

A Linked Data Approach to Online Content Verification and Copyright Enforcement

Pei Zhang

Supervisors: Mr. Lester Gilbert, Dr. Sophie Stalla-Bourdillon

Abstract

The fast development of Internet and Web technologies enables broader and faster distribution and communication of copyright protected works.

- Online copyright infringement and piracy are major concerns for copyright owners, law makers and Internet service providers.
- Many technologies are developed and deployed in order to detect infringement, to track identification of legal usage of copyrighted works, and to stifle users' access to illegal content.
- Using Linked Data principle, to represent, connect, and reuse copyright relevant information from different sources in order to identify the lawfulness usage of online content, is a relatively new topic.

Introduction

- Many technologies have been developed and applied to detect infringing content and control users' access to illegal content.
 - DRM system
 - Content Identification Technology
 - Content Blocking technology
- Semantic Web and Linked Data technologies offer the ability to connect different datasets efficiently and to explore the data by machine in order to enhance information reuse, search, retrieval and management.
- This research will explore how to integrate Linked Data principles within the copyright domain in order to help determine the legality of online content, and how to build technical models and framework to refine copyright enforcement mechanisms.

Copyright issues on the Web

- Main infringement mechanisms on the Web
 - File sharing systems, particularly the use of P2P file sharing software
 - Traditional websites, that illegally offer movies, music, or e-books for download or for streaming
 - Web 2.0 dominant applications that allows UGC to widely sharedbo
- Internet intermediaries liabilities
 - Internet intermediaries, specifically, Web server host providers and content distributors are coming under increasing obligation duty to monitor the alleged illegal activities undertaken through their platform, although there is still a debate upon whether or in what scope Internet intermediaries should have such a duty imposed.
- American 'Notice and Take Down' procedure

Semantic Web and Linked Data in Copyright Domain

Linked Data can be considered as a lightweight Semantic Web

- Top-down vs bottom-up approach
- At current stage, many efforts have been made to ontologising general legal concept including Core Legal Ontology, LKIF-Core (Legal Knowledge Interchange Format) Ontology, Creative Common, Copyright Ontology etc.
- Linked Data focuses more on publishing semi-structured or structured data online without developing universally agreed ontologies, which lowers the technical barrier of applying Semantic Web to the data, so that data can be efficiently published and reused

Conclusion and Future Work

- Copyright of online content is a major concern for consumers, copyright owners, legal professionals and Web service providers. Currently, government, companies and web portals have published legal or copyright related data.
- By using Linked Data as a data substrate, we seek to find the ways to make it possible for distinct datasets to be meshed together in unified and machine-understandable ways, in order to develop at a second stage a new copyright infringement detection application that could be run by many different players.
- Future work
 - Ontologies to describe transparency report data and chilling effect data
 - Convert data into semantic format
 - A simple application to do preliminary test

Infrastructure

