US007912828B2

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

(10) **Patent No.: US 7,912,828 B2**
(45) **Date of Patent: Mar. 22, 2011**

(54) **PATTERN SEARCHING METHODS AND APPARATUSES**

(75) Inventors: **Olivier Bonnet**, Paris (FR); **Frédéric De Jaeger**, Paris (FR); **Toby Paterson**, Paris (FR)

(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 298 days.

(21) Appl. No.: **11/710,182**

(22) Filed: **Feb. 23, 2007**

(65) **Prior Publication Data**

US 2008/0243841 A1 Oct. 2, 2008

(51) **Int. Cl.**
G06F 7/00 (2006.01)
G06F 15/16 (2006.01)
G10L 15/06 (2006.01)

(52) **U.S. Cl.** **707/706**; 704/243; 704/244; 709/246

(58) **Field of Classification Search** 707/6, 102; 709/246; 704/243, 244
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,791,556	A	12/1988	Vilkaitis
5,034,916	A	7/1991	Ordish
5,155,806	A	10/1992	Hoeber et al.
5,189,632	A	2/1993	Paajanen et al.
5,390,281	A	2/1995	Luciw et al.
5,434,777	A	7/1995	Luciw
5,442,742	A	8/1995	Greyson et al.
5,477,447	A	12/1995	Luciw et al.
5,583,921	A	12/1996	Hidaka
5,608,624	A	3/1997	Luciw
5,621,903	A	4/1997	Luciw et al.
5,627,948	A	5/1997	Fukunaga
5,644,735	A	7/1997	Luciw et al.
5,666,552	A	9/1997	Greyson et al.
5,687,333	A	11/1997	Dobashi et al.
5,692,032	A	11/1997	Seppänen et al.
5,708,845	A	1/1998	Wistendahl et al.

(Continued)

FOREIGN PATENT DOCUMENTS

EP 0 458 563 A2 11/1991

(Continued)

OTHER PUBLICATIONS

"Postal Address Detection from Web Documents" Lin et al., IEEE Computer Society, copyright 2005 (hereafter Lin).*

(Continued)

Primary Examiner — Apu M Mofiz
Assistant Examiner — Cindy Nguyen
(74) *Attorney, Agent, or Firm* — Blakely, Sokoloff, Taylor & Zafman LLP

(57) **ABSTRACT**

A computer-based method for identifying patterns in computer text using structures defining types of patterns which are to be identified, wherein a structure comprises one or more definition items, the method comprising assigning a weighting to each structure and each definition item; searching the computer text for a pattern to be identified on the basis of a particular structure, a pattern being provisionally identified if it matches the definition given by said particular structure; in a provisionally identified pattern, determining those of the definition items making up said particular structure that have been identified in the provisionally identified pattern; combining the weightings of the determined definition items and optionally, the weighting of the particular structure, to a single quantity; assessing whether the single quantity fulfils a given condition; depending on the result of said assessment, rejecting or confirming the provisionally identified pattern.

12 Claims, 10 Drawing Sheets

Examples of conventional structure definitions

```
#1:
us_state_code{bind:state} := "AL" | "AK" | ... | "WY";
#2:
us_zip_code{bind:code} := [0-9]{5};
#3:
postal_code :=
// ex: CA. 12345
    us_state_code "."? (some_spaces us_zip_code)?
// ex: 12345 CA
    | us_zip_code (some_spaces us_state_code)?
#4:
number := [0-9]+;
#5:
known_street_type := "Street" | "Boulevard" | ...;
#6:
city := known_city | some_capitalized_words;
#7:
street_address{bind} :=
  number{bind:number} some_spaces some_capitalized_words
{known_street_type{bind:type}}? comma_or_spaces postal_code? some_spaces city;
```