ARTIFICIAL INTELLIGENCE:
Thinking About Law, Law Practice, and Legal Education
Customized AI Techniques for the Patent Field

Dean Alderucci
Carnegie Mellon University
Center for AI & Patent Analysis
Overview

- Patents
- General-purpose AI & NLP
- The gap between AI & the legal field
- Bridging the gap: a framework
- CMU Center for AI & Patent Analysis
What is a Patent?

• A grant of legal rights
  • Right to exclude others from making, using the technology you invented

Also

• A document that describes:
  • the technology, and
  • what exactly others are legally excluded from making, using, or selling
What is a Patent?

**United States Patent**

**Tedesco et al.**

**Patent No.:** US 8,824,784 B2

**Date of Patent:** *Sep. 2, 2014*

**SYSTEM FOR IMAGE ANALYSIS IN A NETWORK THAT IS STRUCTURED WITH MULTIPLE LAYERS AND DIFFERENTIALLY WEIGHTED NEURONS**

**Applicant:** Facebook, Inc., Menlo Park, CA (US)

**Inventors:** Daniel E. Tedesco, Huntington, CT (US); James A. Jorasch, Stamford, CT (US); Geoffrey M. Gelman, Stamford, CT (US); Jay S. Walker, Ridgefield, CT (US); Stephen C. Tulley, Fairfield, CT (US); Vincent M. O’Neil, New York, NY (US); Dean P. Alderucci, Westport, CT (US)

**Assignee:** Facebook, Inc., Menlo Park, CA (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.

This patent is subject to a terminal disclaimer.

**Appl. No.:** 13/731,824

**Filed:** Dec. 31, 2012

**Prior Publication Data**


**Related U.S. Application Data**

Continuation of application No. 13/291,889, filed on Nov. 8, 2011, now Pat. No. 8,456,926, which is a

**References Cited**

U.S. PATENT DOCUMENTS


(Continued)

**Primary Examiner — Gregory M Desire**

(74) **Attorney, Agent, or Firm — Baker Botts L.L.P.**

**ABSTRACT**

Disclosed herein are systems and methods for facilitating the usage of an online workforce to remotely monitor security-sensitive sites and report potential security breaches. In some embodiments, cameras are configured to monitor critical civilian infrastructure, such as water supplies and nuclear reactors. The cameras are operatively connected to a central computer or series of computers, and images captured by the cameras are transmitted to the central computer. After initial training, a computer using the Supervised Learning Algorithm to process the images. This algorithm enables the computer to learn a computational model of the security situation at the site.

(Continued)
1. A method of generating test cases for a text annotator which searches text documents and analyzes them relative to a defined set of tags comprising:

- receiving a corpus of text fragments without any annotations and a description of the text annotator, by executing first instructions in a computer system;

- determining types of inputs to the text annotator from the description, the types of inputs including at least one phrase selected from the group consisting of a person phrase, a date phrase, and a diagnosis phrase, by executing second instructions in the computer system;

- analyzing language structures in the corpus to identify sentence types and grammar constructs, the sentence types including at least one sentence selected from the group consisting of a question, a command, a compound sentence, and a conditional sentence, and wherein said analyzing includes performing a slot grammar parse of the corpus to determine various parse trees of the corpus including a most common parse tree, by executing third instructions in the computer system;

- generating a first test case by performing a grammar tree transformation on a first selected fragment of the corpus based on the sentence types and the grammar constructs wherein the first selected fragment is selected in response to a selection bias towards a sentence type which corresponds to the most common parse tree of the corpus, by executing fourth instructions in the computer system; and

- generating a second test case by replacing at least one starting phrase in the first test case with a substitute phrase from at least one dictionary associated with one of the types of inputs that corresponds to the starting phrase, by executing fifth instructions in the computer system.
What is a Patent?

- The patent is a legal document:
  - Legal doctrines dictate:
    - How the patent is *interpreted*
    - *What exactly* others are excluded from making, using
    - Whether the patent satisfies *all legal requirements* for patenting
What is a Patent?

• Since the patent is a legal document:
  • Patent text encodes the attorney's legal decisions and legal strategies
  • Patent text contains information relevant to various legal determinations
Patent Analysis

- Attorneys and others perform *legal analysis* using the text of patents
  - Does a competitor’s patent cover my company’s product?
  - Does my patent cover a competitor’s product?
  - Can a competitor’s patent be overturned in litigation?
  - Is this patent worth buying?
• **Artificial Intelligence**
  - Software that mimics cognitive functions

• **Natural Language Processing**
  - A subfield of Artificial Intelligence
  - Allow computers to process “natural languages” such as English or Spanish
• **Natural Language Processing**
  • Apple Siri understands spoken commands
  • Google search answers typed questions
Many general-purpose NLP techniques
- Work for any types of text
- Not specific to a domain
- Can be applied to legal documents, patents
AI & NLP

• Many general-purpose NLP techniques
  • “Word vectors”
    • Automatically identify words that are similar or related
    • “negligence”, “duty”, “breach”
• Many general-purpose NLP techniques
  • "Topic Modeling" / "LDA"
  • Automatically group similar documents

Source: Shuai’s AI & data blog
The Gap Between AI & Law

- **General-purpose NLP techniques**
  - Primarily statistical:
    - Uses word frequency and correlation
  - Cannot:
    - “understand” text
    - utilize “common sense”
    - manipulation complex concepts
The Gap Between AI & Law

- **General-purpose NLP techniques**
  - A poor fit for higher-level cognitive tasks
    - e.g., legal decision making
  - Without understanding text, cannot perform legal analysis on that text
• Domain-specific NLP techniques
  • Customized for the text of patents
  • Design software that:
    1. recognizes *text patterns that patent attorneys use*
    2. *connects* those patterns to rudimentary legal analysis
1. Software that recognizes *text patterns that patent attorneys use*
   - Patents have a special structure
   - Patent attorneys use special phrasing / grammar for specific legal goals
1. **Software that recognizes *text patterns that patent attorneys use***
   - If we know *why* attorneys choose particular word patterns
   - then we can tell software *how to* “*understand*” patents
     - Extract small fragments of legal information from patent text
2. **Connect text patterns to legal analysis**
   - How do *courts use these patterns* when interpreting patents?
   - i.e. how are these patterns of text used in legal analysis?
2. How do courts use these patterns when interpreting patents?
   - Need to analyze numerous opinions to determine how text patterns affect legal analysis
Bridging the Gap

• Design software that:
  1. recognizes *text patterns that patent attorneys use*
  2. *connects* those patterns to rudimentary legal analysis

• Both *require* legal experts
• Design software and algorithms customized for the patent field
• Leverage patent structure and knowledge of patent drafting
• Provide tools for different patent tasks
• Tool Category #1
  • Automatically identify, aggregate, and display relevant information to the legal decision maker
  • Software is faster than the attorney searching and aggregating this information
Tool Category #2

- Automatically “score” legal issues
- Count how many pieces of information are in favor of a proposition, and how many are against that proposition

- Weighted, unweighted scores:
  - number for – number against
Example: Analyzing Patent Indefiniteness

- A patent claim must be “definite”
  - i.e. must not be *ambiguous*
• Supreme Court standard:
  • “does the text convey, to the person of ordinary skill in this technical field, a meaning with reasonable certainty?”

• Can software predict how a person would understand certain technical text?
• Potentially relevant pieces of information for indefiniteness:
  1. Are the terms defined?
  2. If not defined, *should* they be defined or are they instead well known?
  3. Are there *inherently ambiguous* terms?
     • e.g., “big”, “fast”, “not unduly difficult”
Example scoring for indefiniteness

- *Definiteness score: 2 out of 10*
  - Claim has 4 undefined terms
    - Of these, 2 appear to be “coined”, and so must be defined
    - The other 2 terms are defined in many other patents
  - Claim includes 1 potentially ambiguous term “heavy”

- Could score fifty thousand patents
Example: Smart Quantity Search

• “Find claims reciting 3 – 8 grams of any hydrocarbon”
  • e.g., “... 2500 mg of a cycloalkane ...”
  • e.g., “... 0.2 – 0.25 ounces of an arene ...”
“Find claims where a means plus function limitation doesn’t appear to have support in the specification”
  * e.g., “... a synthesizing means for synthesizing a hydrocarbon...”
  * “The spec doesn’t appear to disclose ways to synthesize hydrocarbons”
  * “However, the spec appears to disclose synthesis of cycloalkanes”
Example: Patent Law Concept Search

• “Find claims where >3 claim terms are not defined in the specification”
• Legal NLP can leverage the special structure of legal text
• The attorney has a critical role in the design of domain-specific NLP tools