

SMACC - text analyzer for legal assistance

In our daily life we regularly come across various contracts and agreements: real estate, electricity, mobile phone plan, insurance, general conditions of sale, etc. These contracts, in paper or on-line versions, may contain thousands of lines; reading and understanding these lines is a real challenge for many people. The language in the legal domain texts is notoriously hard to digest for non-professionals.

According to several surveys (Scania, 2011; the Deloitte, 2017), less than 10% of people read contracts presented to them in their entirety. Of these 10%, one in five claims to fully understand what they read. In addition, 20% of people admit to having suffered from the lack of information concerning the legality of the contract clauses that they are presented with.

SMACC (Smart Contract Checker) was developed to address those issues. SMACC is a tool that offers legal assistance to users faced with a bidding contract and consequent obligations and who would like to get a better overview and analysis of the document at hand.

SMACC takes a contract in a text or PDF format, analyzes the document and builds a summary of the contents, highlighting the key points and figures, as well as providing a list of relevant legal texts for each of the contract clauses for further examination. The search for relevant documents is performed by a powerful neural network-based algorithm which achieves the accuracy of 92% on the test data from the real estate rental domain.



SMACC analyzer consists of the 4 main components: the splitter which identifies the clauses or (in case of a finer split) smaller self-contained blocks of text; a search engine, which matches the clauses against a legal database in a specific legal domain and retrieves the corresponding legal texts; the main analyzer which verifies the legality of each given clause vis-a-vis the law by performing a comparative analysis of the clause and the retrieved legal text; finally a summarizer compiles a summary of key points that a user should pay attention to. A list of applicable laws corresponding to the contract contents is also provided for further consultation and/or manual check. In case of a PDF input an additional pdf-to-text conversion step is performed in order to obtain bare text which can be processed in an NLP pipeline. The presented version of SMACC handles contracts in the domain of real estate leasing. Each legal domain has a specific database of legal texts. The rest of the pipeline is the same for different legal domains.