based on the selected textual content candidates.

15 Claims, 3 Drawing Sheets



OTHER PUBLICATIONS

Strohmaier et al., "Lexical Postcorrection of OCR-Results: The web as a Dynamic Secondary Dictionary?," pp. 5.

Martins et al., "Spelling Correction for Search Engine Queries," pp. 12.

Xerox, "Xerox Document Imaging Technology Changes the Way People Communicate," Public Relations Department, France, pp. 2, (2004).

Katsuyama et al., "Highly Accurate Retrieval of Japanese Document Images Through a Combination . . .," Proc SPIE vol. 2670, pp. 57-67, (2002).

Saiga et al., "An OCR System for Business Cards," IEEE Comput. Soc., pp. 802-805, 1993.

Likforman-Sulem et al., "Proper Names Extraction from Fax Images Combining Textual and Image Features," IEEE, pp. 545-549, 2003.

Luo et al., "Design and implementation of a card reader based on build-in camera," IEEE, vol. 1, pp. 417-420, 2004.

* cited by examiner

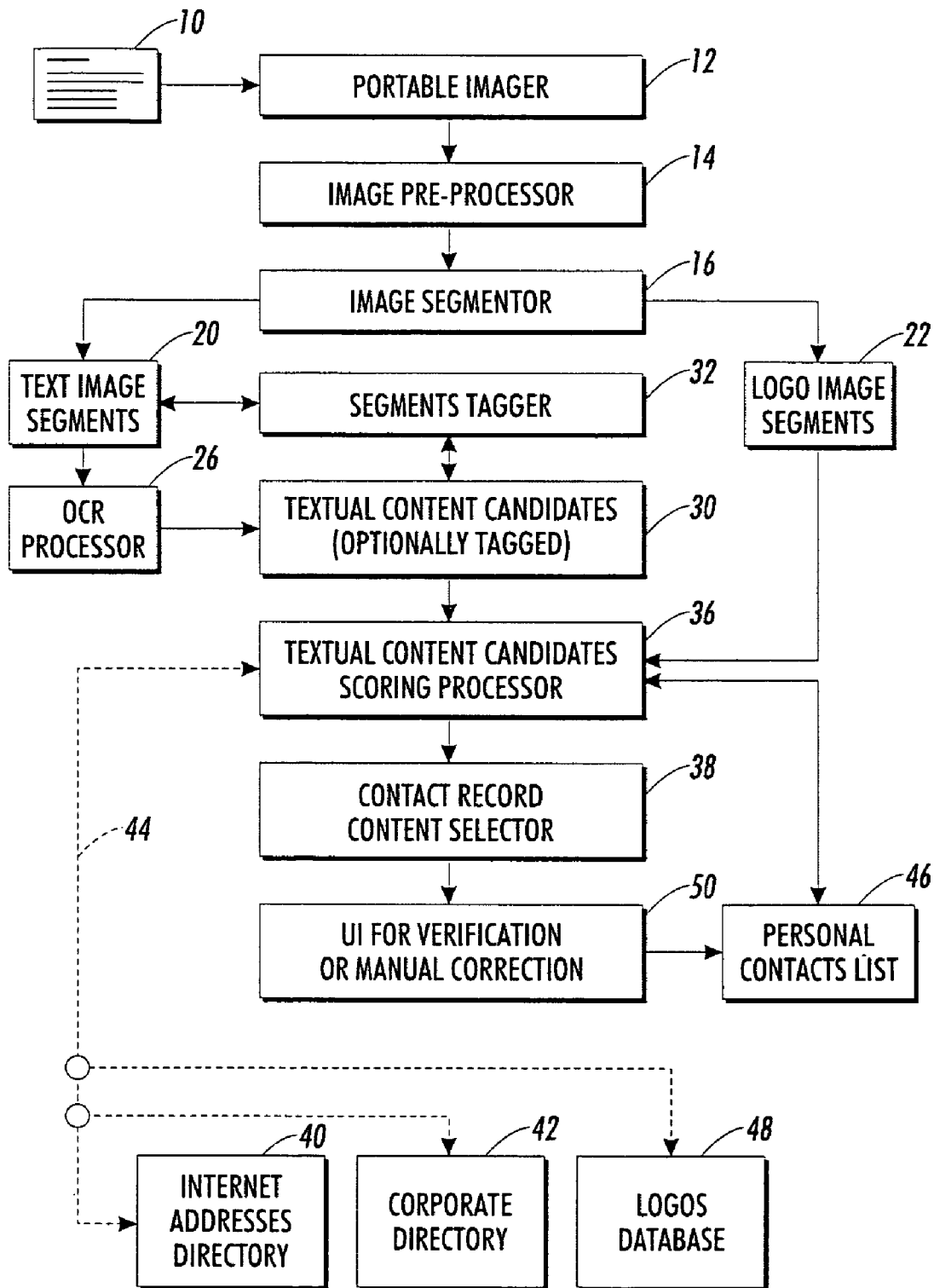


FIG. 1

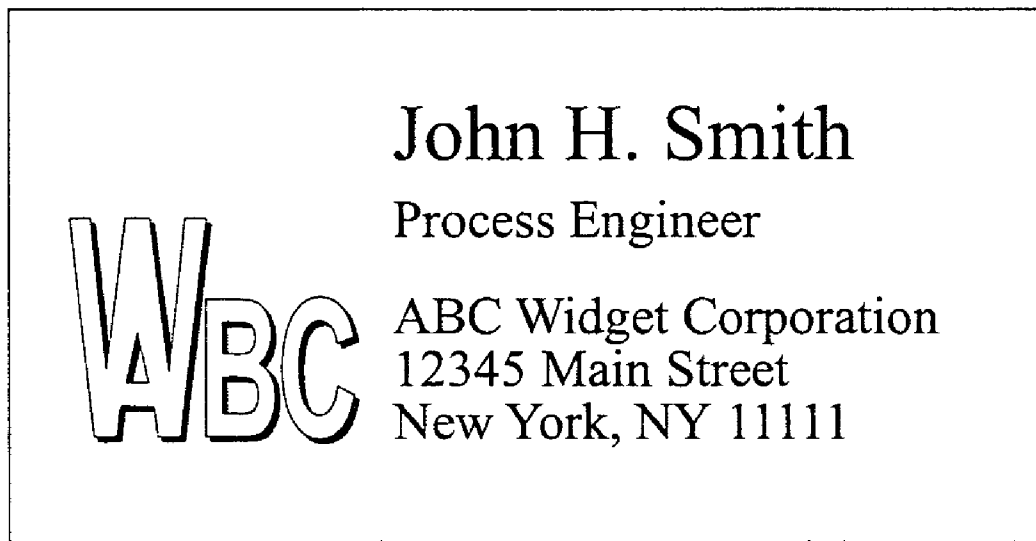
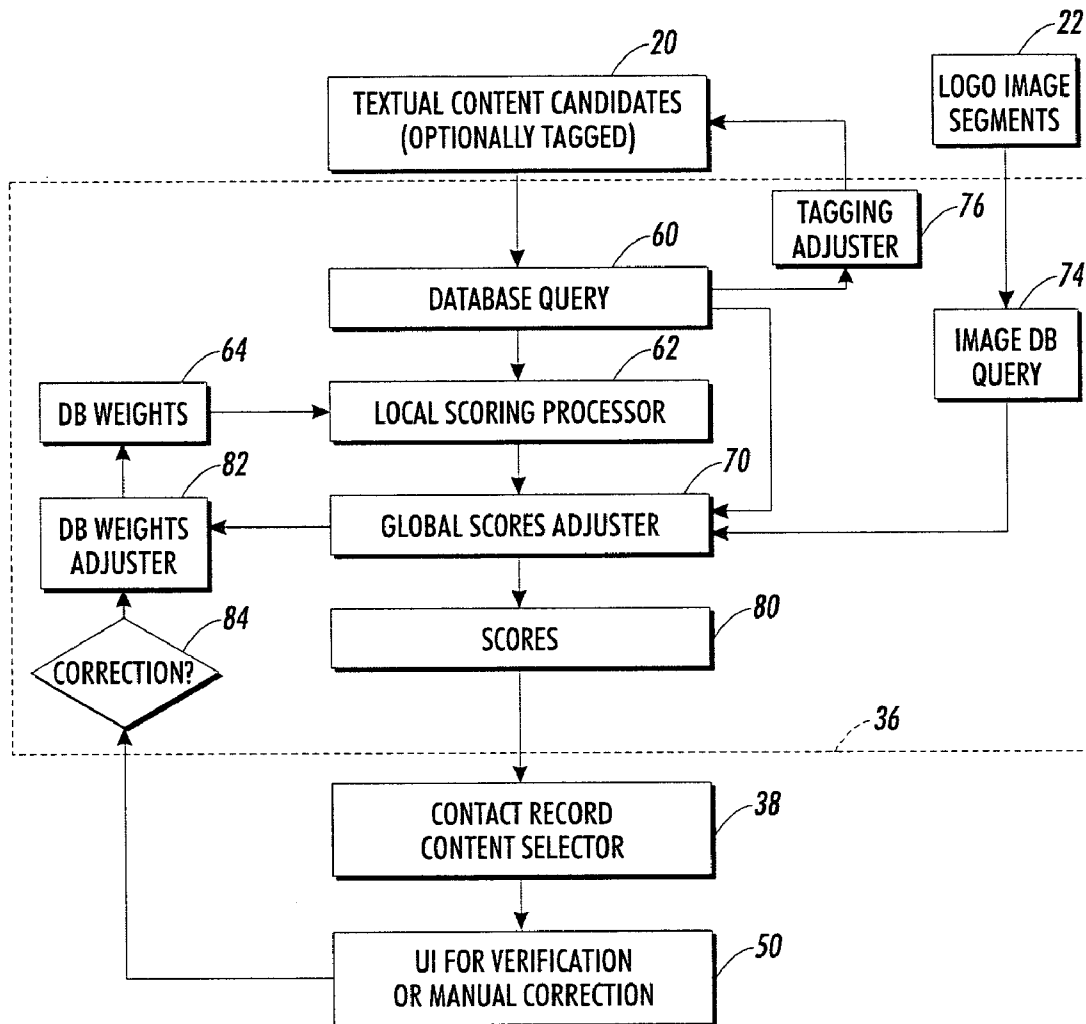


FIG. 2

**FIG. 3**

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts



Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research



With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips



Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.