

Verso un documentale semantico della PA

C. Batini¹, G. Santucci¹, M. Palmonari³, V. Bellandi³, E. Fersini²,
B. Pernici⁵, F. Zanzotto⁴, G. Vecchi⁵ e S. Ronchi⁵

1. Consorzio Interuniversitario Nazionale di Informatica (CINI)
2. Università di Milano
3. Università di Milano-Bicocca
4. Università di Roma Tor Vergata
5. Politecnico di Milano

Abstract

Per erogare i servizi agli utenti, le Pubbliche Amministrazioni (PA) centrali e locali fanno largo uso dei dati; diverse stime qualitative portano a concludere che nelle basi di dati è contenuto il 10-20% della conoscenza disponibile in forma di dati digitali, il restante 80-90% è nei dati semi-strutturati e non strutturati costituiti da documenti, file audio, immagini, video, ecc.

Il presente lavoro ha due scopi: riassumere le esperienze portate avanti negli ultimi quattro anni dal Consorzio Interuniversitario Nazionale di Informatica (CINI) nel progetto Datalake finanziato dalla CRUI nell'ambito di una convenzione con la Direzione Generale dei Sistemi Informativi Automatizzati (DGSIA) del Ministero di Giustizia in sinergia con altri progetti collegati del Ministero; e mostrare come le esperienze effettuate, i Proof of Concept e le specifiche funzionali prodotte, possano costituire un repository di funzionalità per un "documentale semantico della PA", che faccia evolvere i sistemi informativi delle PA verso piattaforme in cui i dati non strutturati possano essere sfruttati e integrati con i dati strutturati, per migliorare e dare valore ai servizi digitali che la PA eroga, e i processi di governo possano essere condotti utilizzando tutta la conoscenza, espressa nei documenti e altre forme di dati non strutturati.

Vengono descritte l'organizzazione giudiziaria, i procedimenti e i processi, le esigenze degli utenti, la struttura funzionale del Datalake, l'architettura di implementazione, in vista di un percorso di progettazione e produzione indirizzato a tutto l'insieme delle PA.

Keywords

Documentale semantico, Datalake, Giurisdizione, Procedimento civile, Procedimento penale

1. Procedimenti, Processi, Organizzazione, Sistemi informativi di Giustizia

Il Ministero di Giustizia espleta funzioni amministrative connesse alla funzione giurisdizionale sia nel campo civile che penale. La Magistratura identifica un complesso di strutture e istituzioni volte all'amministrazione della Giustizia e soggette alla titolarità di singoli individui, i magistrati.

I processi primari¹ del Ministero e della Magistratura (nel seguito denominati assieme con il termine Giustizia) riguardano il procedimento penale

e il procedimento civile; il procedimento penale ha come fasi le indagini preliminari, l'attività di cognizione nei tre gradi di giudizio del processo penale, e l'attività di esecuzione di pene o attività alternative della giustizia minorile e di comunità. Anche il procedimento civile è composto da una fase di cognizione, che prevede tre gradi di giudizio, e una fase di esecuzione.

La fase di cognizione del procedimento civile è da tempo oggetto di automazione nel sistema informativo del processo civile telematico (PCT). Di conseguenza, il grado di digitalizzazione dei dati strutturati e documenti del *fascicolo del* procedimento

¹ Qui e nel seguito useremo il termine *procedimento* quando intendiamo l'attività giurisdizionale come disciplinata da legge, *processo* quando facciamo riferimento alla concreta attuazione del procedimento nell'ambito dei Tribunali, Corti d'Appello e Corte di Cassazione.

civile è nettamente più avanzato rispetto a quelli dei fascicoli delle indagini preliminari e del procedimento penale.

Il Fascicolo digitalizzato di un procedimento civile è costituito da atti e documenti e, per i procedimenti definiti, dalla sentenza; un atto del processo civile è un artefatto documentale inerente un processo, i cui contenuti e forma sono previsti da norme; un documento è un qualsiasi artefatto (testo, registrazione audio, immagine, video, ecc.) inerente il processo e allegato ad atti. Lo svolgimento del processo civile è rappresentato in termini di stati ed eventi.

Nel primo grado del procedimento civile e in misura più ampia nella fase delle indagini preliminari del procedimento penale sono acquisite molte fonti documentali di prova, dai tabulati telefonici alle tracce di carte di credito, alle visure, le intercettazioni telefoniche trascritte e molte altre.

I processi primari delle indagini preliminari e di cognizione del procedimento penale sono diventate solo recentemente oggetto di automazione in un progetto in corso.

La fase di esecuzione è caratterizzata da una maggiore automazione del procedimento penale rispetto al civile, che vede nel Casellario Giudiziale e nelle basi di dati del Dipartimento della Amministrazione Penitenziaria e del Dipartimento della Giustizia Minorile e di Comunità le principali realizzazioni.

Il progetto Datalake nasce nel 2019 per iniziativa della DGSIA, che ha affidato alla CRUI, e questa al CINI, una linea di ricerca rinnovata negli anni, che investiga l'adozione delle tecnologie basate sulla elaborazione del linguaggio naturale, l'approccio simbolico, l'apprendimento automatico e recentemente l'intelligenza artificiale generativa, che possano rivelarsi più utili nell'espletamento dei processi primari di cognizione ed esecuzione civile e penale. Nel 2022 Giustizia ha inserito il Datalake tra i progetti finanziati con fondi PNRR, la cui realizzazione è stata affidata tramite gara a un raggruppamento di tre imprese, Almviva, Almwave e Accenture.

2. Esigenze dei Magistrati nel procedimento penale e civile

Il progetto Datalake si è concentrato sulla fase delle indagini preliminari del procedimento penale, l'esecuzione penale e la fase di cognizione del procedimento civile.

Le funzionalità sviluppate nascono da una attività di elicitazione delle esigenze degli utenti, che per il procedimento penale sono i Procuratori e la Polizia giudiziaria, e per il procedimento civile sono i Giudici.

La raccolta dei requisiti è avvenuta in ambito indagini preliminari tramite la condivisione di Proof

of Concept su procedimenti definiti e successivamente attraverso lo strumento delle consulenze su procedimenti in corso; in ambito civile sono state effettuate interviste a Magistrati della Corte d'Appello di Milano, che a breve sperimenteranno un primo insieme di funzionalità sviluppate dal fornitore Almwave; l'esito della sperimentazione darà luogo ad una nuova versione del sistema che potrà essere adottata in tutte le Corti d'Appello. Le esigenze degli utenti sono descritte sinteticamente nel seguito.

Processi primari – indagini preliminari

- Ricerche puntuali e aggregazioni semantiche per discovery e conferma di indizi e prove.
- Analisi integrata di conoscenza di tipo relazionale con visualizzazione.
- Selezioni di cluster di nodi di grafo semantico con determinate proprietà.
- Ricostruzione delle relazioni mantenute da soggetti sospettati, componendo l'intera rete relazionale del soggetto sospettato.
- Traduzione in testo e arricchimento semantico di messaggi audio.

Processi primari - civile

- Arricchimento ed esplorazione di conoscenza giuridica nel corso del processo.
- Ricerca semantica per entità nominali, menzioni, concetti, termini, sintagmi, frasi semplici a fini di serialità e ricerca di precedenti.
- Ricerca delle sentenze e della giurisprudenza rilevante con unico punto di accesso integrato.
- Selezione delle sentenze riguardanti materie (es. danno da violazione della privacy) non ricomprese nei metadati di accesso.
- Supporto decisionale nella fase di cognizione della sentenza e collegamento di atti e documenti rilevanti con la sentenza.
- Estrazione dalle sentenze civili del risultato finale/outcome del processo (es. liquidazione del danno, assegno di mantenimento nel processo di divorzio giudiziale) e caratteristiche salienti correlate.

Processi di governo nel PCT

- Modelli predittivi delle durate attese e loro variabilità in base alle caratteristiche del procedimento.
- Valutazione della complessità attesa del procedimento, per la distribuzione del carico tra i giudici, individuazione dei "colli di bottiglia", individuazione di segnali ed eventi che hanno un forte impatto sulla durata della procedura, analisi dell'impatto di cambiamenti di leggi, regolamenti e prassi.
- Statistiche descrittive su dati strutturati e sentenze.
- Analisi di correlazione tra caratteristiche salienti e outcome in processi civili e a fini di uniformità (c.d. tabellazione).

- Benchmarking della durata dei processi in funzione della sezione, del distretto

3. Funzionalità

I Proof of Concept sviluppati e le specifiche funzionali prodotte nell'ambito del progetto Datalake riguardano le seguenti macro-funzionalità: Preparazione, Arricchimento semantico e Integrazione della conoscenza, Ricerca semantica e Analisi, Gestione della base di conoscenza e controllo di qualità. Di seguito le funzionalità di dettaglio.

F1 - Preparazione

1. Cleaning dei testi (rimozione caratteri speciali, correzione lettere accentate, rimozione intestazioni, rimozione timbri, gestione della punteggiatura)
2. OCR dei testi e generazione di documenti interpretabili.
3. Strutturazione dei testi in parti: premessa, descrizione del caso e decisione.
4. Classificazione dei testi all'interno dei fascicoli.

F2 - Arricchimento semantico e Integrazione della conoscenza

La caratteristica peculiare dell'approccio adottato (vedi [2], [6], [7]) rispetto a quelli precedenti risiede nello sforzo di consolidare la conoscenza estratta combinando collegamenti e raggruppamenti, invece di limitarsi a identificare le menzioni delle entità. L'impatto di questo approccio è osservabile soprattutto durante la ricerca nei documenti, vedi funzionalità F3.

Processo civile.

Sono state applicate diverse tecniche per estrarre, collegare e consolidare entità da sentenze e produrre annotazioni semantiche che associano le entità estratte a specifiche sezioni delle sentenze. In particolare, è stata sviluppata una pipeline NLP che combina diverse tecniche:

- Named Entity Recognition (NER), che sfrutta approcci basati su regole e approcci neurali, tarato sulla distribuzione dei dati nel dominio (classificatori sequenziali su caratteristiche provenienti da un trasformatore di codifica basato su BERT).

- Named Entity Linking (NEL), che si basa sull'algoritmo di recupero di entità BLINK addestrato sulla Wikipedia italiana nell'ambito del progetto.

- NIL Prediction, che consiste nel decidere se collegare una menzione di entità all'entità ad essa associata da un NEL o etichettarla come entità nuova non presente nella base di conoscenza (NIL); per questo compito si utilizza un classificatore interno basato su caratteristiche.

- NIL Clustering, che consiste nel raggruppare le menzioni di entità che si riferiscono alle stesse entità del mondo reale (tipicamente applicato alle menzioni etichettate come NIL perché le entità collegate a una

base di conoscenza sono implicitamente raggruppate).

- Costruzione dell'Entity Registry - L'Entity Registry è un componente in cui a ogni entità, arricchita con attributi dedotti nella fase di linking, corrisponde una voce unica, evitando duplicati e disambiguando omonimi e sinonimi. Le annotazioni di testo vengono aggiornate con identificatori delle entità.

- Affinamento delle decisioni prese alla fine della pipeline sulla base di alcune regole specifiche per il dominio (in particolare per la classificazione di entità specifiche e a grana fine).

Per eseguire NEL e NIL prediction in una sola volta, è stato anche recentemente esplorato un algoritmo di named entity disambiguation esteso per prevedere NIL come classe.

- Estrazione delle relazioni (es. Relazione vittima colpevole, espressione "a carico di") per mezzo di modelli transformers pre-addestrati di rappresentazione del testo, con training effettuato secondo una politica di cross-validation e modello di estrazione basato sul paradigma entity-relationship e basato su REBEL

E' stata applicata la tipizzazione fine di entità Few-shot, il compito che consiste nell'assegnare entità con tipi specifici da tassonomie; per questo sono stati esplorati sia metodi neuro-simbolici in cui la tassonomia è modellata esplicitamente, sia metodi basati su LLM con prompt impliciti.

Processo penale - Indagini preliminari.

Per i documenti relativi alle indagini preliminari è stata applicata una pipeline molto simile per l'estrazione delle entità e la conseguente annotazione dei documenti, un simile paradigma di ricerca semantica; tuttavia, sono state applicate altre funzionalità e tecniche quali:

- Estrazione di rappresentazioni di grafi dai dati delle applicazioni di messaggistica istantanea (IMA), ad esempio i dump di Whatsapp, e archiviazione in un DB di grafi (Neo4J); i messaggi possono essere interrogati utilizzando un linguaggio strutturato che supporta l'analisi dei dati basata sui grafi.

- Arricchimento dei contenuti con la tecnologia speech-to-text; è stato utilizzato Whisper di OpenAI per trascrivere messaggi audio e rendere questi contenuti ricercabili. Tutti i messaggi e le chat vengono analizzati utilizzando una pipeline NLP come quella descritta in precedenza, che supporta la ricerca semantica alimentata da annotazioni basate su entità.

- Arricchimento semantico e specializzazione di ontologie di annotazione delle entità rispetto a tassonomia specifica (is-overlapping, is-within, ordinamento).

Altre funzionalità sviluppate riguardano la Domain adaptation a fini di riuso di tecniche di annotazione con modelli neurali (per NER è usata un'architettura basata su Adapters), l'argumentation mining, la text summarization, la georeferenziazione delle entità spaziali.

F3 - Ricerca semantica e Analisi

Funzionalità di ricerca comuni a indagini preliminari e processo civile

Sulla base delle funzionalità di arricchimento semantico mostrate nel punto precedente, le entità che in fase di ricerca compaiono nei filtri possono riferirsi a menzioni presenti in documenti diversi; inoltre, quando un utente esplora ad es. una sentenza, può trovare tutte le menzioni di un'entità in tutto il documento, una caratteristica che può diventare particolarmente rilevante per sentenze lunghe o altri documenti.

L'architettura concettuale per la ricerca semantica è rappresentata in Figura 1.



Figura 1 – Architettura del sistema di ricerca

I componenti sono:

- **Keyword search:** consente di effettuare una semplice ricerca per parola chiave. Tale modulo può essere utile come punto di inizio della ricerca, prima di attivare il modulo di faceted search.
- **Faceted search:** combina ricerche per parole chiave e filtri basati su attributi delle sentenze. Il modulo utilizza tecnologie note di indicizzazione e interrogazione di basi documentali (ad es., Elasticsearch). Inoltre utilizza strumenti interattivi di selezione delle facets che permettono di aiutare l'utente a specificare i filtri da applicare sui risultati.
- **LLM-QA:** implementa una ricerca dialogica basata sul modello RAG (Retrieval Augmented Generation). Un LLM generativo gestisce l'interazione con l'utente e la generazione delle risposte; un modulo di *retrieval neurale* permette di indicizzare paragrafi delle sentenze mediante *embedding* e di recuperare paragrafi rilevanti per una domanda dell'utente in fase di ricerca.
- **Document explorer:** consente di esplorare un documento, ovvero una sentenza, guidando la ricerca al suo interno di specifiche entità o concetti menzionati.

- **Annotation editor:** consente di modificare le annotazioni, per supportare un processo di annotazione supervisionata in cui gli utenti possano modificare annotazioni sbagliate o imprecise, e aggiungere nuove annotazioni.
- **Concept search:** consente di cercare o esplorare concetti rilevando secondo logiche di dominio. Tale modulo può essere utile per aiutare l'utente, a selezionare concetti specifici di interesse in una fase esplorativa o di raffinamento della ricerca.
- **Advanced search:** consente di effettuare una ricerca avanzata, ovvero combinando diversi filtri su attributi dei documenti. Tale modulo è inserito in molte applicazioni di ricerca su dati strutturati o semi strutturati, per affiancare i moduli basati su Keyword search e Faceted search; tipicamente, la funzione di questo modulo è la costruzione di interrogazioni precise basate su descrizioni strutturate dei documenti.
- **Features & values search:** ha lo scopo di estrarre valori associati a features (es. valore economico dell'assegno di mantenimento). Sono stati considerati due modelli opensource disponibili, Camoscio e Stambecco (versioni di LLAMA addestrate sulla lingua inglese ed adattati per la lingua italiana) ed il modello pay-per-use noto come ChatGPT. Sono state sperimentate tecniche basate su prompt engineering, con prompt delle seguenti tipologie: Direct Instruction Prompts, Contextual Prompts, Bridging Prompts, Socratic Prompts.

Per tutte le funzionalità sono state condotte analisi di accuratezza, a titolo puramente esemplificativo sono state ottenute le seguenti capacità predittive in termini di accuratezza:

- relazione "a carico di" su menzionata: 83,5%
- importo dell'assegno di mantenimento in favore dei figli nelle use d iseparazione: 77,52%

Funzionalità di analisi per processi di governo - civile

L'organizzazione semantica dei documenti ottenuta attraverso le funzionalità di arricchimento semantico e di integrazione abilitata dall'Entity registry permette di effettuare molteplici statistiche e correlazioni su dati strutturati collegati a documenti annotati, ad es. numero di documenti che coinvolgono entità legali naturali, numero e valor medio di minori coinvolti nelle decisioni di divorzio, correlazione a fini di tabellazione del valore della liquidazione e features collegate nei processi per danno non patrimoniale.

Ulteriori analisi riguardano curve di sopravvivenza dei processi e variabili esplicative della durata temporale e complessità di processi.

Diverse funzionalità di analisi sono state sviluppate nell'ambito del progetto PON Next

Generation UPP² e di altri progetti finanziati dalla CRUI. Sono state condotte le seguenti ricerche basate sui registri del sistema SICID per il PCT (vedi [2], [5]):

- Analisi delle varianti (i cluster di procedimenti aventi la stessa struttura e sequenza di stati) e loro valutazione a scopo di monitoraggio. In particolare, sono stati analizzati i fattori che hanno un impatto maggiore sulla durata dei processi. Per questa attività è stato utilizzato lo strumento di process mining Apromore.

- Identificazione degli eventi critici: viene valutato l'impatto di eventi specifici sulla durata dell'esecuzione di un processo, per individuare gli eventi sistematicamente associati a situazioni anomale. Sono state esaminate sia le fasi che la durata totale dei procedimenti.

- Approcci predittivi per gli alert: a partire da sequenze di stati o di eventi nei registri sono stati costruiti predittori, basati su tecniche di apprendimento automatico con reti neurali LSTM, per prevedere la durata residua dei processi e degli stati durante il loro svolgimento.

Un cruscotto per il controllo di gestione è stato realizzato per la Corte di Cassazione. La soluzione adottata è stata quella di realizzare un cruscotto *alimentato direttamente dal database sottostante al registro SIC della Corte*, con dati *aggiornati quattro volte al giorno*. Sono stati individuati tutti i dati per:

- alimentare le variabili e gli indicatori individuati come necessari per descrivere il percorso del fascicolo nelle diverse fasi e per calcolare indici quali, ad esempio, il *Disposition Time* e l'indice di ricambio;
- costruire la serie storica di tali dati a partire dal gennaio 2019.

Funzionalità di analisi per Indagini preliminari

Analisi di conoscenza di tipo relazionale con visualizzazione (es. Selezioni di cluster di nodi con determinate proprietà) e anomaly detection.

Funzionalità per l'esecuzione penale

Integrazione a fini di analisi sociale dei dati relativi alle restrizioni della libertà/pene alternative vissute dai detenuti nel corso della loro vita.

F4 - Gestione della base di conoscenza e controllo di qualità - Principali metodologie e funzionalità sviluppate

1. Manuale di policy dei processi di pseudonimizzazione - Vengono considerate le diverse tipologie di pseudonimizzazione e individuati diversi tipi di trattamento di dati e documenti in cui è rilevante il processo di pseudonimizzazione, (es. pubblicazione, collegamento di basi di dati, ecc.) e le proprietà che devono essere rispettate nei vari casi. Viene fornito un

metodo generale che può essere seguito per i diversi tipi di trattamento rilevanti nel progetto Datalake.

2. Gestione dell'Entity Registry con creazione, aggiornamento, cancellazione di entità, merge e split di entità.
3. Estrazione di lessici di termini basati su sintagmi nominali da sentenze e organizzazione in ontologia e specializzazione del lessico in ambito legale (fine-tuning).
4. Assessment di qualità del NER e NEL.
5. Benchmarking di modelli di estrazione rispetto a diversi livelli di profondità della tassonomia, e strumenti di annotazione tra diversi modelli di estrazione delle relazioni.
6. Introduzione di guardrail per evitare errori o sentenze non processabili.
7. Manuale di qualità dei dati, documenti e modelli diagnostici e predittivi (accuratezza, completezza, currency, equità, spiegabilità), vedi [4]. Per la accuratezza e equità, il manuale è allineato con i documenti di policy emessi dalla UE (vedi [3]).

4. Ontologie/Tassonomie e loro generazione top down e bottom up

Nelle funzionalità del Data Lake esposte sono utilizzate le seguenti ontologie:

- Ontologia top dei procedimenti di Giustizia (cognizione ed esecuzione), costituita da circa 400 entità, rappresentata per mezzo di circa 40 schemi nel modello Entity Relationship a diverso livello di integrazione/estrazione.
- Ontologia per esecuzione penale, costituita da circa 100 entità e 8 schemi nel modello Entity Relationship comprendente tutte le basi di dati in ambito esecuzione penale.

Le seguenti ulteriori ontologie sono rappresentate in forma di tassonomie a due livelli:

- Top ontology delle indagini preliminari.
- Top ontology del processo civile.
- Ontologie di dominio del processo civile: bancario, lavoro, danno non patrimoniale da violazione della privacy, separazione giudiziale.
- Ontologia per procedimento penale-cognizione: relazione vittima colpevole.

L'ontologia top dei procedimenti di Giustizia e per la esecuzione penale sono state prodotte mediante reverse engineering da schemi logici. Le ontologie per la relazione vittima colpevole e danno non patrimoniale sono state prodotte da esperti di dominio. Le ontologie per bancario e lavoro sono state prodotte a partire da lessici costruiti con analisi delle sentenze.

² <https://www.nextgenerationupp.unito.it/>

5. Architettura a servizi

Le funzionalità sviluppate adottano per il deployment una architettura a servizi. La macro architettura funzionale multinodo è mostrata nella Figura 2.

I componenti della architettura del singolo nodo (cornice in rosso) sono il Multilayer Ingestion Protocol, l' Access Control & User Management, lo Storage Manager, il Document Component, il Metadata Manager, il Service Manager, l'NLP Service Manager, Analysis, il Front End, il Multilayer Export Protocol.

6. Verso un documentale semantico per la PA

Il sistema documentale semantico descritto nel lavoro è di potenziale utilità per tutte le Pubbliche Amministrazioni. Un progetto di diffusione del sistema dovrebbe prevedere due fasi: una prima fase di parametrizzazione, che riguardi la struttura organizzativa, le ontologie, e i processi primari e di governo e una seconda fase di personalizzazione, vedi Figura 3.

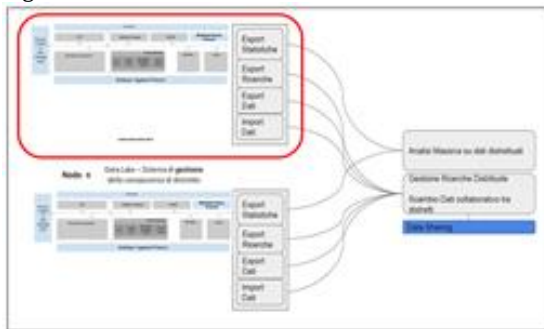


Figura 2 - Architetture multinodo a servizi

7. Verso un documentale semantico per la PA

Il sistema documentale semantico descritto nel lavoro è di potenziale utilità per tutte le Pubbliche Amministrazioni. Un progetto di diffusione del sistema dovrebbe prevedere due fasi: una prima fase di parametrizzazione, che riguardi la struttura organizzativa, le ontologie, e i processi primari e di governo e una seconda fase di personalizzazione, vedi Figura 3. Un tale progetto richiede una governance forte, che si collochi nel percorso strategico di trasformazione digitale del Paese, attualmente in corso di attuazione nel Polo strategico Nazionale. L'inclusione di servizi per un documentale semantico per la PA nella architettura dei servizi del Polo richiederebbe la produzione di una top ontology comune alle PA e la modellazione di alto livello dei processi primari e di governo, con successiva personalizzazione a cura delle singole PA.



Figura 3 - Un documentale semantico generale per la Pubblica Amministrazione italiana

8. Riferimenti principali

- [1] V. Bellandi, C. Bernasconi, F. Lodi, M. Palmonari, R. Pozzi, M. Ripamonti & S. Siccardi (2024). An entity-centric approach to manage court judgments based on Natural Language Processing. *Computer Law & Security Review*, 52.
- [2] A. Campi, S. Ceri, M. Dilettis, B. Pernici, Variants analysis in judicial trials: Challenges and initial results, *ECML PKDD Workshop on Knowledge Discovery and Process Mining for Law (KDPM4LAW)*, Sept. 2023
- [3] CapAI - A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act, <https://artificialintelligenceact.eu/assessment/>
- [4] Manuale di qualità dei dati, documenti, modelli di Giustizia
- [5] B. Pernici, C. Bono, L. Piro, M. Del Treste, G. Vecchi, Improving the analysis of the judiciary performance. The use of data mining techniques to assess the timeliness of civil trials, *International Journal of Public Sector Management*, Vol. 37, N. 1, 2024
- [6] R. Pozzi, R., Rubini, R., Bernasconi, C., M. Palmonari, Named Entity Recognition and Linking for Entity Extraction from Italian Civil Judgements. In *International Conference of the Italian Association for Artificial Intelligence* (pp. 187-201). Cham: Springer Nature Switzerland. Novembre 2023.
- [7] R. Pozzi, R., F. Moiraghi, F. Lodi, M. Palmonari, Evaluation of incremental entity extraction with background knowledge and entity linking. In *Proceedings of the 11th International Joint Conference on Knowledge Graphs* (pp. 30-38). Ottobre 2022.