



DEPARTMENT OF COMPUTER SCIENCE

Natural Language Processing in the Law Domain

Oliver Ryan-George

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Bachelor of Science in the Faculty of Engineering.

Tuesday 5th February, 2019

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of BSc in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

Oliver Ryan-George, Tuesday 5th February, 2019

Acknowledgements

Executive Summary

This project is a contribution to a law academics research into the growing relationship between intellectual property law and human rights law; in particular, the extent to which intellectual property laws involve human rights considerations, and their balance between consideration of creators and users. The first technical objective for the project is to use natural language processing to estimate measures of these two aspects for input intellectual property statutes. This will involve using Python to explore a variety of methods, such as k-means clustering and maximum likelihood, to optimise results. The second is to create the interface for this to allow the input of new data and visualisation of the results.

Currently, there is limited means for doing this type of analysis because it has to be done manually, making it time consuming to analyse significant amounts of data. The project will change this by allowing the user to automatically process legal documents as they become available.

The project is likely to be successful because there have been projects previously successfully completed that involved natural language processing on legal documents for other analysis. It will, however, be made individual from many of these by my iterative approach to the project, collaborating with a law academic to find the most appropriate way to classify and visualise the results.

Should it be successful, the project will further understanding of the legal impact of intellectual property laws.

Contents

1	Background	1
1.1	Introduction	1
1.2	Literature Review	2
1.3	Preliminary Investigation	2
1.4	Project Plan	2

Chapter 1

Background

1.1 Introduction

1.1.1 The Problem

Intellectual Property law is a term typically used to describe the areas of law which establish property protection over intangibles such as ideas, signs and information[1]. This protection is a necessity as it makes the advancement of ideas profitable which therefore incentivises this act[1].

Intellectual property has become a problem as it is being increasingly disassociated from creators to benefit large corporations[2]. One example of this is the expansion of copyright terms such as the controversial Copyright Term Extension Act of 1998 which was heavily lobbied for by Disney just years before Mickey Mouse's copyright ran out[3]. The trend is illustrated by Figure 1.1.

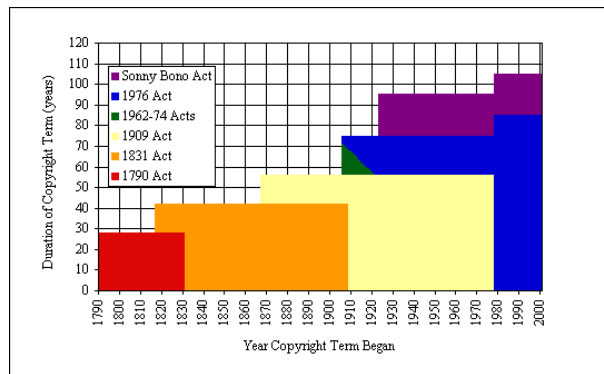


Figure 1.1: Expansion of U.S. copyright term lengths[4].

This implementation of intellectual property law goes against the human rights of access to culture, education and other social and economic rights but is not widely considered so because human rights and intellectual property are regarded as distinct areas of law. By adapting intellectual property law to incorporate human rights, intellectual property law would be reprioritised to promote cultural and scientific progress rather than just reward right owners[2].

To support this, Dr. Megan Rae Blakely of Lancaster University is looking for assistance in analysing journals and legal instruments for overlap in language. Previously, analysis of the intersection between human rights and intellectual property law has been limited to more manual case-by-case methods. Supplying a more systematic method using natural language processing that can cope with large amounts of data would give concrete evidence of the relationship between human rights and intellectual proeprty.

1.1.2 Aims and Objectives

In the early stages of the project, the following requirements for the project were established:

- A simple Graphical User Interface which allows for input of law journals and treaties PDF form.

- A visualisation of all inputted PDF documents with x -axis as time; y -axis as the intent of the document toward human rights to intellectual property; and z -axis as the measure of the extent to which statute protects the owner or the user.
- A code base written well enough for any future researcher to easily understand.

Over the course of the literature review, section 1.2, I will review past literature in order to find the best methods to achieve each of these requirements.

1.1.3 Added Value

The originality of the project stems from its application of natural language processing methods in this domain, rather than the natural language processing methods used, as will be discussed in section 1.2. Therefore, the added value will come in optimising the techniques used for the application, for example, finding the most appropriate visualisation to illustrate the findings.

1.1.4 Scope

Abc

1.1.5 Structure of Literature Review

Abc

1.2 Literature Review

1.2.1 Definitions

Abc

1.2.2 Natural Language Processing Techniques

Abc

1.2.3 Graphical Representation Techniques

Abc

1.2.4 User Interface

Abc

1.3 Preliminary Investigation

1.3.1 Preliminary Investigation

Abc

1.4 Project Plan

1.4.1 Timeline

Abc

1.4.2 Evaluation

Abc

References

- [1] L. Bently and B. Sherman, *Intellectual Property Law*, 4th ed. OUP Oxford, 2014, ISBN: 978-0199645558.
- [2] C. Geiger, *Research Handbook on Human Rights and Intellectual Property*. Edward Elgar Publishing, 2016, ISBN: 978-1786433411.
- [3] V. A. Grzelak, “Mickey Mouse & Sonny Bono Go To Court: The Copyright Term Extension Act and Its Effect On Current and Future Rights,” *John Marshall Review of Intellectual Property Law*, vol. 2, no. 1, pp. 95–115, 2002.
- [4] T. Bell, *Trend of maximum u.s. general copyright term*, tomwbell.com, no changes made, 2008.