

Composizione dell'unità UniGE-DIBRIS nel progetto NGUPP

Nome	Cognome	Ruolo nel progetto
Enrico	Giunchiglia	PO DIBRIS, partecipante
Viviana	Mascardi	PA DIBRIS, partecipante
Luca	Oneto	PA DIBRIS, partecipante
Luca	Verderame	RTD DIBRIS, collaboratore
Chiara	Accinelli	Assegnista DIBRIS
Andrea	Barca	Assegnista DIBRIS
Riccardo	Bianchini	Assegnista DIBRIS
Gianluca	Boletto	Assegnista DIBRIS
Andrea	Gatti	Assegnista DIBRIS
Martina	Pulze	Assegnista DIBRIS
Gianluca	Sommariva	Assegnista DIBRIS

Sintesi dei contributi di UniGE-DIBRIS al progetto NGUPP categorizzati rispetto ai punti elencati nella proposta di progetto

Obiettivi raggiunti sul Punto 5: Realizzare un'anonimizzazione semiautomatica delle decisioni giudiziarie emesse negli uffici giudiziari

- Identificazione delle necessità di anonimizzazione dei documenti giudiziali presso il Tribunale di Genova e identificazione di un caso di studio di interesse per l'anonimizzazione (sentenze di divorzio).
- Studio ed identificazione degli algoritmi e delle tecnologie per la pseudoanonimizzazione di documenti in formato testuale e PDF; sono state analizzate e testate nel dominio giuridico le

tecniche di local differential privacy con particolare riferimento a tecniche di distorsione o randomizzazione e di generalizzazione. Da questa attività, sono state selezionate le tecniche di generalizzazione dei quasi-identifier relativi alle informazioni geografiche e alle generalità anagrafiche e le tecniche di distorsione sui nominativi dei soggetti coinvolti nel documento. Inoltre, sono state testate e valutate le tecniche di soppressione degli identificatori espliciti sui documenti giudiziali.

- Sviluppo di uno strumento prototipale di pseudonominizzazione semiautomatica di documenti giudiziali adottando le tecniche individuate nel punto precedente.

Obiettivi raggiunti sul Punto 6: Realizzare strumenti di analisi semantica delle sentenze

- Identificazione delle necessità di analisi semantica dei documenti giudiziali presso il Tribunale di Genova e identificazione di uno use case di interesse per l'analisi semantica (sentenze di divorzio).
- Studio ed identificazione degli algoritmi e delle tecnologie per l'analisi di documenti in formato testuale e PDF; sono state analizzate le tecnologie PyPdf, PyPDF2, PyPDF4, pdftotext e PyMuPdf; da questa analisi comparata, le tecnologie più adeguate sono risultate pdftotext per l'estrazione del testo puro dal file e PyMuPdf per la modifica dello stesso e la ricerca della posizione geometrica di un determinato lessema all'interno del documento.
- Studio ed identificazione degli algoritmi e delle tecnologie per Natural Language Processing più adatti alla soluzione del problema; sono state analizzate le tecnologie NLTK, spaCy e Regex insieme ad un ampio insieme di librerie Python per la ricerca di nomi, date e luoghi all'interno di testo; da questa analisi comparata, le tecnologie più adeguate sono risultate Regex e spaCy, utilizzando la prima per dedurre informazione a partire dalla struttura della frase e la seconda per verificarne la bontà ed estrarre ulteriori dati studiando la semantica delle singole parole all'interno della frase con un approccio simile alla analisi logica.
- Realizzazione di uno strumento prototipale per l'estrazione di informazioni semantiche nel dominio delle sentenze di divorzio e sua integrazione con lo strumento di pseudonominizzazione descritto al punto 5; lo strumento integrato di analisi semantica e pseudonominizzazione, "Klondike-UniGE", individua i dati anagrafici delle parti e dei figli, li colleziona per fini statistici e li rimuove dal testo originale per fini di anonimizzazione; estrae inoltre dati processuali (es. spese straordinarie, presenza di servizi sociali, violenza casalinga, affido di figli, ecc.) per fini statistici.

Obiettivi raggiunti sul Punto 9: Utilizzare tecniche AI per gestione dei precedenti: ovvero strumento di intelligenza artificiale per l'analisi delle decisioni giudiziarie

- Identificazione delle necessità di analisi semantica delle decisioni giudiziarie espresse dal Tribunale di Genova.
- Studio ed identificazione degli algoritmi e delle tecnologie per individuare e collegare cronologicamente la storia di un procedimento, attraverso le diverse sentenze che lo compongono.
- Realizzazione di uno strumento prototipale basato su Python e SOLR, appoggiandosi alle librerie pySolr, pandas e csv, per la individuazione e la ricostruzione cronologica della storia di un procedimento, attraverso le diverse sentenze che lo compongono.

Obiettivi raggiunti sul Punto 11: Database giurisprudenziali

- Identificazione dei requisiti di una piattaforma per l'immagazzinamento e la ricerca avanzata di sentenze presso il Tribunale di Genova.
- Studio ed identificazione degli algoritmi e delle tecnologie più adeguati alla realizzazione di tale piattaforma. In questa fase sono state analizzate le seguenti tecnologie:

- o PHP – per la realizzazione e la gestione dei servizi della piattaforma lato Backend;
- o MySQL – per la gestione dei dati tramite un Database Relazionale;
- o MongoDB – per la gestione dei dati tramite un Database Documentale.
- o SOLR – per l’ottimizzazione delle query su tipi di dati testuali;
- o Python – per le notevoli librerie e strutture dati a supporto dell’analisi Dati e degli algoritmi di Intelligenza Artificiale.

Da questa analisi comparata, la tecnologia più adatta è risultata essere SOLR, per il supporto nativo alla efficiente esecuzione delle query sul contenuto del loro testo, integrata tramite Python ai servizi Backend necessari alla piattaforma per funzionare costantemente.

- Realizzazione della piattaforma prototipale, “Lexi-UniGe”, e suo popolamento con dati provenienti da più di 100.000 sentenze fornite dal Tribunale di Genova previa convenzione stipulata con il Dipartimento DIBRIS di UniGe.
- Individuazione – in collaborazione con il Tribunale di Genova - dei metadati più rilevanti da associare ai documenti delle sentenze.
- Estensione della piattaforma “Lexi-UniGe” con strumenti basati su ReactJs e Flask per la realizzazione di ricerche avanzate sul testo e sui metadati delle sentenze e per la modifica di tali metadati.
- Estensione della piattaforma “Lexi-UniGe” con la possibilità di commentare e salvare nei preferiti le sentenze più rilevanti.

Obiettivi raggiunti sul Punto 13: Chatbot

- Sperimentazione delle API dei modelli di AI-generativa di OpenAI, ovvero GPT-3.5 (e successivamente GPT-4), con l’obiettivo di fornire i riassunti di interi provvedimenti (sono stati analizzati i limiti di affidabilità all’aumentare della complessità di un provvedimento); rispondere alle domande in riferimento all’esito processuale di un provvedimento; classificare e strutturare in vari campi (quali per esempio: parte, controparte, data, tipologia di provvedimento, oggetto...) un provvedimento a partire dal contenuto testuale dello stesso. Nota: nonostante l’utilizzo di GPT sia risultato promettente, le evidenti ragioni di privacy relative alla raccolta dei dati inseriti in tale piattaforma, hanno imposto la scelta di evitarne l’utilizzo per analizzare documenti con dati sensibili. Per questa ragione, tutti gli strumenti prototipali sviluppati dall’Università di Genova hanno adottato tecnologie che consentissero di avere il pieno controllo sui dati analizzati, assicurandone una gestione compatibile con le normative vigenti in termini di utilizzo e memorizzazione.

Obiettivi raggiunti sul Punto 3.1.2: Sperimentazione dei modelli e delle soluzioni nei testbed

Tutti gli strumenti prototipali di UniGe sono stati sviluppati seguendo metodi agili di ingegneria del software e conducendo una sperimentazione locale basata su unit testing e integration testing da parte degli sviluppatori.

Per quanto riguarda il coinvolgimento degli utenti finali, a luglio 2023 la piattaforma prototipale “Lexi-UniGe” è stata installata sui server di Liguria Digitale, con la quale il Tribunale di Genova ha una convenzione, per consentirne l’accesso remoto sicuro da parte del personale del Tribunale di Genova, al fine di condurre un’adeguata ed estesa sperimentazione. La sperimentazione si è conclusa con successo.

A settembre 2023 lo strumento prototipale “Klondike-UniGE” è stato installato localmente sulla macchina di un giudice presso il Tribunale di Genova per consentirne la sperimentazione. Prima dell’installazione lo

strumento “Klondike-UniGE” era stato valutato dagli sviluppatori rispetto a sentenze annotate manualmente da un esperto giurista.

Obiettivi raggiunti sul Punto 3.1.3: Capacity building e estensione della sperimentazione

La piattaforma prototipale “Lexi-UniGe” è stata installata sui Server di Liguria Digitale in vista di un'estensione della sperimentazione a più soggetti all'interno del Tribunale di Genova e a più Tribunali sul territorio ligure. Questo nell'ottica di trasferire le nozioni apprese e le applicazioni sperimentate a diverse tipologie di uffici giudiziari.