US005933822A

United States Patent [19]

Braden-Harder et al.

[11] **Patent Number:** 5,933,822

[45] **Date of Patent:** **Aug. 3, 1999**

[54] **APPARATUS AND METHODS FOR AN INFORMATION RETRIEVAL SYSTEM THAT EMPLOYS NATURAL LANGUAGE PROCESSING OF SEARCH RESULTS TO IMPROVE OVERALL PRECISION**

[75] Inventors: **Lisa Braden-Harder**, Reston, Va.; **Simon H. Corston**, Seattle, Wash.; **William B. Dolan**, Redmond, Wash.; **Lucy H. Vanderwende**, Bellevue, Wash.

[73] Assignee: **Microsoft Corporation**, Redmond, Wash.

[21] Appl. No.: **08/898,652**

[22] Filed: **Jul. 22, 1997**

[51] **Int. Cl.^{6}** **G06F 17/00**
[52] **U.S. Cl.** **707/5**; 707/3; 707/4
[58] **Field of Search** 707/5, 4, 3; 364/419.19; 395/708

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,278,980	1/1994	Pederson et al.	707/4
5,544,049	8/1996	Henderson et al.	364/419.19
5,642,502	6/1997	Driscoll	707/5
5,671,404	9/1997	Lizee et al.	707/5
5,694,592	12/1997	Driscoll	707/3
5,706,497	1/1998	Takahashi et al.	707/5
5,724,571	3/1998	Woods	707/5
5,794,050	8/1998	Dahlgren et al.	395/708
5,826,261	10/1998	Spencer	707/5

OTHER PUBLICATIONS

B. Katz, "Annotating the World Wide Web using Natural Language", *Conference Proceedings of RIAO* 97, Computer–Assisted Information Searching in Internet, McGill University, Quebec, Canada, Jun. 25–27 1997, vol. 1, pp. 136–155.
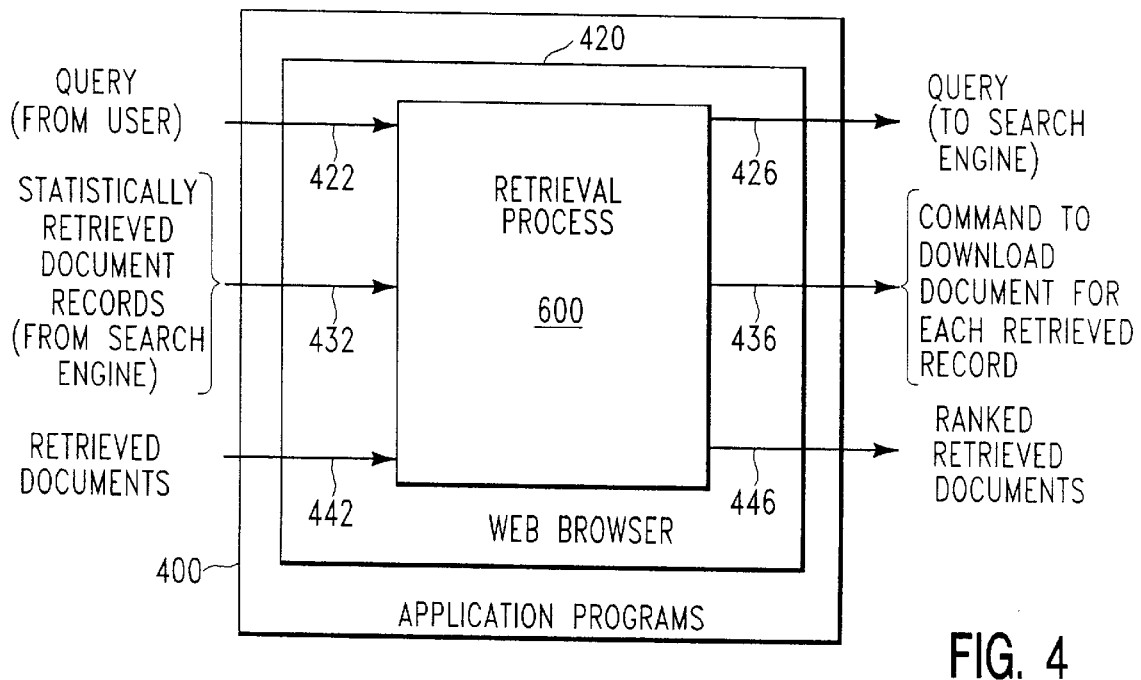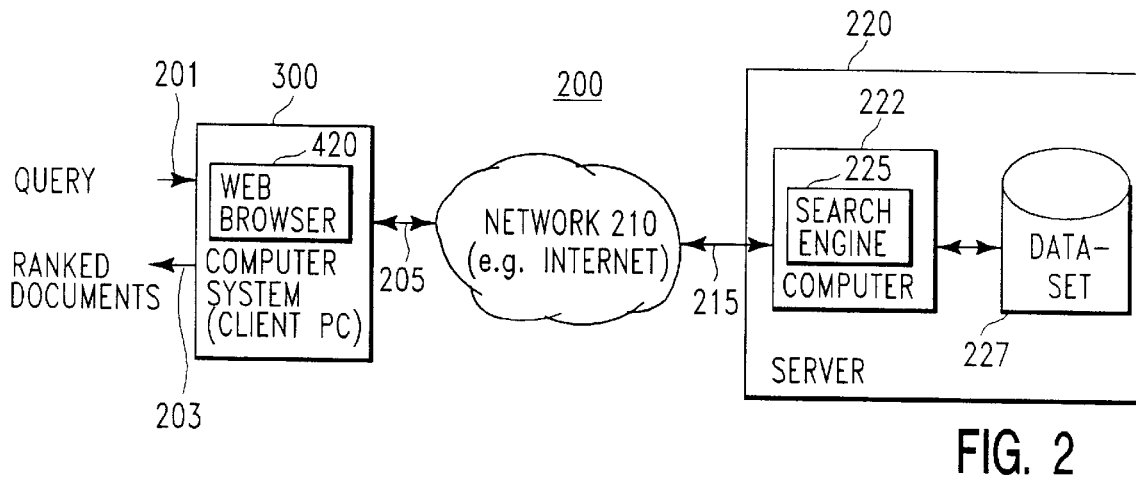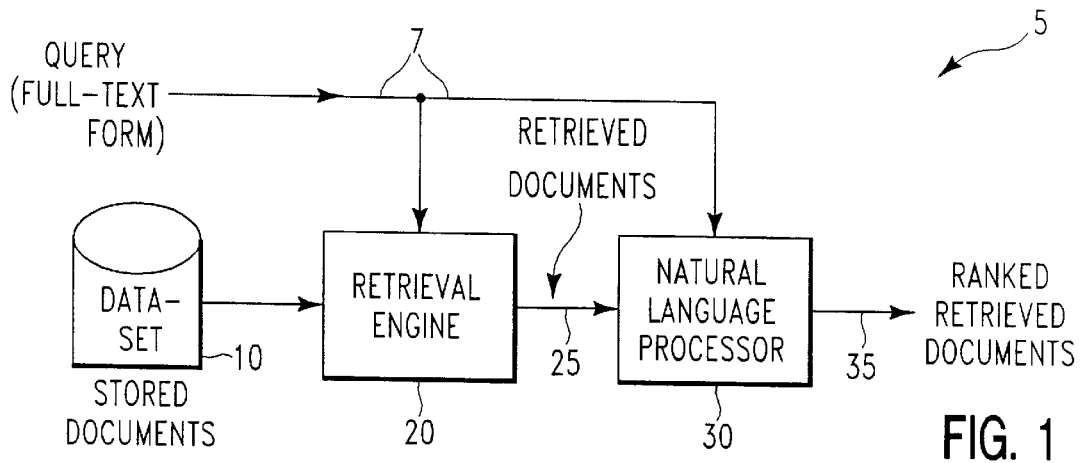
A. T. Arampatzis et al, "IRENA: Information Retrieval Engine based on Natural language Analysis", *Conference Proceedings of RIAO* 97, Computer–Assisted Information Searching in Internet, McGill University, Quebec, Canada, Jun. 25–27, 1997, vol. 1, pp. 159–175.

R. Pohlmann et al, "The Effect of Syntactic Phrase Indexing on Retrieval Performance for Dutch Tests", *Conference Proceedings of RIAO* 97, Computer–Assisted Information Searching in Internet, McGill University, Quebec, Canada, Jun. 25–27, 1997, vol. 1, pp. 176–187.

(List continued on next page.)

Primary Examiner—Wayne Amsbury
Assistant Examiner—Thuy Pardo
Attorney, Agent, or Firm—Michaelson & Wallace; Peter L. Michaelson

[57] **ABSTRACT**

Apparatus and accompanying methods for an information retrieval system that utilizes natural language processing to process results retrieved by, for example, an information retrieval engine such as a conventional statistical-based search engine, in order to improve overall precision. Specifically, such a search ultimately yields a set of retrieved documents. Each such document is then subjected to natural language processing to produce a set of logical forms. Each such logical form encodes, in a word-relation-word manner, semantic relationships, particularly argument and adjunct structure, between words in a phrase. A user-supplied query is analyzed in the same manner to yield a set of corresponding logical forms therefor. Documents are ranked as a predefined function of the logical forms from the documents and the query. Specifically, the set of logical forms for the query is then compared against a set of logical forms for each of the retrieved documents in order to ascertain a match between any such logical forms in both sets. Each document that has at least one matching logical forms is heuristically scored, with each different relation for a matching logical forms being assigned a different corresponding predefined weight. The score of each such document is, e.g., a predefined function of the weights of its uniquely matching logical forms. Finally, the retained documents are ranked in order of descending score and then presented to a user in that order.

123 Claims, 14 Drawing Sheets

OTHER PUBLICATIONS

M. Mitra et al, "An Analysis of Statistical and Syntactic Phrases", *Conference Proceedings of RIAO* 97, Computer–Assisted Information Searching in Internet, McGill University, Quebec, Canada, Jun. 25–27, 1997, vol. 1, pp. 200–214.

P. Bruza et al, "Query ReFormulation on the Internet: Empirical Data and the Hyperindex Search Engine", *Conference Proceedings of RIAO* 97, Computer–Assisted Information Searching in Internet, McGill University, Quebec, Canada, Jun. 25–27, 1997, vol. 1, pp. 488–499.

G. Grefenstette, "SQLET: Short Query Linguistic Expansion Techniques, Palliating One–Word Queries by Providing Intermediate Structure to Text", *Conference Proceedings of RIAO* 97, Computer–Assisted Information Searching in Internet, McGill University, Quebec, Canada, Jun. 25–27, 1997, vol. 1, pp. 500–509.

R. Chandrasekar et al, "Using Syntactic Information in Document Filtering: A Comparative Study of Part–of–Speech Tagging and Supertagging", *Conference Proceedings of RIAO* 97, Computer–Assisted Information Searching in Internet, McGill University, Quebec, Canada, Jun. 25–27, 1997, vol. 1, pp. 531–545.

M.A. Hearst, "TextTiling: Segmenting Text into Multi–paragraph Subtopic Passages", *Computational Linguistics*, vol. 23, No. 1, 1997, pp. 33–64.

O. Etzoni, "The World–Wide Web: Quagmire or Gold Mine", *Communications of the ACM*, Nov. 1996, vol. 39, No. 11, pp. 65–68.

T. Strzalkowski, "Natural Language Information Retrieval: TIPSTER–2 Final Report", *Proceedings of Advances in Text Processing: Tipster Program Phase* 2, DARPA, May 6–8, 1996, Tysons Corner, Virginia, pp. 143–148.

D. D. Lewis et al, "Natural language Processing for Information Retrieval", *Communications of the ACM*, Jan. 1996, vol. 39, No. 1, pp. 92–101.

T. Strzalkowski, "Natural Language Information Retrieval", *Information Processing and Management*, vol. 31, No. 3, 1995, pp. 397–417.

K. Jensen et al (editors), *Natural Language Processing: The PLNLP Approach* (© 1993, Kluwer Academic Publishers), Chapter 3 "PEG: The PLNLP English Grammar", pp. 29–45 and Chapter 16 "PEGASUS: Deriving Argument Structures after Syntax", pp. 203–214.

J. L. Fagan, "Experiments in Automatic Phrase Indexing for Document Retrieval: A Comparison of Syntactic and Non–Syntactic Methods", Ph.D. Thesis, Cornell University, 1988, pp. i–261.
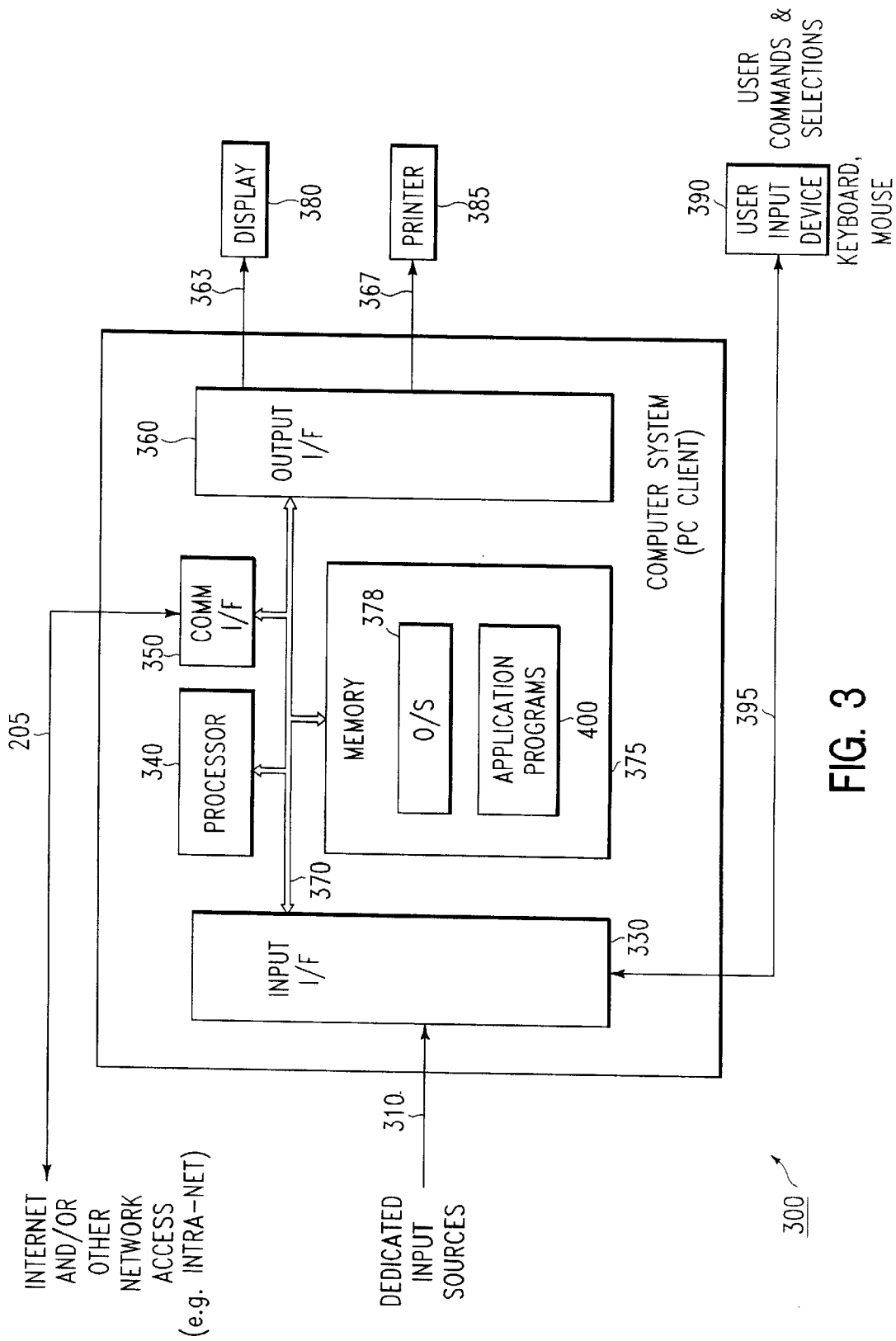
FIG. 1

FIG. 2

FIG. 4

FIG. 3

510 — INPUT STRING: THE OCTOPUS HAS THREE HEARTS.

LOGICAL FORM
GRAPH:

```
HAVE
  ├── Dsub ── OCTOPUS          ╱ 515
  └── Dobj ──  HEART
                  └── Ops ── THREE
```

LOGICAL FORM
TRIPLES:

```
HAVE — Dsub — OCTOPUS ⎫
HAVE — Dobj — HEART   ⎬ 525
HEART — Ops — THREE   ⎭
```

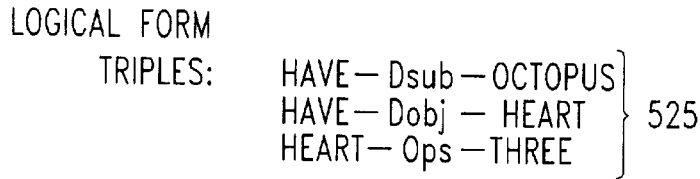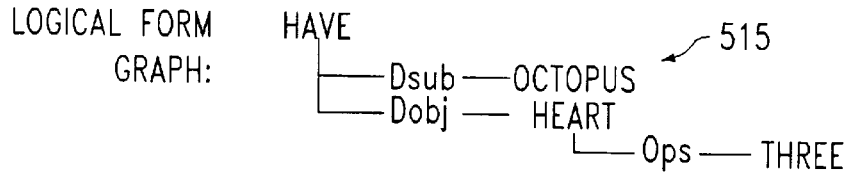## FIG. 5A

530 — INPUT STRING: THE OCTOPUS HAS THREE HEARTS AND TWO LUNGS.

LOGICAL FORM
GRAPH:

```
HAVE
  ├──── Dsub — OCTOPUS          ╱ 535
  └──── Dobj — AND
                 │
              Crds ───── HEART
                           │
                         Ops ── THREE
                 │
              ───── LUNG
                      │
                    Ops ── TWO
```

LOGICAL FORM
TRIPLES:

```
HAVE  — Dsub — OCTOPUS ⎫
HAVE  — Dobj — HEART   ⎪
HAVE  — Dobj — LUNG    ⎬ 540
HEART — Ops  — THREE   ⎪
LUNG  — Ops  — TWO     ⎭
```
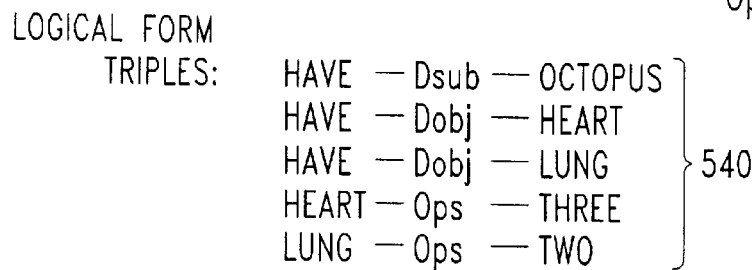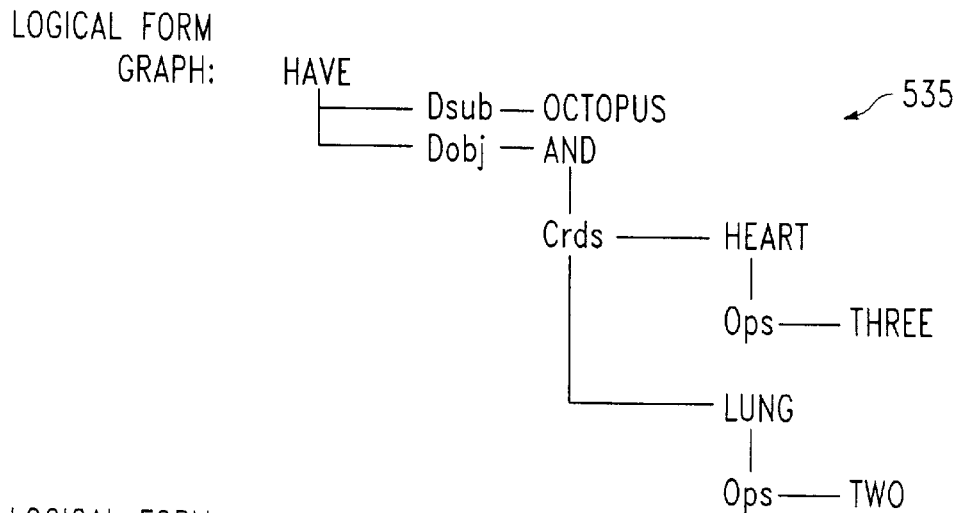
FIG. 5B

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.