Automatic Information Retrieval from Patents for Design Innovation

REU Student: Leo Xuzhang Lin
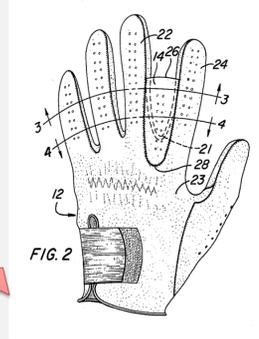
Mentor: Dr. Anthony McCaffrey (Post-Doc)   Faculty Advisors: Dr. Sundar Krishnamurty, Dr. Ian Grosse

Problem

1. Confusing classification in the patent database.

Current U.S. Class:	2/161.4 ; 2/161.5; 2/21; 473/205
Current International Class:	A63B 71/14 (20060101); A63B 71/08 (20060101); A41D 019/00 ()
Current CPC Class:	A63B 71/146 (20130101)
Field of Search:	2/16,20,21,159,161R,161A,163 294/25

US 3930271, the patent for
Golf Gloves
(from uspto.gov)

?

FIG. 2

3. Number of patents?

MORE THAN 8 MILLION !!!

2. No meta-data representing the content of the patent.

Function? Method?

You know nothing unless you've read it!

?

Solution

1. Based on extensive prior work, defining an engineering design's function by energy flow and function.

What does it do?
Described using verbs that describe its function

+

How it does it?
Defined by the type of energy it uses

2. Need for a new method of information retrieval to **AUTOMATICALLY** gain some useful insight about a patent's content.
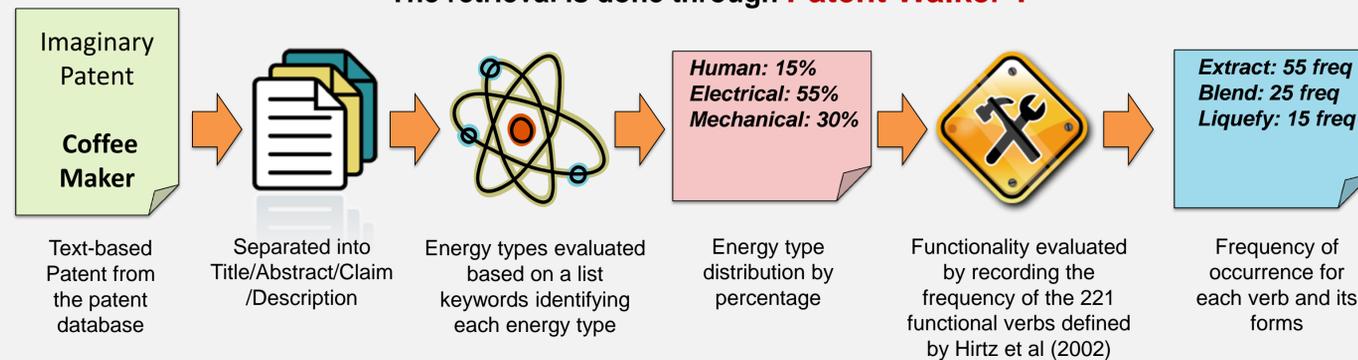
Parse → **Evaluate** → **Aha!**

Abstract

❖ Patent classifications are complex and do not accurately represent the information content of the patent.

❖ Searching for a particular patent is inefficient and inaccurate

❖ There had been extensive previous works by Hirtz et al that defined engineering design's function by its energy flow and set of verbs describing the function.

❖ My research focuses on researching and creating an information retrieval algorithm that can determine a patent's energy distribution and identify its functional verbs.

❖ The extracted information will be used to supplement the **Analogy Finder*** search engine that searches for analogous patents by its goal and synonyms of that goal.

❖ Improving patent search will improve problem solving by finding accurate solutions residing in patents.

❖ Future complex analysis such as ordering by uniqueness can be done on the data set.

* http://demo.innovationaccelerator.com/

Methods/Purpose

The retrieval is done through Patent Walker*:

Imaginary Patent **Coffee Maker**		**Human: 15% Electrical: 55% Mechanical: 30%**		**Extract: 55 freq Blend: 25 freq Liquefy: 15 freq**
Text-based Patent from the patent database	Separated into Title/Abstract/Claim /Description	Energy types evaluated based on a list keywords identifying each energy type	Energy type distribution by percentage	Functionality evaluated by recording the frequency of the 221 functional verbs defined by Hirtz et al (2002)	Frequency of occurrence for each verb and its forms

Ask for a demo!

Purpose? (Why should I care)

1. Patent database is a good resource for existing solutions to be used for problem solving (If you can find them)

2. Possible application in other technical documents to gain valuable data about its contents.

3. When innovating a new design, it is beneficial to search for similar designs to ensure the novelty of the new design.

* A software package I developed.

Future Work

❖ Improving accuracy through natural language processing and linguistics analysis of causality.

❖ Refine technique to calculate semantic distance between two patents

❖ Integration and application in Analogy Finder

Acknowledgements

BYU BRIGHAM YOUNG UNIVERSITY   IOWA STATE UNIVERSITY   Oregon State UNIVERSITY OSU   University at Buffalo The State University of New York   UMASS AMHERST   VirginiaTech Invent the Future®   WAYNE STATE