US008489388B2

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

(10) **Patent No.: US 8,489,388 B2**
(45) **Date of Patent: Jul. 16, 2013**

(54) **DATA DETECTION**

(75) Inventors: **Olivier Bonnet**, Paris (FR); **Frederick de Jaeger**, Paris (FR); **Romain Goyet**, Paris (FR); **Jean-Pierre Ciudad**, San Francisco, CA (US)

(73) Assignee: **Apple Inc.**, Cupertino, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1222 days.

(21) Appl. No.: **12/268,410**

(22) Filed: **Nov. 10, 2008**

(65) **Prior Publication Data**

US 2010/0121631 A1 May 13, 2010

(51) **Int. Cl.**
G06F 17/27 (2006.01)

(52) **U.S. Cl.**
USPC **704/9**; 707/755; 717/143; 717/144

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,227,245	A	10/1980	Edblad et al.
4,791,556	A	12/1988	Vilkaitis
4,818,131	A	4/1989	Sakai
4,873,662	A	10/1989	Sargent
4,907,285	A	3/1990	Nakano et al.
4,965,763	A	10/1990	Zamora
5,034,916	A	7/1991	Ordish
5,146,406	A	9/1992	Jensen
5,155,806	A	10/1992	Hoeber et al.
5,157,736	A	10/1992	Boyer et al.
5,182,709	A	1/1993	Makus
5,189,632	A	2/1993	Paajanen et al.
5,283,856	A	2/1994	Gross et al.
5,299,261	A	3/1994	Bogart et al.
5,301,350	A	4/1994	Rogan et al.
5,346,516	A	9/1994	Alkhas et al.
5,369,778	A	11/1994	San Soucie et al.
5,375,200	A	12/1994	Dugan et al.
5,390,281	A	2/1995	Luciw et al.
5,398,336	A	3/1995	Tantry et al.
5,418,717	A	5/1995	Su et al.
5,424,947	A	6/1995	Nagao et al.

(Continued)

FOREIGN PATENT DOCUMENTS

EP	0 458 563	A2	11/1991
EP	0 458 563	B1	11/1991

(Continued)

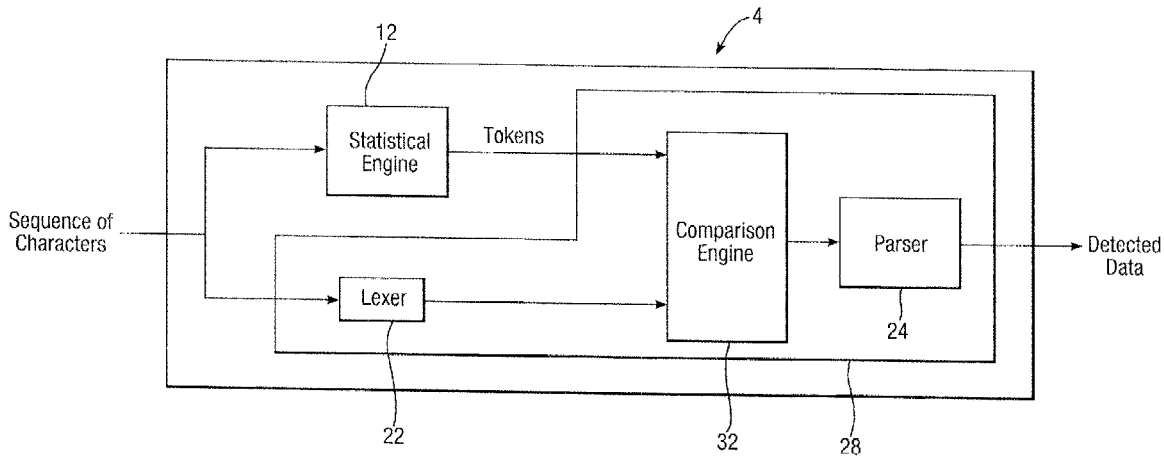
OTHER PUBLICATIONS

Abbasi. "Information Extraction Techniques for Postal Address Standardization" 2005.*

(Continued)

Primary Examiner — Greg Borsetti
(74) *Attorney, Agent, or Firm* — Blakely, Sokoloff, Taylor & Zafman LLP

(57) **ABSTRACT**

A method for detecting data in a sequence of characters or text using both a statistical engine and a pattern engine. The statistical engine is trained to recognize certain types of data and the pattern engine is programmed to recognize the grammatical pattern of certain types of data. The statistical engine may scan the sequence of characters to output first data, and the pattern engine may break down the first data into subsets of data. Alternatively, the statistical engine may output items that have a predetermined probability or greater of being a certain type of data and the pattern engine may then detect the data from the output items and/or remove incorrect information from the output items.

29 Claims, 11 Drawing Sheets