**PROPOSTA DI TESI PRESSO L’AZIENDA ASSIOMA.NET CON TIROCINIO**

# Tesi 1 - Qualità semantica di dati non strutturati (comunicazione aziendale)

In aggiunta alla verifica della qualità dei dati prodotti dalle applicazioni di business dell’azienda è opportuno estendere l’analisi verso le attività relative alla comunicazione aziendale B2B e B2C.

La qualità della comunicazione si concentra sul fatto che il significato di un messaggio sia chiaro. La sola formattazione corretta dei dati non è sufficiente. Se il significato non è chiaro, la comunicazione aziendale non sarà compresa correttamente.

In altre parole, è necessaria la “Qualità Semantica” (non solo la Qualità dei Dati) le cui dimensioni derivano dalla responsabilità dell’emittente della comunicazione nell’inviare l’informazione.

In questo senso il messaggio dovrebbe essere: leggibile, comprensibile, preciso, affidabile, utile e sufficiente.

Obiettivo della tesi è definire un modello generale, ampiamente applicabile per la misurazione della qualità semantica dei dati non strutturati che formano i messaggi, le comunicazioni aziendali; individuare un modello concettuale per esprimere le regole aziendali; formulare un vocabolario che le rispecchi e che possa essere utilizzato per validare le “qualità semantiche” delle comunicazioni in uscita dall’azienda.

# Tesi 2 - Data Quality dei flussi finanziari temporali.

Nei settori *data driven*, le decisioni si basano su applicazioni di analisi che elaborano flussi continui di dati. Tipicamente, i dati sono caratterizzati da un segnale statistico che li rappresenta all'interno di alcuni limiti d’incertezza. Le applicazioni di analisi si basano su questo segnale per fare predizioni e previsioni, creare report, cruscotti e visualizzazioni.

I problemi di Data Quality come dati mancanti (lacune), incompleti o duplicati (picchi) interferiscono con questo segnale distorcendo le informazioni e portando a risultati di analisi fuorvianti e a decisioni potenzialmente sbagliate.

Obiettivo della tesi è definire un modello per la misurazione della qualità dei dati dei flussi e che tenga conto della natura evolutiva dei flussi stessi. Questo sistema consentirà di calcolare un indice di affidabilità delle fonti in termini di qualità dei dati rispetto a specifici vincoli e misurazioni del sistema.

Nell’ambito del progetto, si andranno a classificare le anomalie definendo non meno di tre tipi di vincoli e, utilizzando la distorsione statistica come metrica per misurare la Data Quality, si identificheranno le violazioni che potrebbero essere potenziali difetti.

**Per informazioni gli interessati sono invitati a contattare i docenti di riferimento:**

*Barbara Pernici barbara.pernici@polimi.it*

*Cinzia Cappiello cinzia.cappiello@polimi.it*