

Il GenAI Center of Competence  
di Sistemi Informativi

presenta

Processi di OCR  
aumentati da GenAI  
per estrazione di testo da immagini

# Processi di OCR aumentati da GenAI per estrazione di testo da immagini

Giuseppe Santocchi, Edoardo Croce

January 28, 2025

## Contents

<b>1 Introduzione</b>	<b>2</b>
1.1 Contesto dello studio . . . . .	2
1.2 Ruolo della GenAI nei processi OCR . . . . .	2
<b>2 Obbietivo della ricerca</b>	<b>2</b>
<b>3 Principali problematiche affrontate</b>	<b>3</b>
<b>4 Tecnologie Individuate</b>	<b>3</b>
4.1 Paddle OCR . . . . .	3
4.2 Llama 3.1 8b . . . . .	4
4.3 Llama 3.3 70b . . . . .	4
4.4 MongoDB . . . . .	4
<b>5 Confronto tra approcci OCR</b>	<b>5</b>
5.1 Risultati di Tesseract OCR . . . . .	5
5.2 Miglioramenti osservati con Paddle OCR . . . . .	5
5.3 Impatto sulla qualità dei dati estratti . . . . .	6
5.4 Perché abbiamo optato per Paddle OCR . . . . .	6
<b>6 Step by Step Overview</b>	<b>6</b>
6.1 Estrazione del testo da caricamento immagine con selezione di categoria di appartenenza e rispettive entità da estrarre . . . . .	7
6.1.1 Upload dell'immagine . . . . .	7
6.1.2 Riconoscimento del testo tramite OCR (Paddle OCR) . . . . .	8
6.1.3 Elaborazione del testo estratto . . . . .	8
6.1.4 Output del testo elaborato . . . . .	8
6.2 Estrazione del testo con solo caricamento dell'immagine . . . . .	8
6.2.1 Upload dell'immagine . . . . .	8
6.2.2 Riconoscimento del testo tramite OCR (Paddle OCR) . . . . .	8
6.2.3 Chiamata al servizio GenAI per il recupero della categoria . . . . .	8
6.2.4 Verifica della categoria . . . . .	8
6.2.5 Chiamata al servizio GenAI per l'elaborazione del testo . . . . .	9
6.2.6 Output del testo elaborato . . . . .	9
<b>7 Risultati</b>	<b>9</b>
7.1 Risultati sperimentali . . . . .	9
7.2 Valutazione delle performance . . . . .	9
7.2.1 Affidabilità delle categorie . . . . .	9
7.2.2 Tempi di risposta e variabilità dovuta alle risorse hardware . . . . .	10
<b>8 Conclusioni e sviluppi futuri</b>	<b>10</b>
8.1 Considerazioni finali . . . . .	10
8.2 Possibili miglioramenti al sistema . . . . .	10

## 1 Introduzione

L'analisi delle immagini e l'estrazione del testo da esse rappresentano una necessità fondamentale in molti settori, come la gestione documentale e i processi amministrativi avanzati. La crescente diffusione dell'Intelligenza Artificiale (AI) e dell'Intelligenza Artificiale Generativa (GenAI) ha favorito lo sviluppo di strumenti sempre più sofisticati ed efficienti, migliorando l'accuratezza e l'affidabilità nei flussi di elaborazione delle immagini.

### 1.1 Contesto dello studio

In questo studio, ci siamo concentrati sulla creazione di una pipeline che combini tecnologie consolidate, come i modelli OCR, con soluzioni basate su GenAI per migliorare i risultati finali, soprattutto in contesti complessi in cui gli approcci tradizionali non riescono a garantire performance soddisfacenti.

L'obiettivo principale era quello di affrontare i limiti dei modelli OCR tradizionali e di esplorare come l'integrazione con GenAI possa offrire risultati migliori in termini di accuratezza, leggibilità e flessibilità.

### 1.2 Ruolo della GenAI nei processi OCR

Mentre i Modelli Multimodali di ultima generazione offrono funzionalità avanzate di analisi e generazione di immagini, il loro utilizzo presenta delle limitazioni, tra cui costi elevati, tempi di elaborazione non sempre ottimali e scarsa compatibilità con alcune piattaforme.

Una valida alternativa è rappresentata dai modelli OCR tradizionali, che consentono di estrarre testo dalle immagini in modo rapido ed economico. Tuttavia, questi modelli spesso restituiscono output grezzi, non sempre affidabili o pronti all'uso, specialmente quando si affrontano scenari complessi come documenti con formattazioni variabili, linguaggi differenti o condizioni visive non ottimali.

Per superare tali limiti, abbiamo integrato le capacità dei Large Language Models (LLM) nel processo, utilizzandoli per correggere, normalizzare e arricchire i risultati forniti dall'OCR. Questa combinazione consente di:

- Migliorare la qualità e l'accuratezza del testo estratto.
- Adattare i risultati alle esigenze specifiche del contesto applicativo.
- Aumentare la flessibilità della pipeline di elaborazione.

Questo approccio dimostra come la GenAI possa essere utilizzata per potenziare tecnologie tradizionali, creando soluzioni innovative e scalabili anche per operazioni come l'estrazione di testo da immagini.

## 2 Obiettivo della ricerca

Questo studio si propone di esplorare, approfondire e sviluppare un approccio innovativo per migliorare l'output dei modelli OCR tradizionali. L'obiettivo principale è quello di affrontare le problematiche legate alla qualità e all'affidabilità del testo estratto, introducendo metodologie e strumenti avanzati che combinano le capacità di preelaborazione delle immagini con l'integrazione di Large Language Models (LLM). Il focus della ricerca si articola nei seguenti punti:

1. **Valutare l'efficacia dei modelli OCR tradizionali** rispetto a scenari complessi, identificandone i limiti e le aree di miglioramento.
2. **Definire un flusso di elaborazione avanzato** che riduca la necessità di preelaborazione delle immagini, sfruttando modelli OCR moderni come Paddle OCR.
3. **Integrare LLM per la riorganizzazione e standardizzazione del testo estratto**, garantendo una maggiore leggibilità, coerenza e adattabilità rispetto al contesto applicativo.
4. **Esplorare il confronto tra diversi modelli OCR** (ad esempio Tesseract e Paddle OCR) per evidenziare i benefici e i trade-off dell'approccio scelto.
5. **Proporre una pipeline scalabile e riproducibile**, che possa essere facilmente adattata a scenari e domini diversi.

Attraverso questa ricerca, si intende dimostrare come la combinazione di tecnologie OCR e LLM possa rappresentare una soluzione efficace per superare i limiti delle tecnologie tradizionali, offrendo una maggiore precisione e una flessibilità superiore nei flussi di estrazione di testo da immagini.

### 3 Principali problematiche affrontate

L'estrazione di testo da immagini è un processo che presenta diverse sfide tecniche e operative, soprattutto se si punta a ottenere un risultato che sia al tempo stesso accurato, **replicabile** e **standardizzabile**. Di seguito, vengono evidenziate le principali problematiche affrontate nel corso della ricerca:

- I. **Qualità dell'immagine in input** - La qualità dell'immagine influisce in modo determinante sull'accuratezza del modello OCR. Immagini acquisite con esposizioni alla luce non ottimali, angolazioni scorrette o rumore visivo possono ridurre significativamente la capacità del modello di estrarre correttamente il testo. Questa variabilità rende complessa la gestione uniforme di input provenienti da fonti diverse.
- II. **Struttura del testo estratto** - Il testo ottenuto tramite OCR è spesso grezzo e privo di una struttura chiara, rendendolo inadatto a processi successivi. Questo rappresenta un ostacolo per attività come:
  - Il salvataggio di informazioni chiave.
  - L'applicazione di flussi decisionali automatizzati.
  - L'analisi strutturata del contenuto testuale.
- III. **Personalizzazione eccessiva del flusso** - La creazione di un flusso standard per l'estrazione di informazioni chiave dal testo OCR può richiedere un elevato grado di personalizzazione a livello di codice. Ad esempio, l'uso di espressioni regolari (regex) specifiche per gestire formati particolari penalizza il riutilizzo della codebase. Ciò rende difficile applicare lo stesso approccio a diversi tipi di documenti o fonti.
- IV. **Flessibilità e riutilizzabilità del flusso** - Una delle principali sfide è progettare un sistema che, con modifiche minime a livello di configurazione, sia in grado di:
  - Estrarre informazioni da immagini di formati diversi.
  - Riconoscere e gestire entità variabili in base al contesto.
  - Garantire un output prevedibile e di alta qualità.

Questa flessibilità deve essere bilanciata con la necessità di mantenere un'elevata affidabilità e precisione, evitando la perdita di coerenza nei risultati.

### 4 Tecnologie Individuate

Per affrontare le problematiche evidenziate nel capitolo precedente, sono stati individuati strumenti specifici che combinano i vantaggi offerti dai modelli OCR con la potenza dei Large Language Models (LLM). Questo approccio ibrido consente di superare le limitazioni dei classici sistemi di estrazione di testo, introducendo una fase di rielaborazione, correzione e standardizzazione delle informazioni estratte. L'obiettivo è garantire un output di alta qualità, configurabile dinamicamente e riutilizzabile per diversi scenari operativi.

#### 4.1 Paddle OCR

Paddle OCR è stato scelto come modello base per l'estrazione di testo dalle immagini grazie alla sua efficienza e flessibilità. Tra le caratteristiche principali:

- Supporto per più lingue, con la possibilità di utilizzare dati pre-addestrati specifici.
- Ottime performance anche su immagini con rumore visivo o formati non ottimali.

- Integrazione con strumenti di pre-processing (anche se, nel nostro caso, l'ottimizzazione delle immagini è stata omessa per risultati migliori).

Paddle OCR si è dimostrato particolarmente adatto per la gestione di immagini complesse, riducendo la necessità di interventi manuali nella fase di elaborazione.

## 4.2 Llama 3.1 8b

Llama 3.1 8b è stato utilizzato nelle prime fasi di test e prototipazione dello studio, grazie alla sua bassa complessità computazionale e alla facilità di configurazione per adattarsi a diversi contesti di utilizzo. Le sue caratteristiche principali includono:

- Capacità di processare e riorganizzare il testo estratto, correggendo errori comuni del modello OCR.
- Bassa complessità computazionale, che lo rende ideale per scenari con risorse limitate.
- Configurabilità che permette di adattarsi a vari contesti senza compromettere le performance.

Nonostante la sua semplicità rispetto a modelli più avanzati, Llama 3.1 8b ha fornito risultati promettenti nelle prime iterazioni della ricerca, soprattutto grazie al corretto prompt engineering. Questo approccio ha permesso di ottenere un'accuratezza maggiore nell'interpretazione e rielaborazione del testo estratto, ottimizzando le performance del modello nelle fasi iniziali.

## 4.3 Llama 3.3 70b

Llama 3.3 70b rappresenta una versione aggiornata del precedente Llama 3.1 8b. Le sue principali caratteristiche sono:

- Supporto per lingue multiple e un'ampia finestra contestuale (128K), il che consente di gestire testi più complessi e ampi.
- Uso di strumenti avanzati per una migliore elaborazione e capacità di ragionamento, che risulta in una qualità superiore del testo elaborato.
- Velocità di elaborazione e risposta ridotte, che permettono di ottenere risultati migliori in un tempo contenuto, pur mantenendo un costo computazionale inferiore rispetto a modelli ancora più grandi come llama3.1 405b.

Queste caratteristiche rendono Llama 3.3 70b la scelta ideale per flussi di lavoro complessi e ad alte prestazioni, dove è necessario un livello superiore di precisione e velocità nell'elaborazione del testo.

## 4.4 MongoDB

MongoDB è stato scelto come database per la sua flessibilità e capacità di gestire dati non strutturati o semi-strutturati, caratteristici dei documenti utilizzati in questo studio. Ecco i principali vantaggi che hanno portato a questa scelta:

### I. Flessibilità del modello dei dati:

- MongoDB utilizza un modello di dati basato su documenti JSON-like, che consente di rappresentare informazioni in modo modulare, senza la necessità di strutture fisse e rigide.

### II. Scalabilità orizzontale:

- MongoDB supporta la scalabilità orizzontale, permettendo al sistema di gestire grandi volumi di dati senza compromettere le prestazioni.

### III. Interrogazioni flessibili:

- Le potenti funzionalità di interrogazione consentono di cercare i dati in modo rapido ed efficiente.

#### IV. Integrazione con i microservizi:

- MongoDB si integra facilmente con i microservizi grazie ai driver nativi e alle librerie come Mongoose, utilizzata per definire modelli e gestire la comunicazione tra i servizi e il database.

#### V. Gestione di dati gerarchici:

- MongoDB consente di organizzare e gestire gerarchie di dati complesse senza dover ricorrere a strutture relazionali complicate.

Grazie a queste caratteristiche, MongoDB si rivela un database versatile e potente, in grado di rispondere a una vasta gamma di esigenze di gestione dei dati in applicazioni moderne.

## 5 Confronto tra approcci OCR

Il riconoscimento ottico dei caratteri (OCR) è una tecnologia fondamentale per l'estrazione automatica del testo da immagini. Due dei modelli più utilizzati nel nostro studio sono Tesseract OCR e Paddle OCR. Qui, esamineremo i risultati ottenuti con ciascun sistema, i miglioramenti osservati nell'uso di Paddle OCR e l'impatto sulla qualità dei dati estratti.

### 5.1 Risultati di Tesseract OCR

Tesseract è uno degli OCR open-source più conosciuti e usati nel panorama del riconoscimento dei caratteri. Si tratta di una libreria che, nonostante sia molto versatile e configurabile, ha alcuni limiti nei contesti di estrazione da immagini complesse.

- **Precisione nei documenti chiari:** Tesseract fornisce buoni risultati su documenti con una qualità dell'immagine elevata e testo ben definito. Nei casi di documenti con una risoluzione alta, testo stampato chiaro e formattato correttamente, Tesseract è in grado di restituire buoni risultati, con una precisione elevata.
- **Problemi con documenti di bassa qualità:** Tesseract soffre nei casi in cui l'immagine ha rumore di fondo, distorsioni, o una qualità inferiore. In presenza di immagini con sfondi complessi, caratteri deformati o inclinati, i risultati di Tesseract tendono a essere imprecisi, con errori nei caratteri riconosciuti o perdita di informazioni.
- **Errore nella gestione di testi complessi:** In presenza di formati complessi come ricevute, fatture o documenti con layout non standardizzati, Tesseract potrebbe non essere in grado di identificare correttamente tutti i campi di interesse (come importi, numeri di fattura, date), riducendo la qualità dei dati estratti.

### 5.2 Miglioramenti osservati con Paddle OCR

Paddle OCR è un sistema OCR più recente e avanzato rispetto a Tesseract, sviluppato come parte del framework AI di PaddlePaddle. È progettato per offrire prestazioni superiori rispetto a Tesseract, soprattutto su immagini complesse o di bassa qualità.

- **Maggiore robustezza su immagini di bassa qualità:** Uno dei principali vantaggi di Paddle OCR è la sua capacità di lavorare bene anche su immagini sfocate, con rumore di fondo, caratteri inclinati o distorti. Grazie a una rete neurale profonda e tecniche avanzate di pre-elaborazione delle immagini, Paddle OCR riesce a gestire questi problemi con maggiore efficacia rispetto a Tesseract.
- **Maggiore accuratezza nelle strutture complesse:** Paddle OCR si distingue anche nell'elaborazione di testi complessi, come fatture, contratti e ricevute, dove la disposizione del testo è meno lineare. Il modello è in grado di riconoscere e separare meglio le diverse sezioni (ad esempio, intestazioni, tabelle, campi di importo) migliorando la precisione complessiva dell'estrazione.
- **Capacità di riconoscere caratteri e simboli particolari:** Paddle OCR ha una migliore capacità di riconoscere caratteri speciali, simboli monetari e segni non standard, migliorando la qualità complessiva del testo estratto.

### 5.3 Impatto sulla qualità dei dati estratti

L'utilizzo di Paddle OCR ha avuto un impatto positivo significativo sulla qualità dei dati estratti. Questo si riflette in vari aspetti, che giustificano la decisione di passare a questo sistema, anziché continuare con Tesseract:

- **Maggiore precisione nei dati estratti:** Con Paddle OCR, la qualità dell'estrazione del testo è sensibilmente migliore, con minori errori di riconoscimento dei caratteri. Ciò è particolarmente evidente nei documenti complessi, dove Paddle riesce a estrarre dati come numeri di fattura, date e importi con maggiore affidabilità rispetto a Tesseract.
- **Riconoscimento migliore di categorie e entità:** Grazie a una gestione superiore delle immagini e alla sua capacità di interpretare layout complessi, Paddle OCR consente di estrarre dati significativi per la categorizzazione e l'analisi contestuale dei documenti (come la riconoscibilità di informazioni chiave come numeri di documento, totale, indirizzi, ecc.).
- **Maggiore velocità nel recupero di informazioni precise:** In contesti aziendali dove l'efficienza è cruciale, Paddle OCR ha dimostrato una maggiore velocità nell'estrazione del testo rispetto a Tesseract, grazie a un sistema di elaborazione più ottimizzato e preciso. Questo comporta una riduzione dei tempi di elaborazione, migliorando l'efficienza complessiva del flusso di lavoro.

### 5.4 Perché abbiamo optato per Paddle OCR

La decisione di optare per Paddle OCR è stata presa in seguito a una valutazione approfondita dei risultati ottenuti con Tesseract e delle necessità specifiche della nostra ricerca. Le principali ragioni per la scelta di Paddle OCR sono:

- **Maggiore precisione e affidabilità:** Con Paddle OCR, la qualità dell'estrazione del testo è sensibilmente migliore, con minori errori di riconoscimento dei caratteri. Ciò è particolarmente evidente nei documenti complessi, dove Paddle riesce a estrarre dati come numeri di fattura, date e importi con maggiore affidabilità rispetto a Tesseract.
- **Adattabilità a documenti complessi:** Grazie a una gestione superiore delle immagini e alla sua capacità di interpretare layout complessi, Paddle OCR consente di estrarre dati significativi per la categorizzazione e l'analisi contestuale dei documenti (come la riconoscibilità di informazioni chiave come numeri di documento, totale, indirizzi, ecc.).
- **Supporto a una migliore elaborazione semantica:** Paddle OCR ha estratto dei risultati migliori per lavorare in combinazione con modelli avanzati di Intelligenza Artificiale (come Llama), permettendo di sfruttare tecniche di elaborazione semantica e NLP per estrarre informazioni più precise e contestualizzate dai documenti.

In sintesi, sebbene Tesseract sia un OCR valido per documenti chiari e ben formattati, Paddle OCR ha mostrato risultati superiori per quanto riguarda l'accuratezza e l'affidabilità, rendendolo la scelta preferita per il nostro studio. La sua capacità di gestire immagini di bassa qualità, formati complessi e la sua velocità nell'elaborazione del testo sono stati fattori decisivi per il miglioramento complessivo della qualità dei dati estratti.

## 6 Step by Step Overview

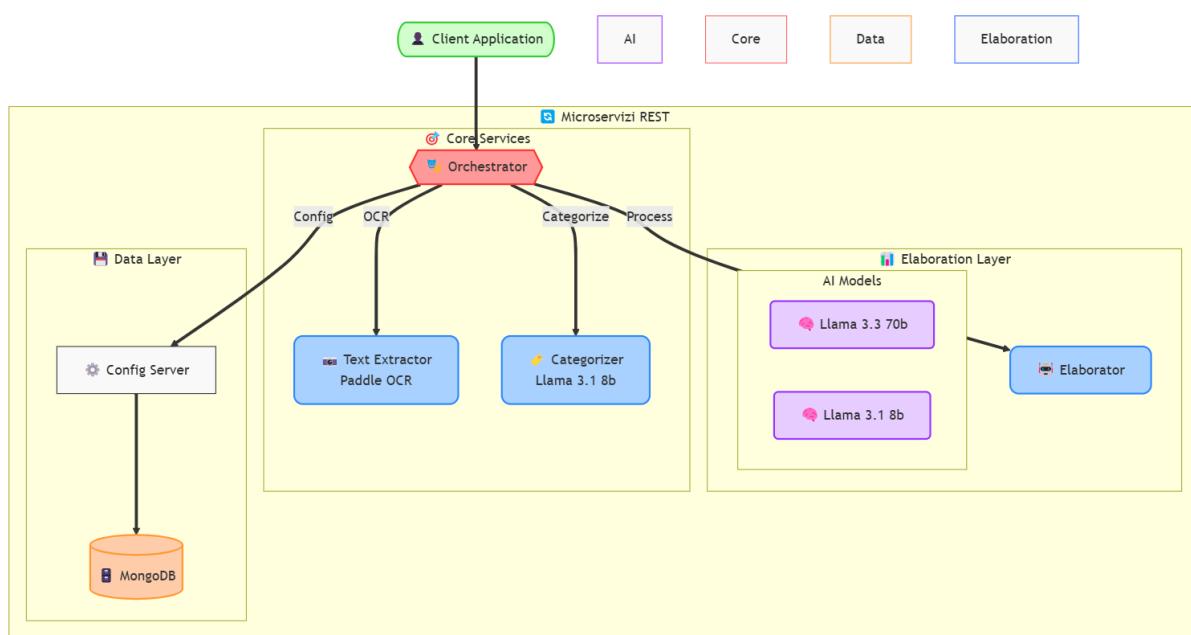
Il flusso di elaborazione per l'estrazione del testo da un'immagine si articola in due principali modalità operative:

- I. Estrazione del testo da caricamento immagine con selezione di categoria di appartenenza e rispettive entità da estrarre
- II. Estrazione del testo con solo caricamento dell'immagine

L'infrastruttura del sistema si basa su una combinazione di microservizi REST e un database MongoDB, progettata per ottimizzare la gestione dei dati e massimizzare l'efficienza e la precisione dell'elaborazione. I principali servizi individuati includono:

- **categorizer:** Si occupa di determinare la categoria del documento in base al testo estratto dall'immagine, utilizzando modelli di intelligenza artificiale.
- **orchestrator:** Coordina l'interazione tra i vari servizi, assicurando che ogni fase del processo sia eseguita in modo corretto e sequenziale.
- **elaborator:** Responsabile dell'elaborazione avanzata del testo estratto, arricchendolo e strutturandolo in formato JSON attraverso l'AI per utilizzi successivi.
- **text-extractor:** Integra Paddle OCR per il riconoscimento del testo nelle immagini, fornendo un output di alta qualità in termini di accuratezza e precisione.
- **config-server:** Gestisce la connessione e l'interazione con MongoDB.

Il database MongoDB, elemento chiave dell'architettura, contiene i riferimenti, le chiavi, e gli esempi per le possibili categorie coinvolte nell'elaborazione delle immagini.



## 6.1 Estrazione del testo da caricamento immagine con selezione di categoria di appartenenza e rispettive entità da estrarre

Questa modalità consente all'utente di specificare manualmente la categoria del documento caricato, migliorando la precisione dell'estrazione grazie a un targeting mirato delle entità rilevanti. Questo flusso assicura che l'estrazione del testo sia non solo accurata, ma anche ben strutturata, pronta per essere utilizzata in diversi contesti aziendali.

### 6.1.1 Upload dell'immagine

- L'utente carica l'immagine tramite il servizio orchestrator, che agisce come punto centrale per gestire il flusso di lavoro.
- Il buffer dell'immagine viene trasmesso al servizio paddle, specializzato nell'elaborazione OCR.

### 6.1.2 Riconoscimento del testo tramite OCR (Paddle OCR)

- Il servizio paddle, basato sul framework Paddle OCR, riconosce ed estrae il testo presente nell'immagine.
- Il testo estratto viene restituito all'orchestrator, che lo inoltra per ulteriori elaborazioni.

### 6.1.3 Elaborazione del testo estratto

- Il servizio elaborator riceve il testo estratto e utilizza la categoria scelta dall'utente per guidare l'elaborazione.
- Interrogando il database MongoDB, il servizio config-server recupera chiavi, descrizioni ed esempi specifici della categoria selezionata. Questi dati vengono utilizzati per compilare dinamicamente un prompt ottimizzato per l'interazione con il modello Llama 3.1 8b.
- L'elaborazione si concentra sull'estrazione di entità specifiche, come ad esempio numeri di fattura, date o importi ecc., assicurando che il risultato sia accurato e contestualizzato.

### 6.1.4 Output del testo elaborato

- Il testo elaborato viene formattato in JSON e restituito all'utente tramite l'orchestrator.
- Questo output è pronto per essere integrato in altri sistemi aziendali o archiviato per usi futuri.

## 6.2 Estrazione del testo con solo caricamento dell'immagine

In questa modalità, il sistema identifica automaticamente la categoria del documento caricato, utilizzando i modelli GenAI per analisi semantiche e contestuali avanzate.

### 6.2.1 Upload dell'immagine

- L'utente carica un'immagine contenente il testo tramite il servizio orchestrator.
- Il buffer dell'immagine viene trasmesso al servizio paddle per il riconoscimento OCR.

### 6.2.2 Riconoscimento del testo tramite OCR (Paddle OCR)

- Paddle estrae il testo dall'immagine e lo restituisce all'orchestrator.
- Se il testo estratto è vuoto o non valido, il sistema restituisce un errore e l'operazione termina.

### 6.2.3 Chiamata al servizio GenAI per il recupero della categoria

- Il testo estratto viene inviato ed esaminato dal servizio categorizer, che utilizza il modello Llama 3.1 8b per determinare la categoria del documento in base a un'analisi semantica e contestuale.
- Il config-server interroga il database MongoDB per verificare se la categoria estratta è presente nelle categorie disponibili nella collection.

### 6.2.4 Verifica della categoria

- Se il modello riconosce una categoria valida e presente nel database, il sistema prosegue con l'elaborazione del documento.
- Se la categoria non è identificata, il sistema utilizza comunque il servizio elaborator con il modello Llama 3.3 70b per estrarre le entità chiave dal testo.

### 6.2.5 Chiamata al servizio GenAI per l'elaborazione del testo

- Una volta identificata la categoria, il servizio elaborator si occupa di elaborare il testo utilizzando le stesse operazioni svolte nella prima modalità e le informazioni della categoria recuperate dal database.
- Se invece il testo non appartiene a una categoria predefinita, il modello Llama 3.3 70b viene utilizzato per eseguire un'analisi diretta e approfondita ed estrarre le entità chiave. Si utilizza quello da 70b perché essendo più potente, è in grado di gestire testi con minori informazioni contestuali e restituire un'analisi più accurata.

### 6.2.6 Output del testo elaborato

- Una volta completata l'elaborazione, il testo elaborato viene formattato in formato JSON dall'elaborator e restituito tramite l'orchestrator.
- L'output è pronto per essere archiviato o integrato in altri sistemi aziendali.

## 7 Risultati

### 7.1 Risultati sperimentali

I risultati sperimentali ottenuti dimostrano l'efficacia del sistema implementato nell'estrazione automatica del testo e nella categorizzazione dei documenti. Durante le prove su un dataset eterogeneo, composto da fatture, ricevute, patenti e altri documenti con layout variabili e fissi, il sistema ha raggiunto i seguenti risultati:

- **Categorizzazione affidabile:** Il sistema ha dimostrato un'alta affidabilità nella classificazione dei documenti, valutata attraverso feedback manuali su un set di test diversificato.
- **Alta precisione nell'OCR:** I testi estratti dai documenti analizzati presentano un grado di accuratezza elevato, rendendo il sistema adatto a scenari di utilizzo pratico.

### 7.2 Valutazione delle performance

La valutazione delle performance del sistema è stata condotta analizzando principalmente l'affidabilità delle categorie. I risultati ottenuti sono riportati di seguito.

#### 7.2.1 Affidabilità delle categorie

L'affidabilità delle categorie è stata calcolata come percentuale di feedback positivi rispetto al totale dei feedback (positivi e negativi). I risultati per ciascuna categoria sono i seguenti:

- **Ricevuta:** l'affidabilità si attesta intorno all'85% e il 90%, con alcuni margini di errore legati alla variabilità dei formati delle ricevute, che possono influire sulla precisione del sistema in alcuni casi.
- **Fattura:** l'affidabilità si aggira tra l'88% e il 91%. Questa categoria ha mostrato buone prestazioni grazie alla struttura abbastanza standardizzata delle fatture, ma alcuni formati complessi hanno causato imprecisioni.
- **Carta d'identità:** l'affidabilità è molto alta, tra il 96% e il 97%. Essendo un documento con layout fisso e ben definito, il sistema ha ottenuto ottimi risultati nella sua classificazione.
- **Patente:** l'affidabilità varia tra il 91% e il 93%. La patente ha un layout abbastanza standard, ma alcuni documenti potrebbero contenere variazioni che riducono lievemente la precisione.
- **Tessera Sanitaria:** l'affidabilità è compresa tra il 95% e il 97%. La tessera sanitaria, essendo un documento con un formato uniforme, ha permesso al sistema di raggiungere un'alta precisione.

Questi valori evidenziano che le categorie “Carta d’identità” e “Tessera Sanitaria” hanno ottenuto le migliori performance, con un’affidabilità compresa tra il 95% e il 97%, seguite dalla categoria “Patente”, che ha raggiunto un’affidabilità tra il 91% e il 93%. Le categorie “Fattura” e “Ricevuta” hanno mostrato performance più variabili, con valori compresi rispettivamente tra l’88% e il 91% per le “Fatture” e tra l’85% e il 90% per le “Ricevute”.

Le categorie selezionate per lo studio sono state scelte tenendo conto della loro rilevanza nei contesti applicativi in cui il sistema può essere impiegato, come la gestione documentale aziendale o i servizi pubblici. Si è optato per documenti comunemente utilizzati, quali fatture, ricevute e tessere sanitarie, che rappresentano scenari concreti e frequenti. Inoltre, la scelta di includere documenti con layout fissi, come la carta d’identità e la patente, ha permesso di testare la robustezza del sistema anche su formati più standardizzati. Questo approccio ha consentito di ottenere risultati significativi, pur limitando il campo di applicazione del sistema a queste specifiche categorie.

In sintesi, l’analisi ha dimostrato che il sistema è particolarmente efficace nelle categorie selezionate, ma ulteriori studi potrebbero estendere l’applicazione a tipologie documentali differenti per valutarne la generalizzabilità.

### 7.2.2 Tempi di risposta e variabilità dovuta alle risorse hardware

È importante notare che i tempi di risposta del sistema non sono definiti in modo preciso, poiché dipendono dalle risorse hardware della macchina su cui vengono eseguiti i vari modelli, sia per quanto riguarda il modello di OCR, sia per quanto riguarda gli LLM coinvolti. Infatti, i tempi di elaborazione possono variare in modo significativo a seconda della potenza di calcolo, della memoria disponibile e della velocità del processore della macchina utilizzata. Pertanto, i tempi necessari per ottenere i risultati possono differire tra diverse configurazioni hardware, con macchine più performanti che forniranno risultati più rapidi. Inoltre, il numero di utenti che utilizzano il sistema simultaneamente può influire sui tempi di risposta. In scenari di utilizzo con un elevato carico di utenti, i tempi di elaborazione potrebbero aumentare, poiché il sistema potrebbe dover gestire risorse condivise e soddisfare richieste simultanee. Data la natura dei modelli di IA utilizzati, questi fattori influiscono in modo diretto sull’efficienza e la velocità del sistema, ma non possono essere facilmente standardizzati senza tener conto delle specifiche caratteristiche del dispositivo e del carico di lavoro.

## 8 Conclusioni e sviluppi futuri

### 8.1 Considerazioni finali

Il sistema sviluppato ha raggiunto un buon livello di automazione nell’elaborazione dei documenti. L’integrazione di Paddle OCR con i modelli GenAI, tra cui Llama 3.1 e Llama 3.3, ha consentito di ottenere risultati soddisfacenti in termini di accuratezza e scalabilità, riducendo il carico di lavoro manuale. Le scelte tecnologiche, come l’utilizzo di Paddle OCR e del modello Llama 3.3 per gestire scenari con informazioni limitate, si sono rivelate appropriate per garantire una categorizzazione automatica efficace e dati estratti di buona qualità.

### 8.2 Possibili miglioramenti al sistema

Nonostante i progressi compiuti, esistono diverse aree in cui il sistema potrebbe essere ulteriormente migliorato:

- **Miglioramento della gestione degli errori:** Implementare algoritmi più sofisticati per gestire meglio i casi in cui l’OCR o il modello GenAI non producono risultati ottimali, con soluzioni di fallback più complete.
- **Ottimizzazione delle risorse:** L’efficienza del sistema potrebbe essere migliorata riducendo i tempi di elaborazione, attraverso ottimizzazioni software o l’uso di macchine più performanti.
- **Integrazione multilingue:** L’introduzione del supporto per più lingue permetterebbe di adattare il sistema a esigenze aziendali internazionali.

- **Feedback continuo:** Sviluppare un meccanismo di apprendimento continuo che sfrutti il feedback degli utenti potrebbe contribuire a migliorare la precisione del sistema nel tempo.

Questi miglioramenti potrebbero ulteriormente potenziare la precisione, l'efficienza e l'affidabilità del sistema, rendendolo sempre più utile nelle operazioni aziendali quotidiane, con la possibilità di affrontare una varietà di documenti e scenari in modo ancora più efficiente.