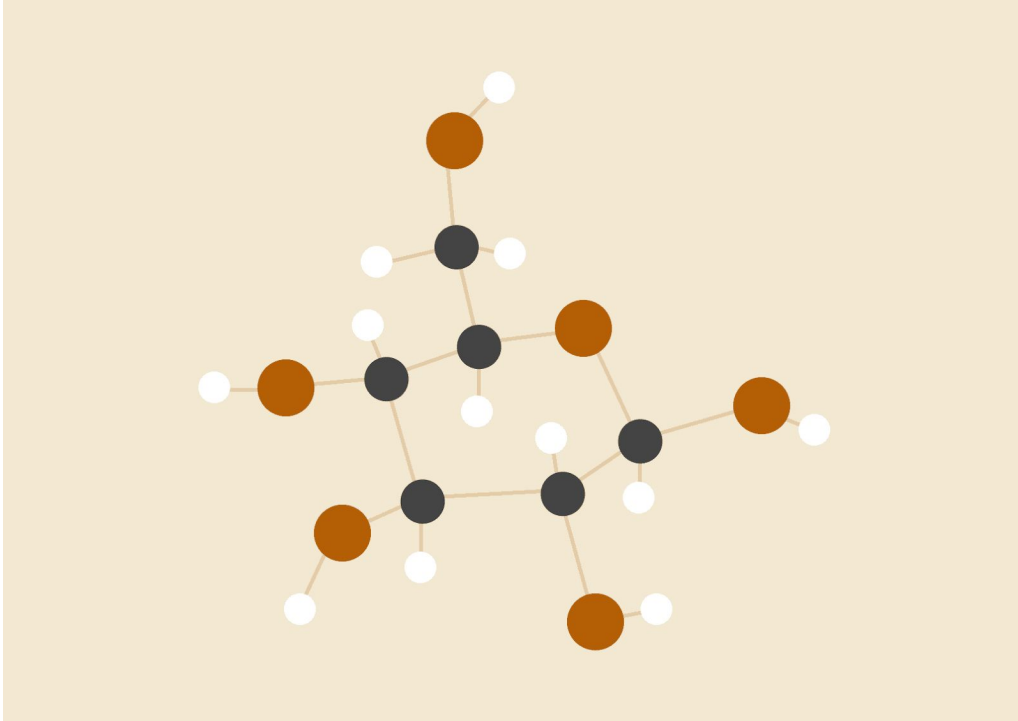


PATENTCONNECT PROPOSAL

Empowering Developers, Designers, Hobbyists, & Alike



DEVELOPERS:

Yousef Hashemi <alhashey@mcmaster.ca>

Jatin Chowdhary <chowdhaj@mcmaster.ca>

Steven Gonder <gondes1@mcmaster.ca>

Yiding Li <liy443@mcmaster.ca>

Varun Verma <vermav6@mcmaster.ca>

LAB: 2 → GROUP: 5

The developers of PatentConnect electronically sign
& date that the work being done is our strictly our own

ABSTRACT

The world of intellectual property (IP) does not have a free and open source solution to searching patents and their citations. PatentConnect aspires to present a visually appealing and accessible representation of the connections between patents. Users can search for key terms to find collections of patents. In addition, this will provide insight to potential inventors about the process of creating patents. This is based on the Patent Citation Network, which shows pairs of patents where the first cites the second. We can verify the output of our program using the dates of the patents to ensure against impossible patent relations.

OBJECTIVE

PatentConnect empowers developers, designers, hobbyists, enthusiasts, and anyone interested in finding inspiration in patents.

MOTIVATION

1. What Is The Problem That Motivates You To Build This Product?

Our motivation for building PatentConnect is to empower potential inventors. PatentConnect will provide designers and developers with great detail on the patent system.

2. Why Is The Problem Important To Be Addressed

Allowing potential inventors to easily see the connections between patents will further enable the creation of innovative patents. This will help advance any field of technology with a pre-existing trove of patents.

3. Who Are The Users Of This Product?

PatentConnect will empower inventors, designers, developers, and enthusiasts. In addition, it will allow anyone and everyone to gain more insight into the patent system.

4. What Are The Expected Outcome(s) & Utility(ies) Of The Product

PatentConnect aims to empower inventors and tech enthusiasts by allowing them deeper insight into the patent system's citation hierarchy and interconnectivity. Inventors and designers will be able to visually see the connections between patents and the process. In addition:

- ☐ Inventors can draw inspirations from existing patents, and explore how ideas behind those patents came to be.
- ☐ Inventors can make sure that their ideas/patents do not infringe on other patents.
- ☐ Developers can easily discover patents that match their objectives.
- ☐ Customers can easily understand the progression of patents and how they came to be from other patents.

PRIOR WORK

Patent landscaping is defined as “exercise of identifying all the patents that are relevant to a topic” (Abood & Feltenberger, 2018). A model for effective patent landscaping was proposed in a Google paper by Aaron Abood and Dave Feltenberger; the paper presented an approach called Automated Patent Landscaping (Abood & Feltenberger, 2018). Our app's specification will be requiring ideas similar to Google's APL approach in that it largely automates searching patent metadata and patent IDs, but the most major difference is that we aren't currently using machine learning as an aid for interpreting

the patent landscape. Furthermore, our specification differs in that we don't need to consider or implement class codes, seeds, anti-seeds, or family IDs.

There is also a practical similarity with Teqmine. Teqmine uses AI for patent searching; however, their service costs roughly €79 monthly ("Patent Similarity", 2019). We aim for our service to be free; our app will be economical, easy-to-maintain, open-source, and will hopefully democratize patent searching. We are giving the general public a more accessible alternative. One similarity with Teqmine is that we will be using a similar hierarchy system for graphical and data representation.

I.O. & SOLUTION

- 1. Include the name and URLs of the dataset that you will be using (from Canada or US government open datasets or other sources). Add one sentence for each dataset describing why that dataset is needed.**

The first open dataset PatentConnect will use is the [Patent citation network](#). It is maintained by the National Bureau of Economic Research. This dataset is required because it contains all utility patents granted from 1963 to 1999; a period of 37 years (Leskovec, n.d.). It contains 16,522,438 citations. The dataset has 3,774,768 nodes and 16,518,948 edges (Leskovec, n.d.).

The second open dataset PatentConnect will use is [Google Patents](#). It will be used to retrieve information about the patents from the first dataset. Google Patents has an API we can leverage that will allow us to quickly search and retrieve information about a particular patent.

- 2. Describe the outputs that your product will be generating in terms of the data and not the interface (e.g., Distance between point A and B, number of cured patients, etc.). For each output add one sentence description of the output.***

PatentConnect will generate collections of related patents based on queries.

PatentConnect will generate the shortest path that connects patents based on citations; for instance, when two utility patents about Bluetooth pairing are provided, PatentConnect will connect them using citations in the shortest possible path.

PatentConnect will generate graphs showing the relations between patents in a collection.

- 3. Briefly describe the proposed solution in terms of how the input will be used to generate the output. You can describe this part as a usage scenario. You may add a figure if it helps to describe how your proposed solution is working. (max 500 words)***

PatentConnect will work with the dataset in the following manner:

1. The patent dataset will be searched and sanitized for errors and malformed data/input.
2. The dataset will be grouped together based on citations..
3. A searchable database containing the sanitized and grouped dataset will be created. This database will be the backbone of PatentConnect. Upon a successful patent query, a list of citations will be returned. Each citation will be hyperlinked and lead to its own information page.
4. The information page will contain an option to show more information. This will fetch information from Google Patent's database and return it.
5. The bottom of the information page will have a textbox where a patent ID can be entered. Then, it will find the shortest path from one patent to the other using citations. The citations will connect the patents together.

ALGORITHMIC CHALLENGES

The program will use union-find to put related patents into collections. These collections would then be put into a database, made of an efficient binary search tree, such as a red-black tree. The red-black tree would use keys that could be searched for by keyword. This may change as we understand more about the challenges of searching. Each of the collections would be sorted by date of patent upon access. In order to get the data surrounding the patent, the Google Patents API would be used to construct the database. As this would likely take a long time for many entries, the constructed database would be saved and could be restored upon restarting the program. Some challenges to this design would be searching by keywords and accessing patent data using Google API. Some research may need to be done on the Patent API to see how we can ensure we are retrieving the correct patent based on patent ID. Formatting the keys to the search tree will also be a difficult task and may need to be changed, and the database type may need to be optimized for this type of work.

PROJECT PLAN

Milestones	Deliverable	Date
Patent Collections	Class for a collection of patents	Week of February 21
Database Backend	A searchable database containing the collections	Week of February 28
Requirements Specification	Document based on CS 2ME3 instructions	Week of March 7
Project Progress Check	A rough project, with functional back end searching and a database feature	Week of March 9
Sorting Collections	A functional back end searching mechanism, and a database feature that can return a sorted collection of patents	Week of March 27
GUI Finish	A graphical representation of the sorted collections, and a graphical means of searching	Week of April 6
Final Project Presentation	Full demo of software	April 12

REFERENCES

- Abood, A., & Feltenberger, D. (2018). Automated patent landscaping. *Artificial Intelligence and Law*, 26(2), 103–125. doi: 10.1007/s10506-018-9222-4
- Leskovec, J. (n.d.). Patent citation network. Retrieved February 7, 2020, from https://snap.stanford.edu/data/cit-Patents.html?fbclid=IwAR2dhge7ee45EauA4rB_gdif98CDmNXvnGaCWJRLM8was6ffLhLRLktRRPk
- Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005, August 21). Graphs over time: Densification Laws, Shrinking Diameters ... Retrieved February 7, 2020, from <http://www.cs.cornell.edu/home/kleinber/kdd05-time.pdf>
- PatentInspiration. (2020). Retrieved February 7, 2020, from <https://www.patentinspiration.com/blog/Crowdsource-The-World-Open-Innovation-Through-Patent-Inpiration?fbclid=IwAR0XRhQqDkpz-3NHSt6rsJ33fB7duX5z6WZ6bkvv9d-GeIAO83S9MxI11-s>
- Toivanen, H., & Suominen, A. (2019, September 25). Patent Similarity. Retrieved February 7, 2020, from <https://teqmine.com/patent-similarity/>

[Fork Us On GitLab](#)