

PROGETTI CONCLUSI 2018

Titolo progetto	Web mining per integrare e validare informazioni del registro statistico delle imprese attive ASIA.
Descrizione	Sperimentazione dell'utilizzo di nuove metodologie di trattamento di dati provenienti dal web, al fine di una loro possibile integrazione con le fonti amministrative e statistiche che alimentano il registro statistico delle imprese. L'uso di questa nuova fonte per l'aggiornamento del registro risponde ad una domanda crescente di dati tempestivi, aggiornati ed interconnessi a livello multidimensionale.
Obiettivi	<p>Impiego di tecniche di Web Mining per:</p> <ul style="list-style-type: none"> i) aggiornare/completare informazioni presenti nel Registro; ii) convalidare informazioni del Registro; iii) aggiungere nuove informazioni nel Registro anche individuando tassonomie di impresa diverse dalle classificazioni ufficiali, a supporto di una migliore interpretazione dei nuovi fenomeni economici emergenti.
Metodologia	<p>Il progetto è stato uno dei primi a partecipare al LabInn. In questo modo la sperimentazione ha beneficiato delle infrastrutture messe a disposizione per testare le metodologie in uno spazio dedicato.</p> <p>Principale fattore del successo finora ottenuto è stato il contesto multidisciplinare, hanno partecipato e lavorato fianco a fianco gli esperti metodologici (MEA - G. Bianchi, F. Scalfati; MEC - D. Summa) insieme a quelli statistici-tematici (SEA - M. Consalvi, B. Gentili, F. Pancella, C. Viviano) ed informatici (DCIT – M. Amarone, D. Aprile).</p> <p>E' stata attuata una strategia complessa che si compone di tre fasi principali: <i>web scraping</i>, <i>text mining</i> e <i>machine learning</i>; ed è conforme ai criteri di autonomia e indipendenza dell'istituto, in quanto è preservato il controllo sul processo e sui dati, compreso l'accesso ai dati internet, senza l'uso di servizi di terze parti.</p> <p>La metodologia ed i risultati ottenuti sono stati presentati alla 13^a Conferenza nazionale di statistica e al Wiesbaden Group on Business Registers, un gruppo di esperti internazionali sotto l'egida della Commissione statistica delle Nazioni Unite impegnato nell'ulteriore sviluppo dei registri delle imprese e dei frames di indagine:</p> <p>http://www.wiesbaden2018.bfs.admin.ch/wp-content/uploads/2018/08/Paper_Bianchi.pdf</p>
Risultati ottenuti	<p>La realizzazione del processo di data mining fornisce maggiore efficacia ed efficienza alle statistiche ufficiali, attraverso un processo di trasformazione del modello di produzione dei dati di tipo <i>register-based</i>. Inoltre, rende visibile, attraverso un approccio <i>data driven</i>, una nuova dimensione delle imprese, ad oggi non nota, risultato di un'attività di analisi e predizione di fenomeni nascosti.</p> <p>Il risultato ottenuto nella fase di sperimentazione riguarda un prototipo di circa 100mila imprese, ma l'obiettivo di lungo termine di questo lavoro è la predisposizione di una piattaforma di informazioni organizzate per area tematica, un repository di dati che va ad affiancare i registri statistici ed i dati dell'indagine ed è accessibile dalle rilevazioni di quel dominio tematico.</p> <p>Le informazioni saranno, quindi, a utilizzo e beneficio trasversale nell'ottica di favorire l'integrazione dei dati e qualsiasi nuova attività di analisi volta a produrre le statistiche necessarie.</p>