

Inserzioni pubblicitarie de “Il Resto del Carlino” 1939 – 45



Lo Scopo

- Creare una base dati per utile **all'analisi comparativa di tendenze socio-economiche storiche** attraverso la valutazione di inserzioni a stampa su quotidiani storici
- Rendere fruibile e facilmente analizzabile un **dataset ampio, omogeneo ed organico, di inserzioni pubblicitarie**

Le Fonti

- Corpus complessivo composto da più di **2000 uscite de «Il Resto del Carlino»** nell'intervallo **01/09/1939 – 27/02/1945**.
- Digitalizzazioni ad alta risoluzione messe a disposizione **dall'archivio digitale di Storia e Memoria di Bologna ***

CINECA

* <https://www.storiaememoriadibologna.it>

L'elaborazione

- A partire dalle digitalizzazioni, si è analizzato il **layout del quotidiano** identificando:
 - Titoli
 - Testi degli articoli
 - Fotografie
 - Grafici e disegni
 - Annunci
- Ogni **annuncio** è stato corredato dalla sua **descrizione e trascrizione**.

Layout Analysis

- Lo strumento **open** allo stato dell'arte identificato, LayoutParser, seppur performante, poco si adatta all'analisi di quotidiani storici italiani.
- Si è sviluppata una **piattaforma** per aiutare gli esperti di dominio nella creazione di un dataset annotato, base che si è utilizzata per il **fine-tuning** dello strumento.

Entrambi gli strumenti sono a disposizione fra i servizi richiedibili all'interno di DPaaS:

- *Piattaforma: LWP Data Preparation*
- *Modello specializzato: lp_carlino40*

Descrizione e Trascrizione

- Ogni annuncio identificato è stato corredato da:
 - Descrizione
 - Trascrizione
- Per entrambi i task è stato utilizzato **Qwen**

Dataset

- Il dataset è composto da tutti gli annunci identificati all'interno del quotidiano
- Sono presenti tutti i **dettagli dell'annuncio** come: nome del quotidiano, data e numero di pagina, posizione (in pixel) dell'annuncio
- Ogni annuncio è corredato dalla sua descrizione e trascrizione

Il Dataset in numeri

- Annunci estratti dalle uscite de «Il Resto del Carlino» nel periodo 1939-45
 - più di **2'000** uscite
 - circa **100'000** annunci identificati