

Università Degli Studi Di Torino
Master di I Livello in Analisi Dati per
la Business Intelligence e Data Science
A.A. 2017/2018

TITOLO TESI

Analisi della qualità dei dati di ContrattiPubblici.org

Candidato: Sofia Benedetta Rosati

Tutor aziendale: Alessio Melandri

Relatore: Prof.ssa Maria Teresa Giraudo

Abstract

Il progetto di tesi è stato svolto presso Synapta S.r.l., un'azienda tecnologica di Torino specializzata nel campo dei linked data e nello sviluppo di software.

L'obiettivo del progetto è l'analisi del database di ContrattiPubblici.org¹ – database online di tutti i contratti pubblici italiani dal 2013, sviluppato da Synapta S.r.l. - al fine di valutare la qualità dei dati e porre le basi per l'automatizzazione del data cleaning degli stessi.

La tesi si articola in due principali capitoli, oltre le conclusioni. Nel primo capitolo si illustrano: la provenienza dei dati nell'ambito di applicazione della legge 6.11.2012, n. 190 "*Disposizioni per la prevenzione e la repressione della corruzione e dell'illegalità nella Pubblica Amministrazione*"; ContrattiPubblici.org e l'azienda; la struttura del database e dei dati e infine il software e le tecniche utilizzate. Nel secondo capitolo si presenta passo passo tutto il percorso di studio fatto, presentando i risultati di ciascuna fase.

¹ www.contrattipubblici.org