

Auto linker of laws references in the network

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

The purpose of the 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 explorer and want to read more about a specific law.

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

We had couple of lessons with two representors of the Ministry of Justice of Israel, who introduce us their work and an advanced processing of digitization of the legislation. In addition, they represented us some subjects 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 of them (e.g. the Knesset site).

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 remains flexible 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 downloading. When the extension is turned on, every page the user route to is 'scanned' by the linker, and if the linker finds a match for a law name in the page, then a link for the law's official source is created in the same position of the law's reference.



In addition, each law's reference which the linker caches, is binds to a popup that contain the law's future Uri. There exists a target to unite the global law's URLs into one form in the network. This feature isn't useful for the user yet, but it was added to the linker in purpose to get this reorganization idea into the public knowledge.

\

