



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.

Monday 29th April, 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, Monday 29th April, 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. Classifications will be evaluated using F1-scores and compared against current research's achievements.

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 using 'matplotlib' for plotting and tKinter for the user interface. Success in this objective will be based upon typical human computer interaction practices.

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 give that systematic method and therefore allow a more comprehensive argument to be made about the relationship.

The project is likely to be successful because there have previously been successfully projects that involved similar topic classification tasks. 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 how past events have impacted intellectual property laws.

Contents

1	Introduction	1
1.1	Chapter Overview	1
1.2	Project Context	1
1.3	Requirements	2
1.4	Added Value	3

Chapter 1

Introduction

1.1 Chapter Overview

To open the thesis, this chapter gives a brief background of why the project is being pursued in the law domain. It continues by explaining the projects aims with regards to the given background and then finishes with how the aims intended to add value to the law domain once they had been met.

1.2 Project Context

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.

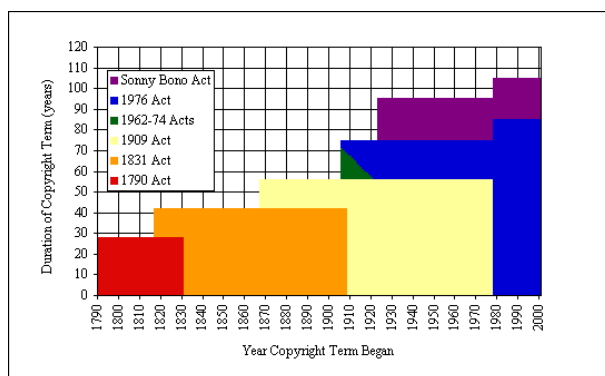


Figure 1.1: Expansion of U.S. copyright term lengths[4].

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.

1.3 Requirements

In the early stages of the project, I worked with Dr. Blakely to establish a set of requirements for the project. These requirements can be split into three categories: natural language processing, visualisation, and usability.

Each of the following subsections discusses the requirements of a different category. Each subsection starts by specifying the deliverable elements for the category, followed by high-level criteria of how they will be evaluated. All goals needed to be completed by 16th April 2019, when Dr. Blakely presents to the British and Irish Law Education and Technology Association (BILETA) Conference.

1.3.1 Natural Language Processing

The natural language processing in the project will take the form of two models. Each model will take the same corpus of training PDF documents as input but will classify different characteristics of the corpus.

- The first model will use topic classification to identify whether a document is related more to human rights or intellectual property;
- The second model will use sentiment analysis to identify whether the document indicates that the current legal climate more strongly benefits the user of intellectual property of the creator of it.

It is notable that Dr. Blakely expressed that the former is the more important model in terms of the projects success as there is more scholarly work done on this relationship. The latter models relationship is less widely covered and was requested by Dr. Blakely as a personal preference.

These models will be successful if they meet the following criteria:

- The model accurately represents a large proportion of the documents in the corpus with regards to its classification characteristics;
- The model deduces any trends in the corpus with regards to its classification characteristics, if those trends exist;

The model must only consider the language of the content of the article.

1.3.2 Visualisation

The previously mentioned models will then have to be visualised. Dr. Blakely stated her belief that a modification of Ebochs tone graph[8] shown in Figure 1.2 would be a suitable visualisation for the models. This would involve an x-axis representing time, a y-axis representing human rights-intellectual property and a z-axis representing user-creator.

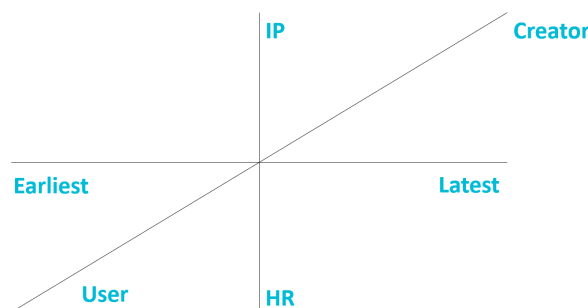


Figure 1.2: An adaptation of Ebochs tone graph for this domain.

The visualisation will be successful if it meets the following criteria:

- The visualisation and its axes meanings are self-explanatory;

- It is clear where a point lies on the visualisations axes;
- The visualisation has all the possible information that Dr. Blakely requires, on it;
- The visualisation clearly indicates any trends deduced by the models;
- The visualisation is aesthetically pleasing.

1.3.3 Usability

There will be two key types of stakeholders in this project. Both their circumstances will have to be considered in order to deliver a suitable final product.

The first type of stakeholder are the users of the projects end product. This will primarily be Dr. Blakely but is likely to extend to other law academics. It is important to note that these will likely be people with standard information technology skills and no computer science skills. Therefore, the end product will need to be in the form of a tool that is accessible to people of this calibre.

The tool must have the following features:

- Allow for input of new documents;
- Display visualisations based on models with given documents as input;
- Display information about success of model.

The tool will be successful if it meets the following criteria:

- The tool does not require any prior Computer Science knowledge to setup or use;
- The tool has a graphical user interface which is intuitive;

The second type of stakeholder is the future developers furthering Dr. Blakely's research. These developers will be experts in natural language processing and computer science and so will be able to both use and change the tool to their liking. However, changes to the tool will only be easily changeable if the codebase for it is well written.

The codebase will be successful if it meets the following criteria:

- It is easy to understand the purpose of each section of code;
- It is easy to expand upon and adapt the overall codebase.

1.4 Added Value

This project is original in its application of natural language processing methods to find the relationship between intellectual property and human rights language. The natural language processing methods used were not original, but the added value appears through the iterative process which found the most appropriate visualisation of results. The outcome of the project will help add value to its domain by illustrating how historical changes in technology have impacted the tone of intellectual property 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.

References

- [1] Lionel Bently, Bred Sherman, *Intellectual Property Law*, 4th ed. OUP Oxford, 2014, ISBN: 978-0199645558.
- [2] Christophe Geiger, *Research Handbook on Human Rights and Intellectual Property*. Edward Elgar Publishing, 2016, ISBN: 978-1786433411.
- [3] Victoria 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] Tom Bell, *Trend of Maximum U.S. General Copyright Term*, tomwbell.com, 2008.
- [5] Laurence R. Helfer, *Mapping the Interface Between Human Rights and Intellectual Property*. Cambridge University Press, 2011, ISBN: 978-0521711258.
- [6] Megan R. Blakely, *BILETA 2019 Abstract*, Not published - available upon request, 2019.
- [7] Laurence R. Helfer, “Human rights and intellectual property: Conflict or coexistence?” *Minnesota Intellectual Property Review*, vol. 5, no. 1, pp. 47–62, 2003.
- [8] Doug Eboch, *Some Thoughts About Tone*, letsschmooze.blogspot.com, 2015.