Published in Technology Hits

Yogesh Haribhau Kulkarni

Jan 15 · 1 min read · ▶Listen

Save

Representing documents using document keys

A US patent

Documents are made up of paragraphs, which in-turn are made up of sentences, which in-turn are made up of words. That's the usual representation of documents.

More abstract way could be that documents are made of themes or concepts or topics, such that a document is, say, 40% politics, 30% sports and 30% entertainment.

Following patent is about representing a contract document by a sequence of clause categories, say ['preamble', 'recitals', 'term',….,'signature']. Almost like a genome sequence, only difference is instead of letters A, T, C, G, the contract document is made up of sequence of clause categories-symbols.

US011361034B1

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

(10) **Patent No.:**   **US 11,361,034 B1**
(45) **Date of Patent:**   **Jun. 14, 2022**

(54) **REPRESENTING DOCUMENTS USING DOCUMENT KEYS**

(71) Applicant: **Icertis, Inc.**, Bellevue, WA (US)

(72) Inventors: **Yogesh Haribhau Kulkarni**, Pune (IN); **Sunu Engineer**, Pune (IN); **Amitabh Jain**, Pune (IN); **Ravi Kothari**, New Delhi (IN); **Monish Mangalkumar Darda**, Pune (IN)

(73) Assignee: **Icertis, Inc.**, Bellevue, WA (US)

(*) 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.: **17/538,852**

(22) Filed: **Nov. 30, 2021**

(51) **Int. Cl.**
G06F 16/93   (2019.01)

(52) **U.S. Cl.**
CPC **G06F 16/93** (2019.01)

(58) **Field of Classification Search**
CPC G06F 16/93; G06N 20/00
See application file for complete search history.

(56)   **References Cited**

U.S. PATENT DOCUMENTS

6,505,195 B1	1/2003	Ikeda et al.
7,519,589 B2	4/2009	Charnock et al.
7,617,164 B2	11/2009	Burges et al.
7,689,531 B1	3/2010	Diao et al.
8,411,935 B2	4/2013	Eaton et al.
8,429,103 B1	4/2013	Aradhye et al.
8,626,682 B2	1/2014	Malik et al.
8,868,472 B1	10/2014	Lin et al.
8,965,814 B1	2/2015	Rangan
9,235,812 B2	1/2016	Scholtes
9,384,450 B1	7/2016	Cordes et al.
9,922,285 B1	3/2018	Glode et al.
10,133,791 B1	11/2018	Chan
11,157,475 B1 *	10/2021	Cobb G06F 40/30

(Continued)

FOREIGN PATENT DOCUMENTS

EP	1946254 B1	2/2012
WO	2009061917 A1	5/2009
WO	2018042548 A1	3/2018

OTHER PUBLICATIONS

Office Communication for U.S. Appl. No. 15/949,905 dated Jun. 27, 2018, pp. 1-19.
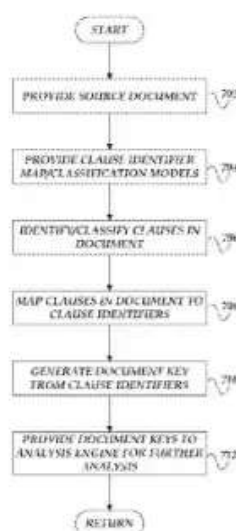
(Continued)

Primary Examiner — Hau H Hoang
(74) *Attorney, Agent, or Firm* — John W. Branch; Branch Partners PLLC

(57)   **ABSTRACT**

Embodiments are directed to representing documents using document keys. Documents that include one or more clauses may be provided. Each clause type for the one or more clauses in documents may be determined based on one or more classification models. One or more clause identifiers may be associated with the one or more clauses based on one or more clause types of each clause. A document key may be generated for each document based on an ordered collection of the one or more clauses included in each document such that each clause identifier may be positioned in the document key based on an order of its location in a corresponding clause of a document. The documents may be analyzed based on comparisons of one or more document keys corresponding to the documents. One or more reports may be generated based on one or more results of the analysis.

26 Claims, 9 Drawing Sheets

(Source: https://patents.justia.com/patent/11361034)

This terse representation is effective in various downstream machine learning based natural language processing workflows such as document similarity, contract type classification, etc.

Here is the full patent.

This info was originally published at LinkedIn

Patents   Naturallanguageprocessing   Machine Learning   Artificial Intelligence

Contracts