

Scheda progetto:
KlonDikE@SemAnonim-UniGE
Realizzazione:
Università degli Studi di Genova - DIBRIS

Titolo progetto	KlonDikE@SemAnonim-UniGE								
Obiettivi	Sviluppo di un'applicazione software prototipale per l'estrazione di informazione semantica da sentenze di divorzio, loro anonimizzazione ed estrazione di statistiche.								
Descrizione	KlonDikE@SemAnonim-UniGE individua i dati anagrafici delle parti e dei figli nella sentenza di divorzio, 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.								
Metodologia	<div>1. Identificazione delle necessità di analisi semantica dei documenti giudiziari presso il Tribunale di Genova e identificazione di uno use case di interesse per l'analisi semantica (sentenze di divorzio).</div> <div>2. Studio ed identificazione degli algoritmi e delle tecnologie per l'analisi di documenti in formato testuale e PDF (PyPdf, PyPDF2, PyPDF4, pdftotext e PyMuPdf).</div> <div>3. Studio ed identificazione degli algoritmi e delle tecnologie per Natural Language Processing (NLTK, spaCy e Regex).</div> <div>4. Realizzazione di uno strumento prototipale open source e che operi senza accesso alla rete, in conformità con le Linee guida acquisizione e riuso software PA e con le ulteriori restizioni di accesso a Internet delle macchine collocate nei tribunali.</div>								
Risultati	<div>I risultati del progetto sono descritti nell'articolo "Mining Information from Legal Sentences in KlonDikE" di A. Gatti, V. Mascardi e D. Pellegrini, inviato a NL4AI@AIxIA 2023 e attualmente sotto revisione. A titolo di esempio riportiamo sotto i dati correttamente estratti da una sentenza fittizia e alcuni grafici che visualizzano statistiche su 100 sentenze di divorzio selezionate dalle 100.000 sentenze fornite dal Tribunale di Genova previa convenzione stipulata con il Dipartimento DIBRIS.</div> <div><table><tr><td> Name: Montecchi Romeo Birthdate: 1971-06-02 Birthplace: Verona, VR Sex: M Codice Fiscale: MNTRM071H02L7810 Role: part Note:</td><td> Name: Capuleti Giulietta Birthdate: 1975-09-11 Birthplace: Verona, VR Sex: F Codice Fiscale: CPLGTT75P51L781Z Role: part Note:</td><td> Name: Tebaldo Capuleti Birthdate: 1986-08-21 Birthplace: Cremona, CR Sex: M Codice Fiscale: CPLTLD86M21D150A Role: avv Note:</td><td> Name: Valentino Birthdate: 2007-12-15 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:</td></tr><tr><td> Name: Giovanna Birthdate: 2009-04-07 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:</td><td> Name: Marisa Birthdate: 2012-01-30 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:</td><td> Name: Domenico Pellegrini Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:</td><td> Name: Viviana Mascardi Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:</td></tr></table></div> <div><div><div>Percentage of separations with minor age sons</div><div><div>Sons with minor age</div><div>Not sons with minor age</div></div></div><div><div>Percentage of Domestic Violence</div><div><div>Presence of Violence</div><div>No Violence</div></div></div><div><div>Separations by Age</div><div><div>Frequency</div><div>Age</div></div></div></div>	Name: Montecchi Romeo Birthdate: 1971-06-02 Birthplace: Verona, VR Sex: M Codice Fiscale: MNTRM071H02L7810 Role: part Note:	Name: Capuleti Giulietta Birthdate: 1975-09-11 Birthplace: Verona, VR Sex: F Codice Fiscale: CPLGTT75P51L781Z Role: part Note:	Name: Tebaldo Capuleti Birthdate: 1986-08-21 Birthplace: Cremona, CR Sex: M Codice Fiscale: CPLTLD86M21D150A Role: avv Note:	Name: Valentino Birthdate: 2007-12-15 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:	Name: Giovanna Birthdate: 2009-04-07 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:	Name: Marisa Birthdate: 2012-01-30 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:	Name: Domenico Pellegrini Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:	Name: Viviana Mascardi Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:
Name: Montecchi Romeo Birthdate: 1971-06-02 Birthplace: Verona, VR Sex: M Codice Fiscale: MNTRM071H02L7810 Role: part Note:	Name: Capuleti Giulietta Birthdate: 1975-09-11 Birthplace: Verona, VR Sex: F Codice Fiscale: CPLGTT75P51L781Z Role: part Note:	Name: Tebaldo Capuleti Birthdate: 1986-08-21 Birthplace: Cremona, CR Sex: M Codice Fiscale: CPLTLD86M21D150A Role: avv Note:	Name: Valentino Birthdate: 2007-12-15 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:						
Name: Giovanna Birthdate: 2009-04-07 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:	Name: Marisa Birthdate: 2012-01-30 Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: son Note:	Name: Domenico Pellegrini Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:	Name: Viviana Mascardi Birthdate: None Birthplace: Unknown Sex: U Codice Fiscale: Unknown Role: judge Note:						
Output	SemAnonim-Unige è stato installato sulla macchina di un giudice del Tribunale di Genova per consentire una accurata sperimentazione.								