

Auto linker of laws references in the network

This project has been created as part of the 'Digital Humanities' course, in collaboration with the Ministry of Justice of Israel.

The purpose of this project is to tag and mark the Israeli laws references all over the network. This way the details of all the laws will become accessible to every user which has an interest to explore and want to read more about a specific law.

During the course we have learned about some concepts in NLP and OCR, and we raised some questions about how the legislation can be converted to digital platform.

We had couple of lessons with two representees of the Ministry of Justice of Israel, who introduced us their work and an advanced processing of digitization of the legislation. In addition, they presented us some topics we can take as our final project, which will contribute to the processing.

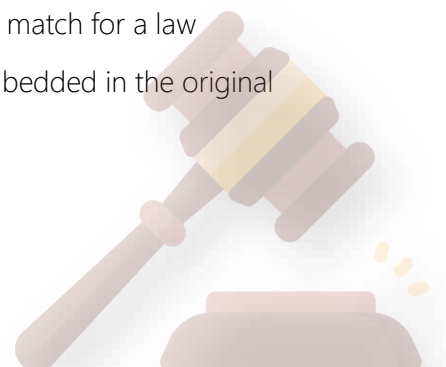
I chose to develop an automatic linker which bind a reference of laws names, with a link to an official source (e.g. Knesset.gov.il).

The main issue of this work is the fact that the laws references in the network aren't always the same as their official's names. Hence, the linker must consider an editing distance when comparing some text to a law's name, and for that purpose must tokenize and normalize the laws names.

The implementation I chose to overcome this problem, uses regular expression for Hebrew words. For each law name the code generates a regular expression, which consider vowels or prepositions (sometimes prepositions in Hebrew are letters).

This way, the code adjusts for most of the references appear in the network.

To make the linker accessible for everyone who is interests, I created a chrome extension that is available for download. When the extension is installed, every website page loaded to the browser being 'scanned' by the linker, and if the linker finds a match for a law name, the linker generate a link to the law's official source which embedded in the original text.



In addition, each law's reference which the linker caches, binds to a popup which contains the law's future URI. This additional feature is inactive yet, but it was added to the linker in purpose to establish a new reorganization idea of the laws URL's.

