



United States Patent [19]

[11] Patent Number: **6,076,051**

Messerly et al.

[45] Date of Patent: **Jun. 13, 2000**

- [54] **INFORMATION RETRIEVAL UTILIZING SEMANTIC REPRESENTATION OF TEXT**
- [75] Inventors: **John J. Messerly**, Bainbridge Island; **George E. Heidorn**, Bellevue; **Stephen D. Richardson**; **William B. Dolan**, both of Redmond; **Karen Jensen**, Bellevue, all of Wash.
- [73] Assignee: **Microsoft Corporation**, Redmond, Wash.
- [21] Appl. No.: **08/886,814**
- [22] Filed: **Mar. 7, 1997**
- [51] Int. Cl.⁷ **G06F 17/27**; G06F 17/30
- [52] U.S. Cl. **704/9**; 704/10; 707/2; 707/5
- [58] **Field of Search** 704/1, 9, 10; 707/3, 707/4, 5, 6, 1, 2, 104, 530, 531, 532

OTHER PUBLICATIONS

Van Zuijlen, Job M., "Probabilistic Methods in Dependency Grammar Parsing," *International Parsing Workshop*, 10375:142-151, 1989.

Gerard Salton, "Automatic Information Organization and Retrieval," McGraw Hill Book Company, pp. 168-178 (1968).

Fagan, Joel L., Ph.D., "Experiments in automatic phrase indexing for document retrieval: A comparison of syntactic and non-syntactic methods," Cornell University, UMI Dissertation Information Service, pp. 1-261 (1987).

James Allen, "Natural Language Understanding," The Benjamin/Cummings Publishing Company, Inc., Chapter 8, pp. 227-238, 1995.

Primary Examiner—Joseph Thomas

[57] ABSTRACT

The present invention is directed to performing information retrieval utilizing semantic representation of text. In a preferred embodiment, a tokenizer generates from an input string information retrieval tokens that characterize the semantic relationship expressed in the input string. The tokenizer first creates from the input string a primary logical form characterizing a semantic relationship between selected words in the input string. The tokenizer then identifies hypernyms that each have an "is a" relationship with one of the selected words in the input string. The tokenizer then constructs from the primary logical form one or more alternative logical forms. The tokenizer constructs each alternative logical form by, for each of one or more of the selected words in the input string, replacing the selected word in the primary logical form with an identified hypernym of the selected word. Finally, the tokenizer generates tokens representing both the primary logical form and the alternative logical forms. The tokenizer is preferably used to generate tokens for both constructing an index representing target documents and processing a query against that index.

[56] **References Cited**

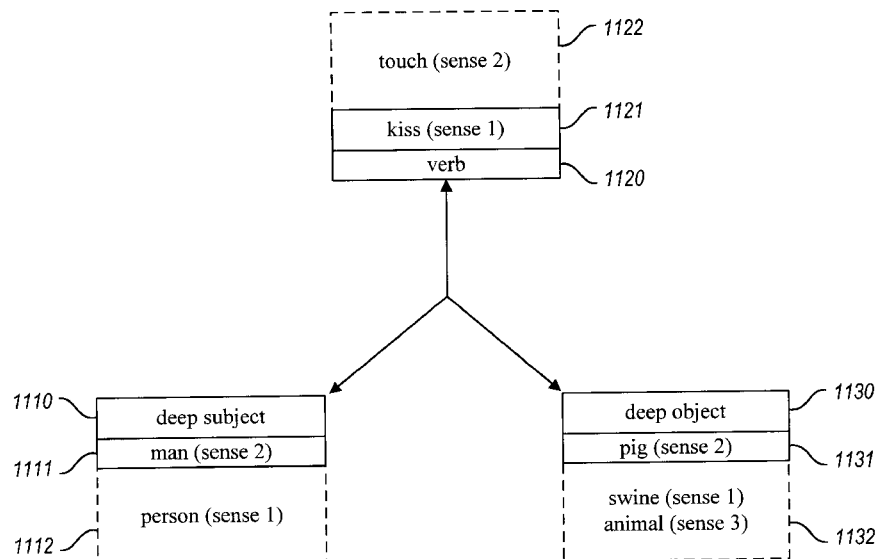
U.S. PATENT DOCUMENTS

| | | | |
|-----------|---------|----------------------|---------|
| 4,839,853 | 6/1989 | Deerwester et al. | 707/5 |
| 5,278,980 | 1/1994 | Pedersen et al. | 707/4 |
| 5,325,298 | 6/1994 | Gallant | 704/9 |
| 5,592,661 | 1/1997 | Eisenberg et al. | 707/104 |
| 5,619,709 | 4/1997 | Caid et al. | 707/532 |
| 5,675,745 | 10/1997 | Oku et al. | 705/7 |
| 5,675,819 | 10/1997 | Schuetze | 704/10 |
| 5,794,050 | 8/1998 | Dahlgren et al. | 395/708 |
| 5,794,178 | 8/1998 | Caid et al. | 704/9 |
| 5,799,308 | 8/1998 | Dixon | 707/100 |
| 5,873,056 | 2/1999 | Liddy et al. | 704/9 |
| 5,933,833 | 8/1999 | Braden-Harder et al. | 707/5 |

FOREIGN PATENT DOCUMENTS

| | | |
|--------------|---------|----------------------|
| 0 304 191 A2 | 2/1989 | European Pat. Off. . |
| 0 304 191 A3 | 2/1989 | European Pat. Off. . |
| 0 386 825 A1 | 9/1990 | European Pat. Off. . |
| 0 687 987 A1 | 12/1995 | European Pat. Off. . |

23 Claims, 18 Drawing Sheets



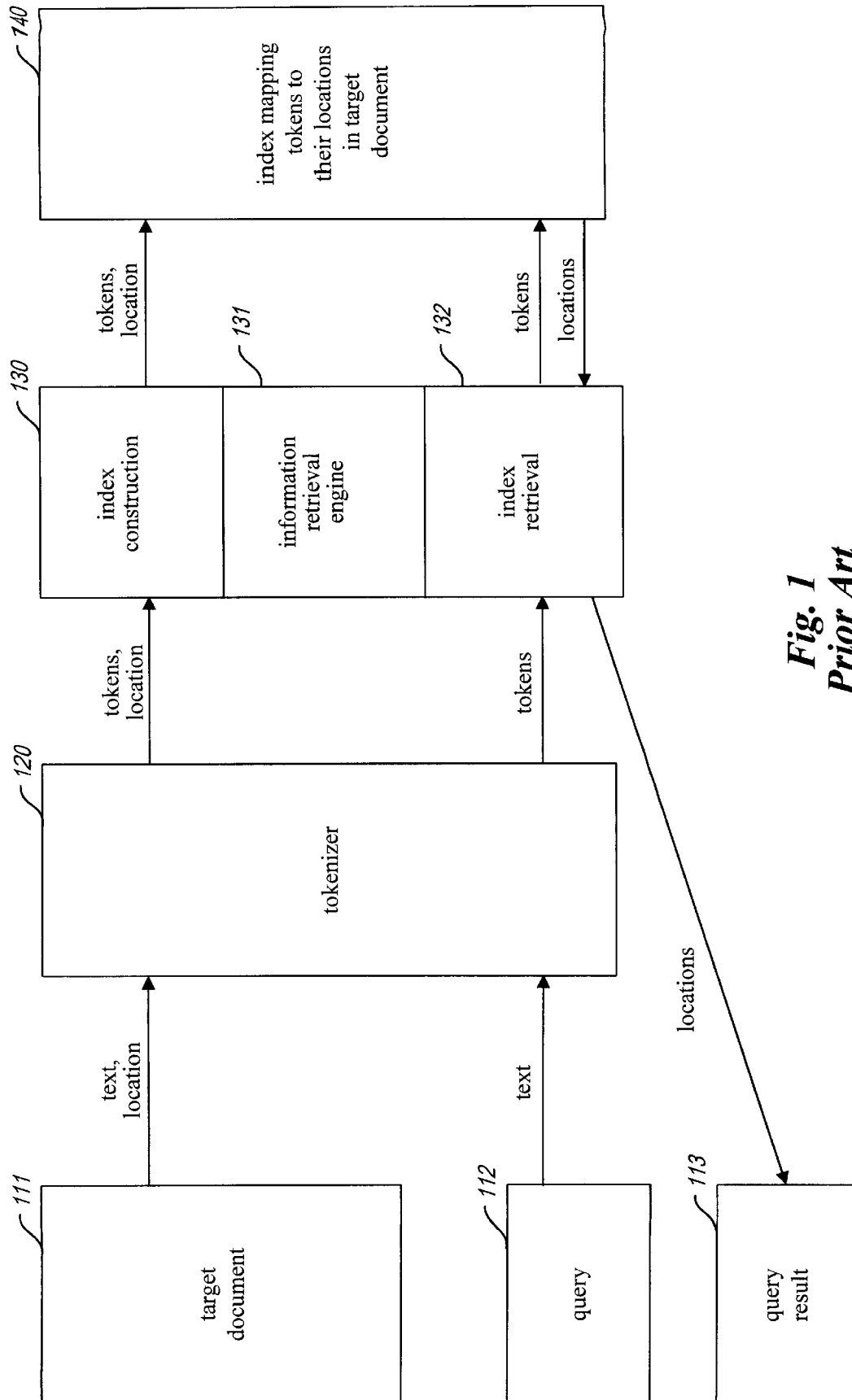


Fig. 1
Prior Art

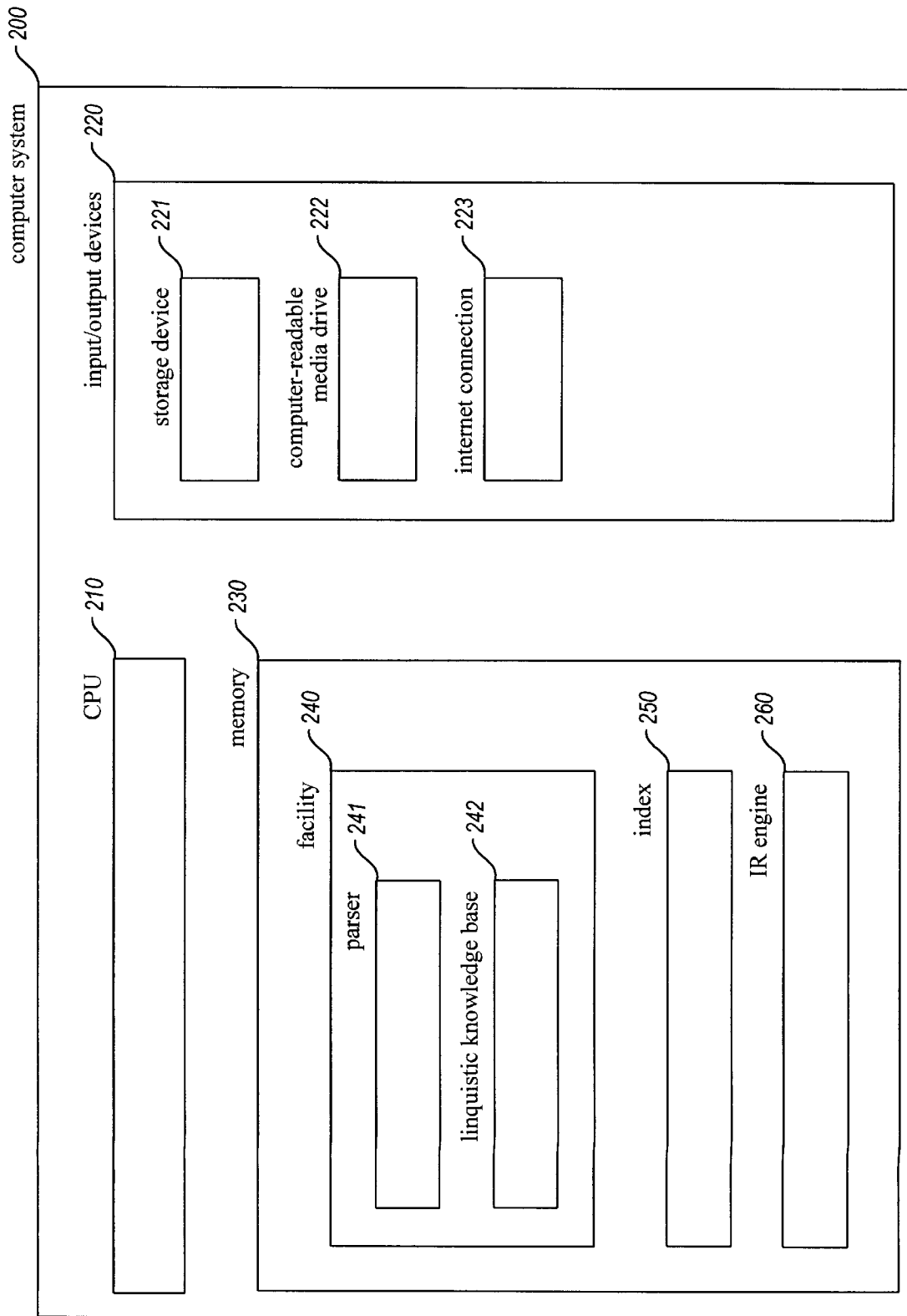


Fig. 2

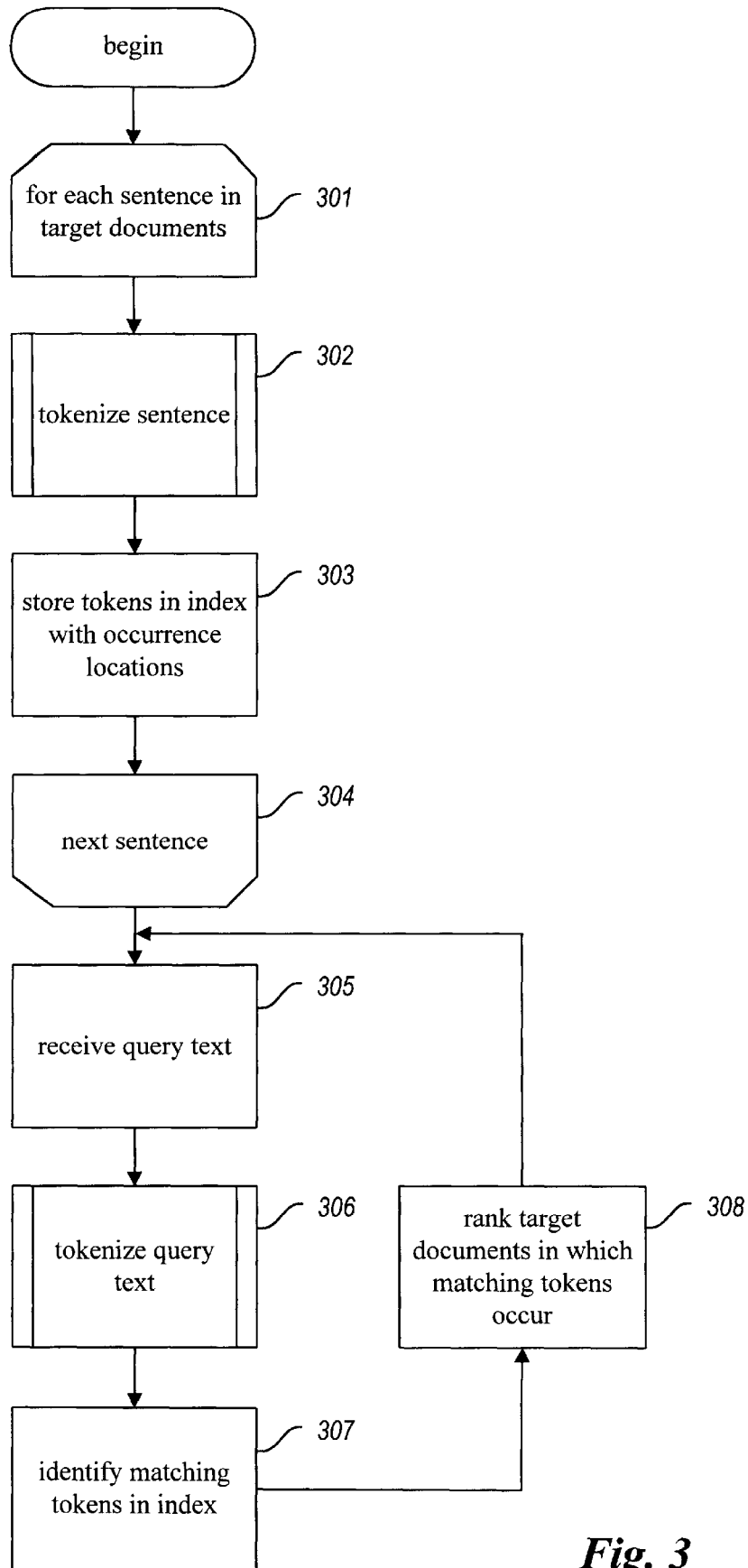


Fig. 3

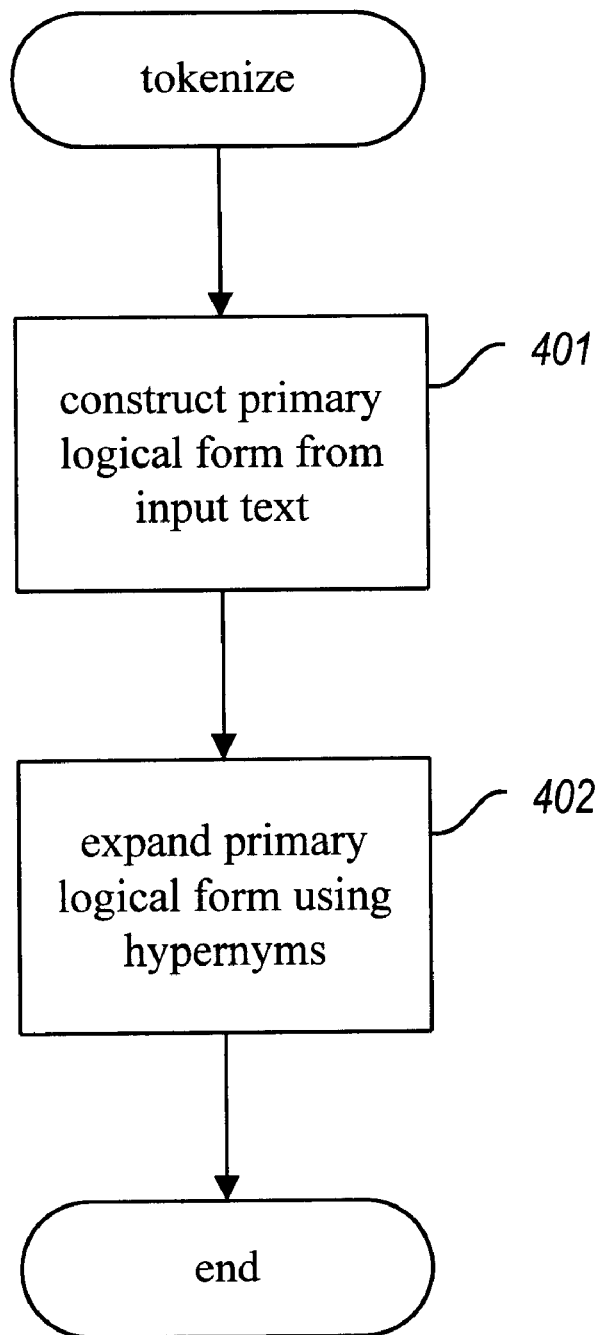


Fig. 4

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.