LegiCrowd (Onto) Project

Building an ontology of Online Legal Documents for annotation, machine learning, DataViz, Citizen empowerment, schema.org extension…. And more

Alain Couillault

National Technical University of Athens

FUTURE NOW

Alexandros Nousias

Sofia Almpani, Theodoris Mitsikas, Petros Stefaneas

Terms of Services and Privacy Notices

- Online Legal Documens (OLDs) legally bind a platform with their users
  - Rights and Obligations of both parties
  - What happens with users' data
  - Actions, complaints, jurisdiction...
- How many times (a day) do we agree on ToS or PN ?
  - Implicitely
  - Explicitely
- How many OLDs do we actually read?

Alain Couillault

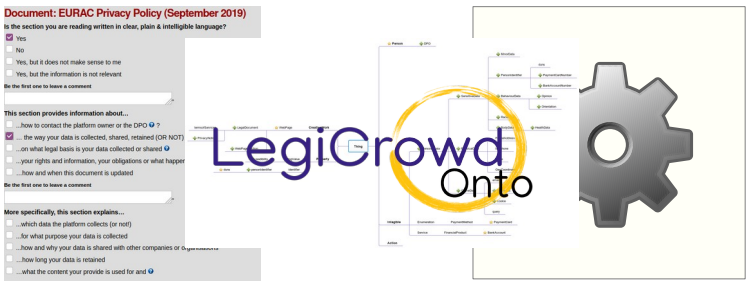
LegiCrowd

Readability and Transparency

- OLDs are too difficult to read
  - Complicated legal content
  - Platforms penchant for transparency
- Legal and ethical issues
  - For the user: Citizen empowerment
  - For the Platform: GDPR compliance
  - For both: Trust between platform and users

Make OLDs readable for humans and for computers

LegiCrowd



Collect OLDs → Annotate → Evaluate
- infer

Render
- DataViz
- Icons
- Grades
- simplified language
- ...

Alain Couillault

LegiCrowd

Previous and useful projects

- ToS;DR
  - A platform for crowdsourced OLD annotation
  - Partners of LegiCrowd
  - Useful input but no formal representation
- Platform for Privacy Preferences (P3P)
  - Highly formalized vocabulary for Privacy Notices
  - W3C (suspended) standard
  - But relies on the platform willingness

1st step : a set of tags

- How to annotate
  - Annotate segments
  - questions/answers
  - Tag/value pairs
- What to annotate
  - List of questions (lawyer's input)
  - Add cases from ToS;DR
  - Add nodes from extended P3P
  - Test and repeat

Alain Couillault

LegiCrowd

Annotation platform

LegiCrowd

Document: EURAC Privacy Policy (September 2019)

Is the section you are reading written in clear, plain & intelligible language?

- [x] Yes
- [ ] No
- [ ] Yes, but it does not make sense to me
- [ ] Yes, but the information is not relevant

Be the first one to leave a comment

This section provides information about…

- [ ] ...how to contact the platform owner or the DPO ?
- [x] … the way your data is collected, shared, retained (OR NOT) by the platform
- [ ] ...on what legal basis is your data collected or shared
- [ ] ...your rights and information, your obligations or what happens in case of conflicts
- [ ] …how and when this document is updated

Be the first one to leave a comment

More specifically, this section explains…

- [ ] ...which data the platform collects (or not!)
- [ ] ...for what purpose your data is collected
- [ ] ...how and why your data is shared with other companies or organisations
- [ ] …how long your data is retained
- [ ] …what the content your provide is used for and

Submit (3/14)

*2.** **Types of Personal Data Subject to Processing*

"Personal data" means any information relating to an identified oridentifiable natural person (the "Data Subject). An identifiable naturalperson is one who can be identified, directly or indirectly, inparticular by reference to an identifier such as a name, anidentification number, location data, an online identifier or to one ormore factors specific to the physical, physiological, mental, economic,cultural or social identity of that natural person.The personal data that may be processed includes browsing data, dataprovided voluntarily by the data subject and cookies.

Alain Couillault

LegiCrowd

Adding some semantics

- Schema.org
  - Shared vocabulary for webpages (and more) mark up
  - Founded by Google, Microsoft, Yahoo and Yandex
  - Open Community process
  - Provides out of the box descriptions for persons, actions, organizations…
  - Supports RDF

Full Hierarchy

Schema.org is defined as two hierarchies: one for textual property values, and one for the things that they describe.

This is the main schema.org hierarchy: a collection of types (or "classes"), each of which has one or more parent types. Although a type may have more than one super-type, here we show each type in one branch of the tree only. There is also a parallel hierarchy for data types.

- Thing
  - Action
    - AchieveAction
      - LoseAction
      - TieAction
      - WinAction
    - AssessAction
      - ChooseAction
        - VoteAction
      - IgnoreAction
      - ReactAction
        - AgreeAction
        - DisagreeAction
        - DislikeAction
        - EndorseAction
        - LikeAction
        - WantAction
      - ReviewAction

schema.org

Home Schemas Documentation

WebPage

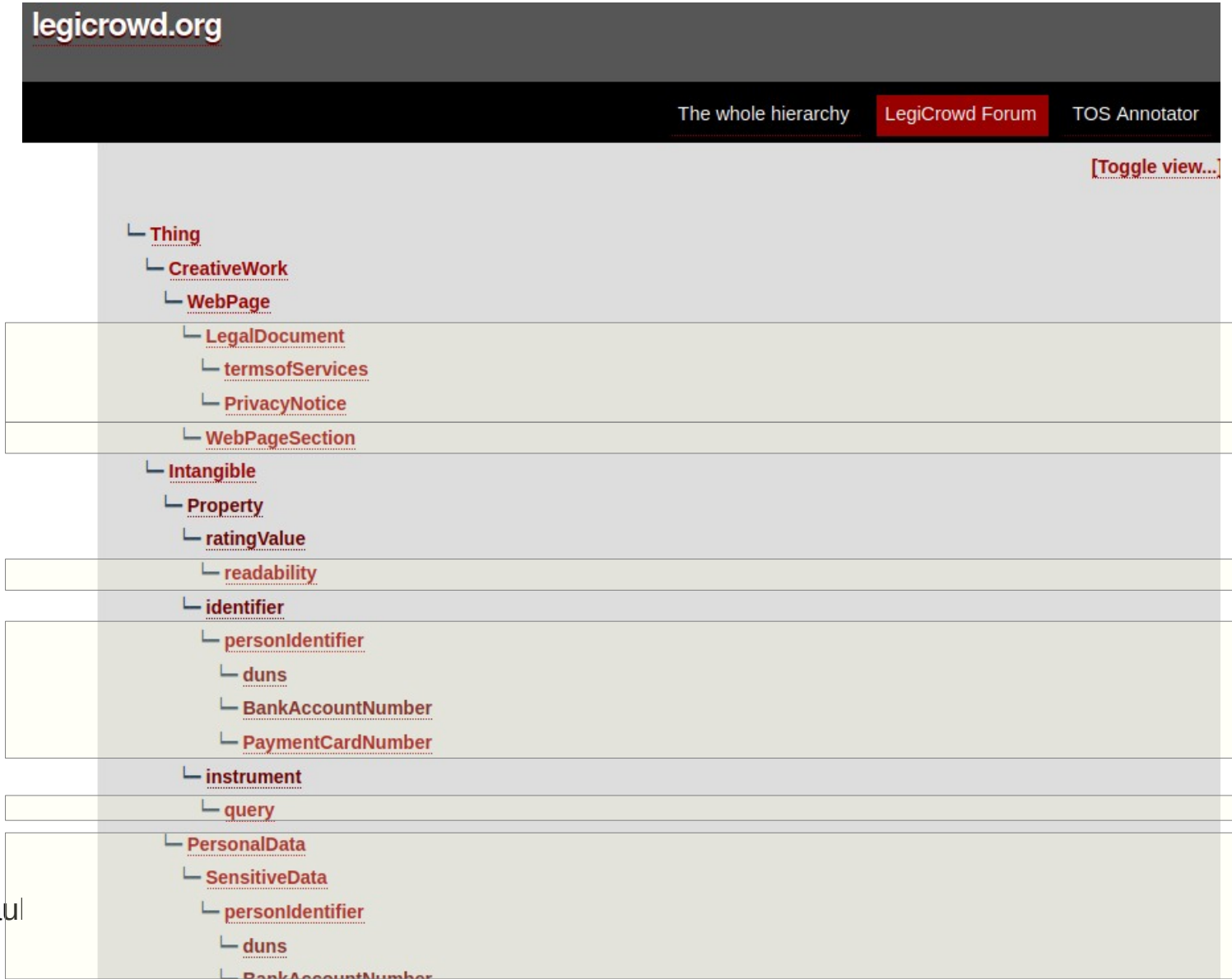
Thing > CreativeWork > WebPage

A web page. Every web page is implicitly assumed to be declared to be of type WebPage, so the various properties about that webpage, such as `breadcrumb` may be used. We recommend explicit declaration if these properties are specified, but if they are found outside of an itemscope, they will be assumed to be about the page.

[more...]

Property	Expected Type	Description
Properties from WebPage		
`breadcrumb`	BreadcrumbList or Text	A set of links that can help a user understand and navigate a website hierarchy.
`lastReviewed`	Date	Date on which the content on this web page was last reviewed for accuracy and/or completeness.
`mainContentOfPage`	WebPageElement	Indicates if this web page element is the main subject of the page. Supersedes aspect.
`primaryImageOfPage`	ImageObject	Indicates the main image on the page.
`relatedLink`	URL	A link related to this web page, for example to other related web pages.
`reviewedBy`	Organization or Person	People or organizations that have reviewed the content on this web page for accuracy and/or completeness.
	URL	One of the more significant URLs on the page. Typically, these are the

Alain



Alain Couillaul

legicrowd.org

The whole hierarchy | LegiCrowd Forum | TOS Annotator | Login

termsofServices

Human-readable terms of service documentation.

Thing>CreativeWork>WebPage>LegalDocument>termsofServices>

[more...]

Attached to the following types	
Type	**Description**
LegalDocument	A legal document enforcing the relationship between the platform and its users such as Terms of Services or a Privacy Notice

Properties from WebPage		
Property	**Expected Type**	**Description**
WebPageSection		A section of a WebPage.
readability		The easiness/difficulty of reading the document (can be an text or an integer, for example a Fleish Readability Index)

Current status of LegiCrowd Onto

- Annotation environment
- Ontology
- Preprocessing
  - Collection of OLDs
  - Splitting

Next steps

- Complete LegiCrowd Onto
  - Launch Bêta test
  - Complete and propose schema.org extension
- Move to LegiCrowd
  - Work with Katerina Zdravkova on OLD evaluation and on inferences
  - Expose to the crowd
- Later
  - Machine Learning
  - Standardisation

Special thanks to the LegiCrowd partners, to the NGI and enet-collect teams who made this project possible, to Petros Stefaneas from the National Technical University of Athens and to Mariachiara Tallacchini from the Università Cattolica del Sacro Cuore in Milano who hosted STSMs.

The LegiCrowd project has received funding from the European Union's Horizon 2020 research and innovation programme under the NGI_TRUST grant agreement no 825618.

Alain Couillault