(12) **United States Patent**　(10) **Patent No.:**　　**US 6,366,908 B1**
Chong et al.　(45) **Date of Patent:**　　**Apr. 2, 2002**

(54) **KEYFACT-BASED TEXT RETRIEVAL SYSTEM, KEYFACT-BASED TEXT INDEX METHOD, AND RETRIEVAL METHOD**

(75) Inventors: **Kyung Taek Chong**; **Myung-Gil Jang**; **MiSeon Jun**; **Se Young Park**, all of Taejon (KR)

(73) Assignee: **Electronics and Telecommunications Research Institute**, Taejon (KR)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/475,743**

(22) Filed: **Dec. 30, 1999**

(30) **Foreign Application Priority Data**

Jun. 28, 1999　(KR) 99-25035

(51) **Int. Cl.^{7}** **G06F 17/30**
(52) **U.S. Cl.** **707/3**; 707/5; 707/6; 707/101; 707/201; 704/7; 704/9; 704/10; 382/177; 382/306
(58) **Field of Search** 707/1, 2, 3, 4, 707/5, 6, 102, 103, 104.1; 706/12, 45, 47; 704/7, 9, 10, 270.1

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,839,853	A		6/1989	Deerwester et al. 364/900
5,289,375	A	*	2/1994	Fukumochi et al. ... 364/419.02
5,535,382	A		7/1996	Ogawa 395/600
5,541,836	A	*	7/1996	Church et al. 364/419.07
5,598,557	A	*	1/1997	Doner et al. 395/605
5,708,829	A		1/1998	Kadashevich et al. 395/793
5,721,902	A	*	2/1998	Schultz 395/604
5,771,378	A	*	6/1998	Holt et al. 395/605

OTHER PUBLICATIONS

Jang, Ho Wook and Se Young Park, "Keyfact Concept For An Information Retrieval System," Proceedings of Natural Language Proc. Pacific Rim Symposium, pp. 510–513, 1995.

Jun, MiSeon and Se Young Park, "Keyfact–Based Information Retrieval System," International Symposium on Digital Library, pp. 521–524, 1997.

Arampatzis et al., "Phrase–Based Information Retrieval," *Journal of Information Proc. & Management 34*(6):1–19, 1998.

Robert Krovetz and W. Bruce Croft:"Lexical ambiguity and information retrieval", Apr. 1992, ACM, vol. 10, pp. 115–141.*

Michael Sussna: "Word sense disambiguation for free–text indexing using a massive semantic network", 1993, ACM, pp. 67–74.*

* cited by examiner

Primary Examiner—Thomas Black
Assistant Examiner—Jacques Veillard
(74) *Attorney, Agent, or Firm*—E. Russell Tarleton; SEED IP Law Group PLLC

(57) **ABSTRACT**

A keyfact-based text retrieval method and a keyfact-based text index method that describes the formalized concept of a document by a pair comprising an object that is the head and a property that is the modifier and uses the information described by the pairs as index information for efficient document retrieval. A keyfact-based text retrieval system includes keyfact extracting, keyfact indexing, and keyfact retrieving. The keyfact extracting analyzes a document collection and a query and extracts keywords and keyfacts. The keywords do not have part-of-speech ambiguity and the keyfacts are extracted from the keywords. The keyfact indexing calculates the frequency of the keyfacts and generates a keyfact list of the document collection for a keyfact index structure. The keyfact retrieving receive a keyfact of the query and keyfacts of the document collection and defines a keyfact-based retrieval model in consideration of a weight factor of the keyfact pattern and generates a retrieval result. The retrieval result is a document similar to the query.
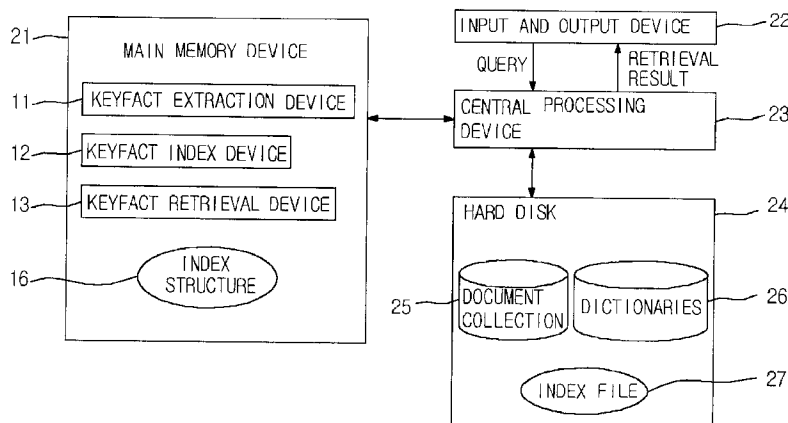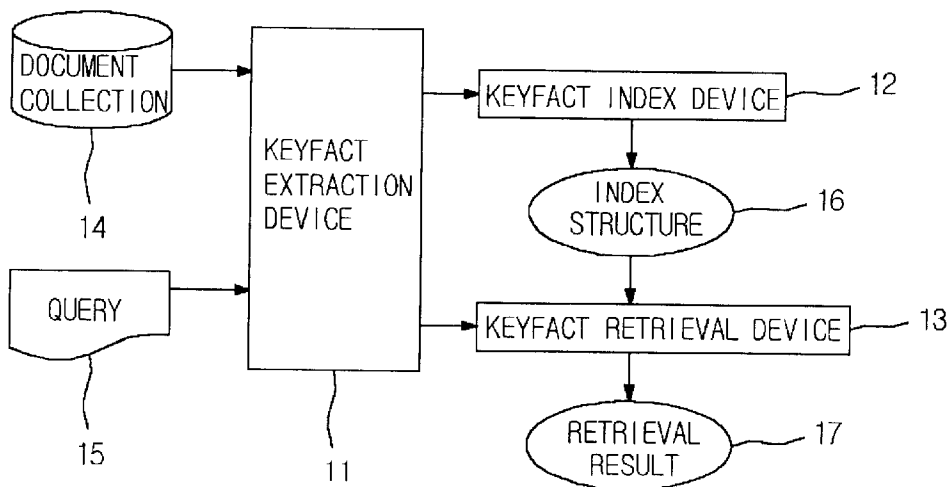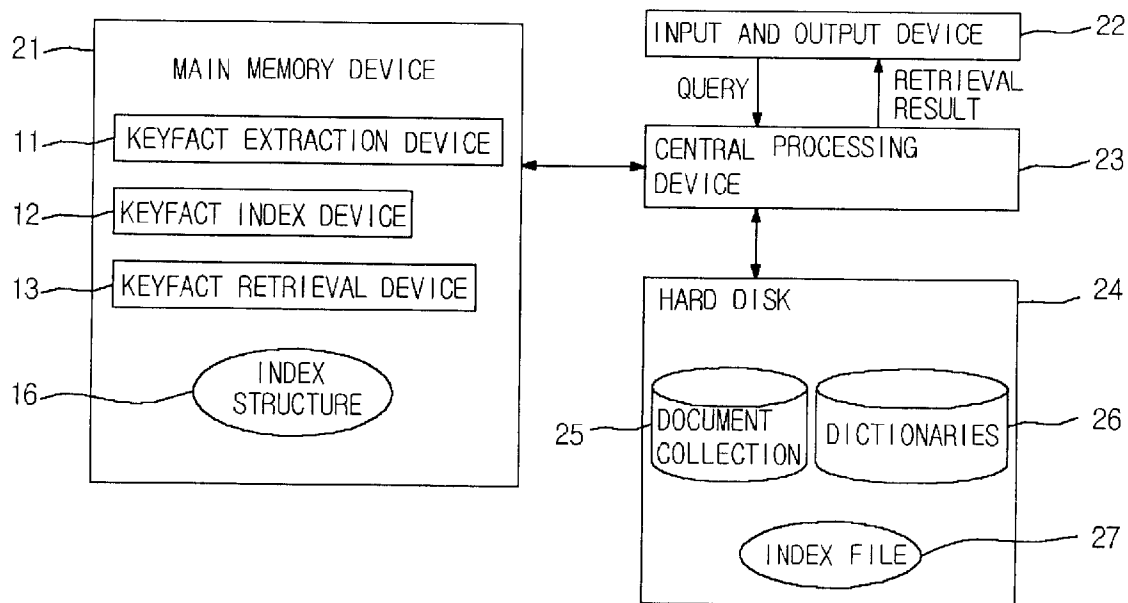
12 Claims, 5 Drawing Sheets

FIG. 1

FIG. 2

FIG. 3

START

31 — INPUT DOCUMENT

32 — MORPHOLOGICAL ANALYSIS

33 — PART-OF-SPEECH TAGGING

34 — KEYFACT PATTERN EXTRACTION

35 — KEYFACT GENERATION

39 — KEYFACT

END

36 — DICTIONARIES

37 — KEYFACT PATTERN RULE

38 — KEYFACT GENERATION RULE

FIG. 4

FIG. 5

51

KEFACT INDEX DEVICE

START

52 — INDEX FILE

QUERY

KEYFACT EXTRACTION DEVICE — 53

55

FORMING DOCUMENT,
QUERY VECTOR — 54

DETERMINING
KEYFACT WEIGHT
CONSTANT C$_{kfType\#}$

CALCULATING DOCUMENT
QUERY WEIGHT — 56

57 — RANKED RETRIEVAL RESULT
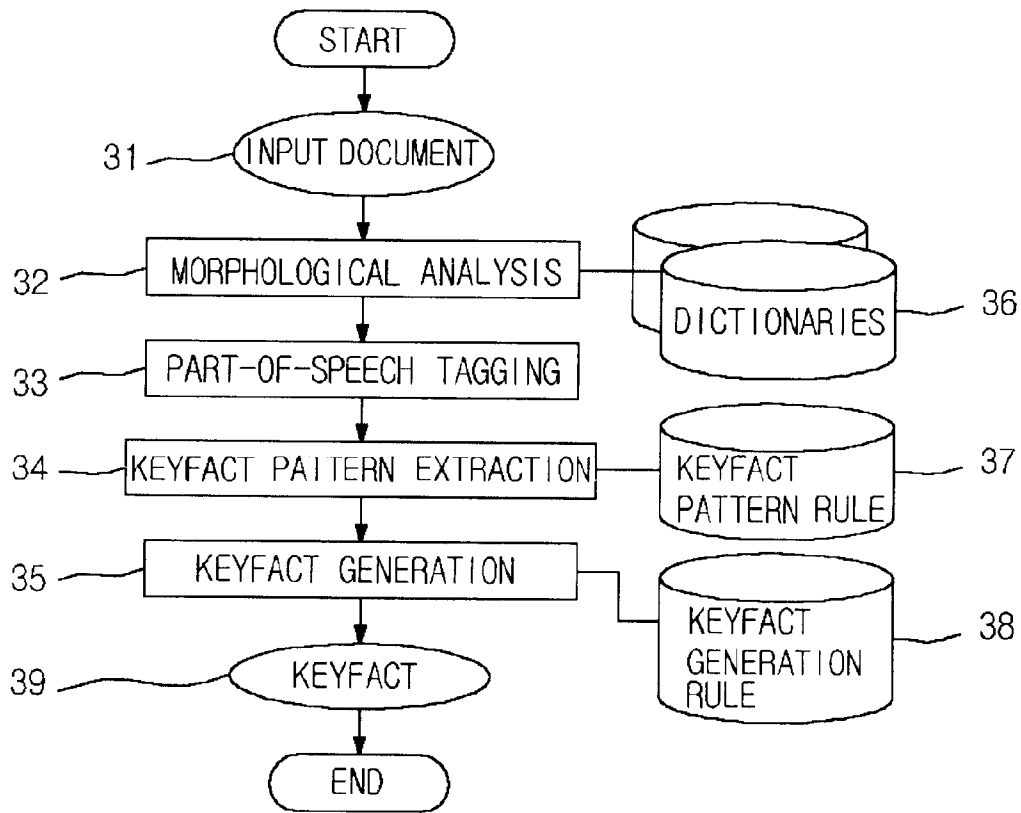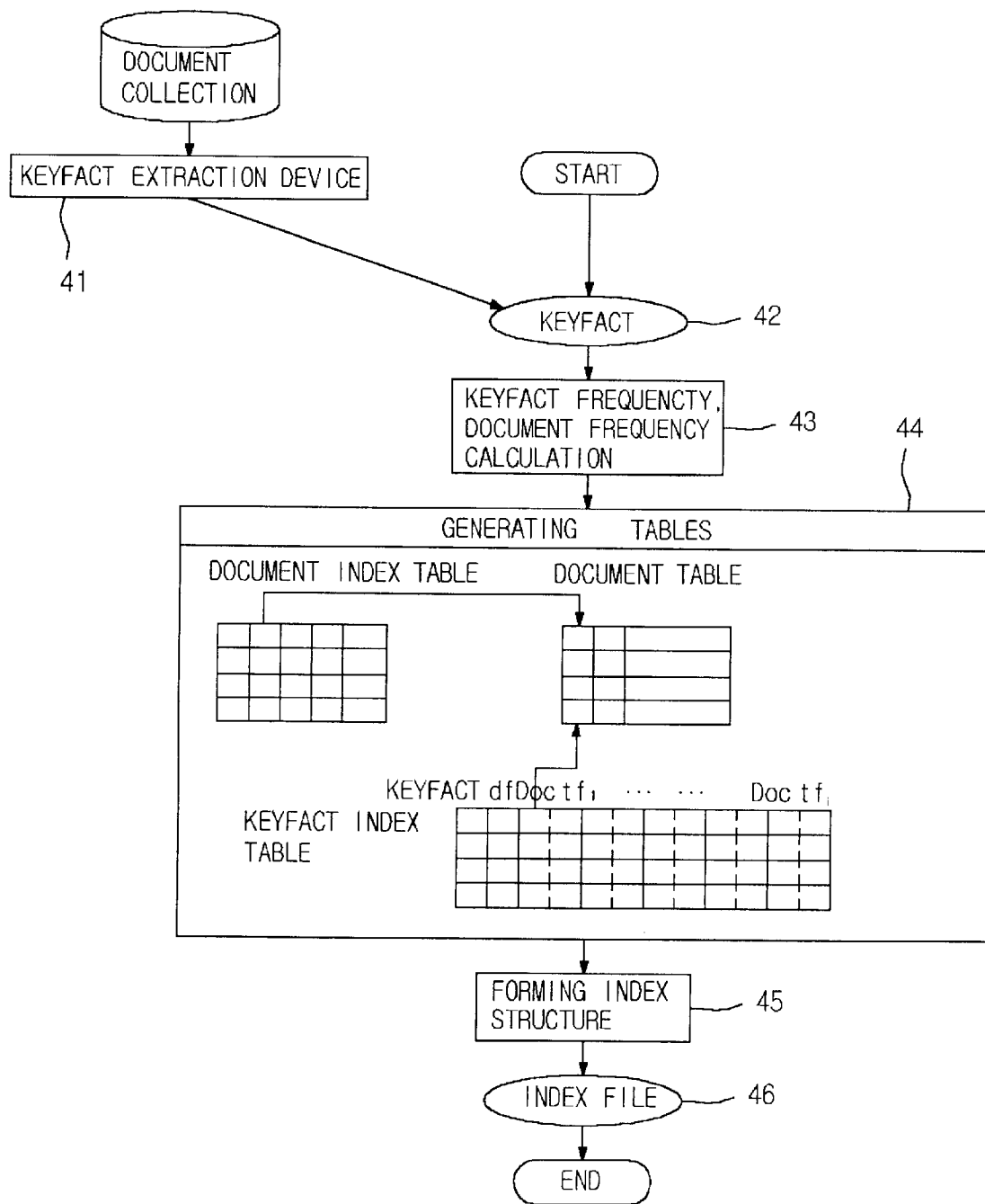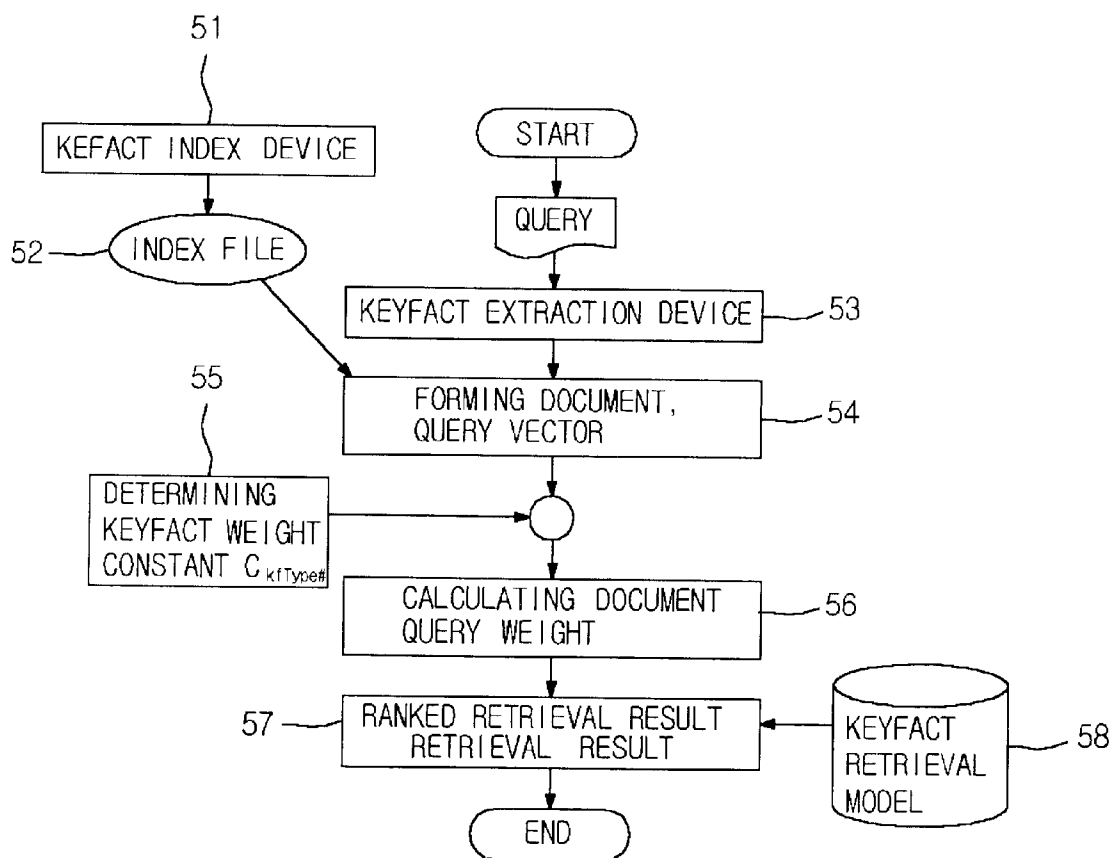RETRIEVAL RESULT

KEYFACT
RETRIEVAL
MODEL — 58

END

DOCKET ALARM

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.