SHORT TERM SCIENTIFIC MISSION (STSM) SCIENTIFIC REPORT

This report is submitted for approval by the STSM applicant to the STSM coordinator

Action number: CA19102
STSM title: An ontology based approach to Websites'Privacy Policies ethical scoring
STSM start and end date: 23/05/2022 to 28/06/2022
Grantee name: Alain Couillault

PURPOSE OF THE STSM:

Online Legal documents (OLDs), such as Privacy Notices and Terms of Services, of web platforms are contracts which legally bind users and platform owners. These documents are lengthy and difficult to read. For these reasons, online users usually agree to these contracts without even reading them, and hence without knowing the implications of giving their consent in terms of rights, privacy or empowerrment.

During the LegiCrowd Onto Project (www.legicrowd.org), experts in law, mathematics and linguistics collaborated to design an ontology for the description of such OLDs. A simple user interface has been developped in the course of this project to allow the annotation of OLDs with the LegiCrowd Ontology: while reading a short section of the OLDs, users have to answer some simple questions and are driven narrow into the ontology (www.legicrowd.org/annotate).

During this STSM, we put this ontology at work to score OLDs on ethical grounds.

DESCRIPTION OF WORK CARRIED OUT DURING THE STSMS

As part of our previous work, we have developped an online environment for OLD annotation (see www.legicrowd.or/annotator) thanks to which a set of OLDs have been annotated.
In our case, annotation consists in associating a set of tags describing the content of each OLD. This set of tags is a subset of the set of tags which have been identified during the LegiCrowd Onto project and which have been organized in an ontology *a la* schema.org.

During this STSM, we considered a set of 6 OLDs which supposetly highly differ with regards to ethics.

The first step consisted in assigning three scores to each tag of the set of tags. Each score is a value among {-2,-1,0,1,2}. This has been a collaborative task which led to much discussion. These scores reflect:
- **P**rivacy: does the platfom state that it respects privacy
- **E**mpowerment: does the platfom state that it provide users with ways to control the platform behaviour
- **T**ranscparency: does the platfom provide information about the way it processes data.

In this way, a triple {P,E,T} is assigned to each tag.
As an example, the tag CPDPROFILILNGYES means that the platform collects data for profiling purposes.

We associated this tag with a low score (-2) on Privacy as well as on Empowerment (-2). Hence, the corresponding {P,E,T} score is {-2,-2,0}.

Scoring a whole OLD on this PET score is achieved by computing the ratio between the sum of the positive and negative scores for each index with, respectively, the positive and negative possible scores for each index.

By doing so, each OLD is scored with two triples, one with positive values and one with negative values, which is reduced to a unique score for each of element of the {P,E,T} triple by summing up each value.

The work has been conducted collaboratively with members of NTUA who, for most of them, had been involved in the LeigCrowd Onto project. Other PhD students attended some of the meetings. The meetings and discussion were organized during the STSM, with a final meeting on Sunday 29th of may.


DESCRIPTION OF THE MAIN RESULTS OBTAINED

The main results obtained consists in a method to evaluate the content of OLD with an ethic point of view.
We coined this method the **PET INDEX** method.
 We plan to demonstrate the results during the next LITHME WAC 2022 meting.

These results can be sum up with the following figure where:
- the horizontal axis is the **empowerment** vector
- the vertival axis is the **Transparency** vector
- the two axes cross at the (0,0) coordinate
- each document (D2, D76...) is positionned on these two axes
- the up and down arrows for each document represent their values (respectivelcy positive and negative) on the **Privacy** vector.
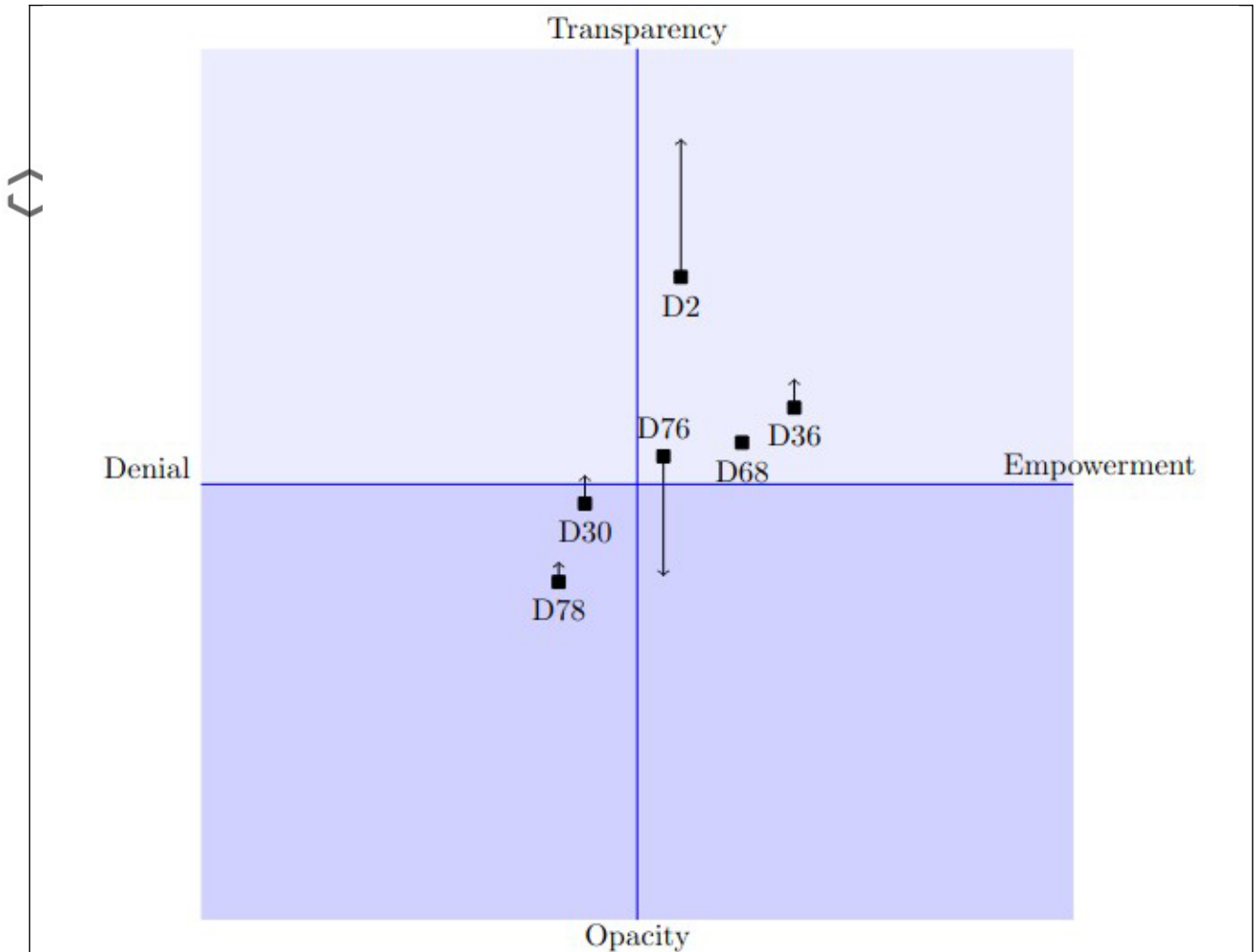
This figure gives a good overview of the ethical scoring for each document. For example, D2 is more transparent and values privacy more then D76.

FUTURE COLLABORATIONS (if applicable)

Other collaboration, potentially with other research partners, could be launched on the following tracks;
- designing a simplified global score based on these three factors, geared towards end users and citizens
- evaluate how the LegiCrowd onto could be linked or merged with other legal ontologies
- work on the standardisation of icons representing the ethics of OLDs.