



ATC
anzanigroup

AI FOR
KNOWLEDGE

PROGETTO AI4KNOWLEDGE

Analisi semantica dei testi

● **ATG Artificial Intelligence**

- Obiettivo
- Pipeline
- Image Processing
- Text extraction & OCR
- Text validation
- Question answering

ATG Artificial Intelligence

Anzani Group da più di 25 anni si occupa di soluzioni software e BI.

Ci siamo formati facendo esperienza con multinazionali e grossi gruppi e negli ultimi 10 anni ci siamo presi cura anche di piccole e medie aziende, con le stesse esigenze ma budget diversi.

ATG AI è una nuova realtà che fa parte di Anzani Group, specializzata nell'innovativo campo dell'Intelligenza Artificiale, con già diversi progetti in attivo in diversi mercati: Business, Financial, Industria 4.0, HR/Recruiting e sicurezza.



+50 **Direct People
Worldwide Team**

+300 **Worldwide
Customers**

☐ ATG Artificial Intelligence

☒ **Obiettivo**

☐ Pipeline

☐ Image Processing

☐ Text extraction & OCR

☐ Text validation

☐ Question answering

Obiettivo

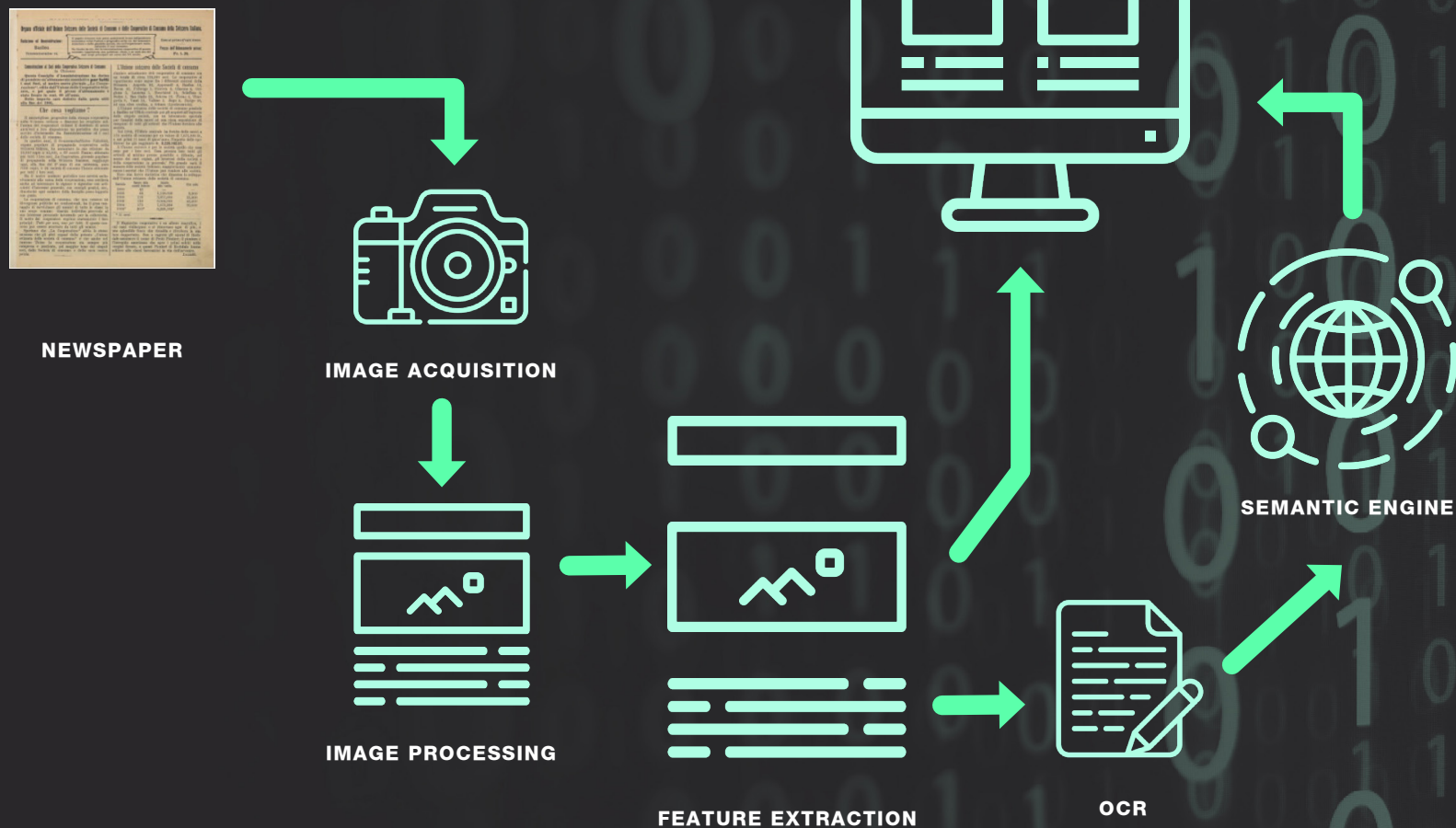
Lo scopo di questa soluzione è la creazione di uno strumento, basato su tecniche di intelligenza artificiale, che permetta di:

- Estrarre testo, tabelle, immagini e altri elementi costitutivi da pagine di giornale, pubblicazioni scientifiche, manuali, schede tecniche, etc.;
- Tradurre, tramite OCR, immagini contenenti testo in testo vero e proprio;
- Sottoporre i testi ottenuti ad analisi semantica, con il duplice scopo di ottenere l'indicizzazione dei contenuti e di ricostruire il testo in forma web friendly;
- Creare un sistema di risposta alle domande che risponda automaticamente alle domande poste attraverso il linguaggio naturale, estraendo i contenuti dalla knowledge base creata nel punto precedente.

Analisi semantica dei testi

- ☐ ATG Artificial Intelligence
- ☐ Obiettivo
- ☒ Pipeline
- ☐ Image Processing
- ☐ Text extraction & OCR
- ☐ Text validation
- ☐ Question answering

Pipeline

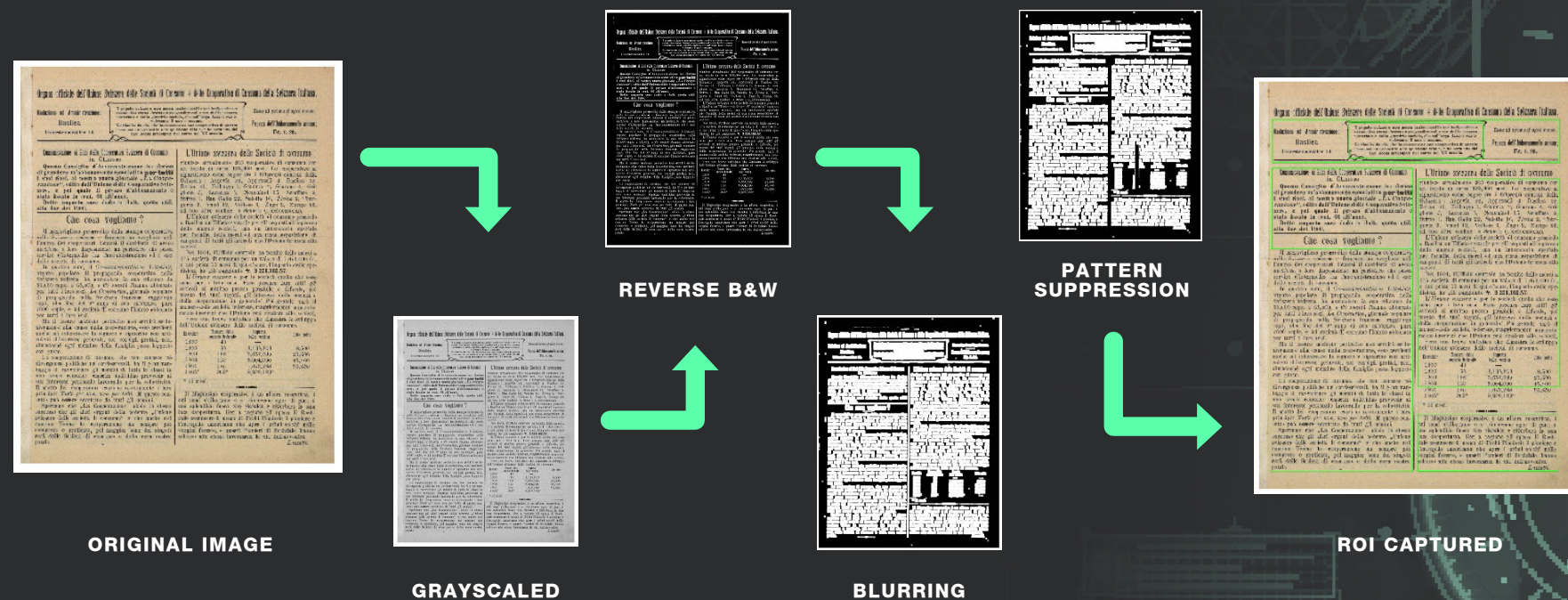


Analisi semantica dei testi

- ATG Artificial Intelligence
- Obiettivo
- Pipeline
- Image Processing
- Text extraction & OCR
- Text validation
- Question answering

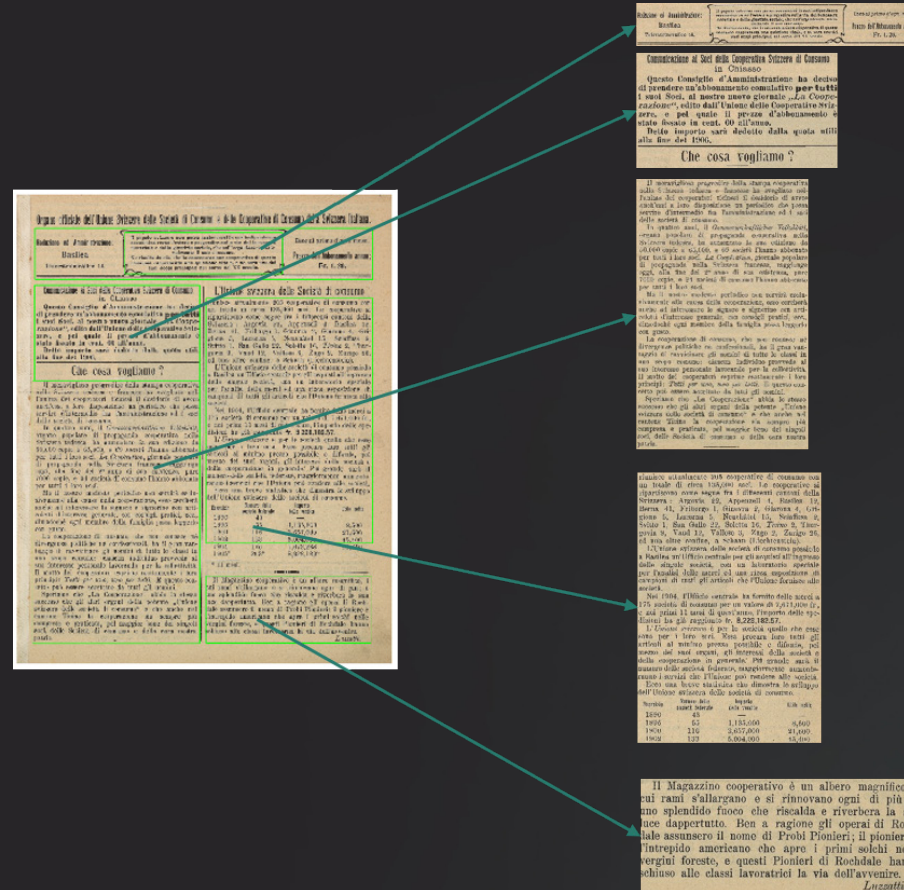
Image Processing

L'immagine subisce una serie di trasformazioni che servono a identificare le regioni di interesse.



L'immagine viene scomposta e le parti di testo passate a un OCR.

- ☐ ATG Artificial Intelligence
- ☐ Obiettivo
- ☐ Pipeline
- ☐ Image Processing
- ☒ **Text extraction & OCR**
- ☐ Text validation
- ☐ Question answering



Tesseract è un OCR open source. La versione utilizzata in AIVision è basata su reti neurali di tipo LSTM. È in grado di riconoscere 33 lingue.

- ☐ ATG Artificial Intelligence
- ☐ Obiettivo
- ☐ Pipeline
- ☐ Image Processing
- ☐ Text extraction & OCR
- ☒ **Text validation**
- ☐ Question answering

Text validation

Il Magazzino cooperativo è un albero magnifico,
i rami s'allargano e si rinnovano ogni di più;
lo splendido fuoco che riscalda e riverbera la sua
luce dappertutto. Ben a ragione gli operai di Rochdale
ne assunsero il nome di Probi Pionieri; il pioniere
intrepido americano che apre i primi solchi nelle
vergini foreste, e questi Pionieri di Rochdale hanno
schiuso alle classi lavoratrici la via dell'avvenire.
Luzzatti.



Il Magazzino cooperativo é un albero magnifico, i **cul** rami
s'allargano e si rinnovano ogni di **pil**; 6 uno splendido
fuoco che riscalda e riverbera la sua luce dappertutto.
Ben a ragione gli operai di Rochdale assunsero il nome
di Probi Pionieri; il pioniere é intrepido americano che
apre i primi solchi nelle vergini foreste, e questi Pionieri
di Rochdale hanno schiuso alle **elassi** lavoratrici la via
dell'avvenire.

Luzzatti

Il testo viene ricomposto su un'unica riga, senza ritorno
a capo, i caratteri non alfanumerici e la punteggiatura
vengono rimossi le stringhe di meno di due caratteri
vengono rimosse le parole chiave vengono rimosse,
riducendo il testo a un elenco di parole chiave ogni parola
chiave è convalidata su un dizionario di circa 1 milione
di parole. Le parole non valide vengono sostituite dalla
parola del dizionario più vicina in base a criteri misurabili.



ELENCO FINALE DI PAROLE CHIAVE

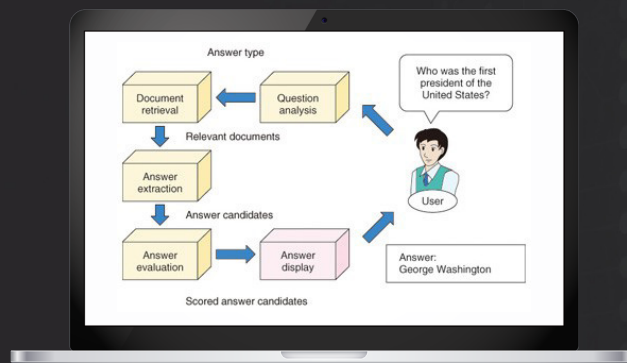


DOMINIO SEMANTICO DEL FRAMMENTO DI TESTO

- ATG Artificial Intelligence
- Obiettivo
- Pipeline
- Image Processing
- Text extraction & OCR
- Text validation
- Question answering

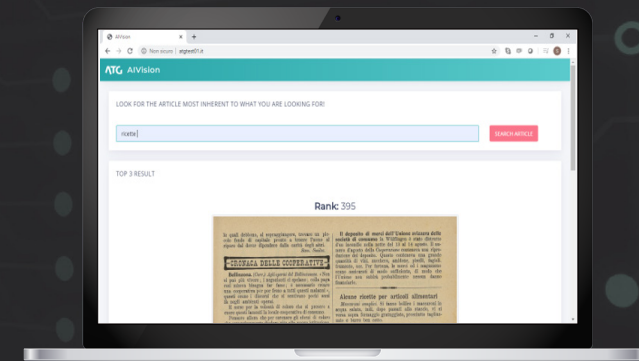
Question answering

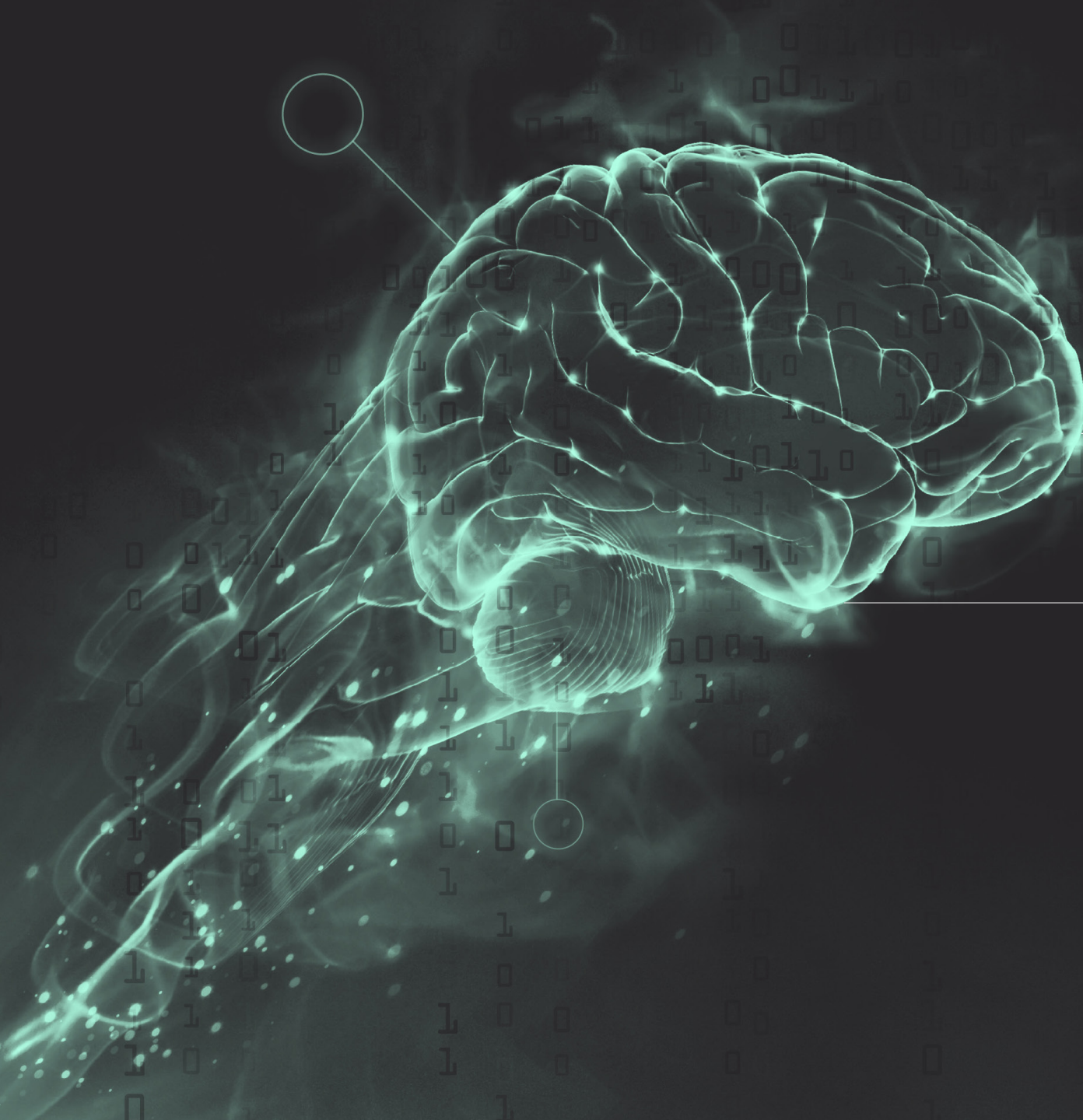
Rispondere alle domande non è più cercare una stringa in un testo, ma un concetto in un corpo di conoscenza (ontologia), tenendo conto del contesto.



Le domande vengono inoltrate al database, che utilizza la ricerca a testo completo per cercare le risposte ordinate in base alle classifiche.

L'insieme di tutti i contesti semantici raccolti dai vari frammenti di testo è immagazzinato in un database con il livello di accuratezza della pagina, e costituisce l'ontologia su cui si basano le risposte fornite all'utente.





ATG
anzani group

AI FOR
KNOWLEDGE

Tel. +39 031 611.860

Email info@atgartificialintelligence.com

SEDE PRINCIPALE

Corso XXV Aprile, 145
22036 Erba (CO)

SEDE DI SONDRIO

Via Stelvio, 24
23020 Poggiridenti Piano (SO)

SEDE DI STABIO

Via Gaggiolo, 27
6855 Stabio, Svizzera

www.atgartificialintelligence.com