



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.

---

Thursday 7<sup>th</sup> 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, Thursday 7<sup>th</sup> February, 2019

---

## Acknowledgements

---

## Executive Summary

This project is a contribution to a law academic's 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 law journals and statutes. This will involve using Python to explore a variety of machine learning models, starting with support vector machines, to optimise results. The second is to collaborate with the law academic iteratively to find the most appropriate way to visualise the results for the project's domain.

Currently, the relationship between intellectual property and human rights is analysed on a case-by-case basis due to a lack of systematic means of analysis. The project will allow large amounts of evidence.

The project is likely to be successful because there have been previously successfully projects 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 significant past events on intellectual property laws.



---

# Contents

<b>1</b>	<b>Background</b>	<b>1</b>
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. This protection is in order to make the advancement of ideas profitable which therefore incentivises this act[1].

There is, therefore, a balance to be struck between limited exclusive rights and benefits to the public. While limiting exclusive rights may facilitate progress, benefiting the public, the overprotection of the exclusive property may restrict their access[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 in extension of copyright terms, illustrated by Figure 1.1, illustrates the appearance that the balance is tipping towards exclusivity rights.

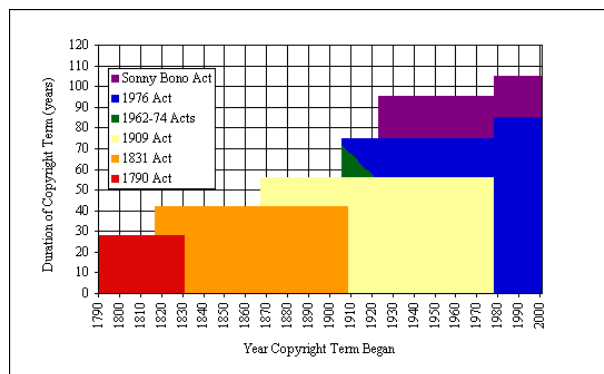


Figure 1.1: Expansion of U.S. copyright term lengths[4].

This implementation of intellectual property law brings in the question of human rights of access to culture, education and other social and economic rights, but traditionally this has not been included in discussion of intellectual property law[5]. However, recently scholarship and legislation has progressively begun to incorporate each other's linguistics[6].

Dr. Megan Rae Blakely of Lancaster University is looking for assistance in analysing journals and legal instruments for overlap in languages. In this analysis, I will map the dynamics of this change in legal and social perspective, to make evident moments of significant shifts in language.

Previously, analysis of the intersection between human rights and intellectual property law, such as Helfer's[7], 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 property.

### 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 the document suggests the protection of the creator/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. The added value will be the unique way in which the findings are best illustrated for this application through the visualisation.

The outcome of the project will help add value to its domain. As the first systematic, easily scalable technology in this domain, it will help illustrate how historical changes in technology have impacted the tone of intellectual law. Blakely points out that this, in turn, will allow legal professionals to consider how future technological changes will impact their work and adapt accordingly.

## **1.2 Literature Review**

### **1.2.1 Natural Language Processing**

Abc

### **1.2.2 Graphical Representation**

Abc

### **1.2.3 User Interface**

Abc

### **1.2.4 Documentation**

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, 2008.
- [5] L. R. Helfer, *Mapping the Interface Between Human Rights and Intellectual Property*. Cambridge University Press, 2011, ISBN: 978-0521711258.
- [6] M. R. Blakely, *BILETA 2019 Proposal*, Not published - available if upon request, 2019.
- [7] L. R. Helfer, “Human rights and intellectual property: Conflict or coexistence?” *Minnesota Intellectual Property Review*, vol. 5, no. 1, pp. 47–62, 2003.