

SERVIZIO · 03 / 05

# Guida di personalizzazione

Quando un commerciale interrompe un tecnico per chiedergli un prezzo che e' scritto a pagina 23 di un capitolato, il costo non e' i 5 minuti della risposta. E' la concentrazione persa dal tecnico che doveva finire un disegno, e la fiducia che chi cerca un dato lo trova solo chiedendo a qualcuno. DocBot smonta quel pattern. Cinque scelte da fare al kickoff su quale dominio coprire, come, e da quale chat. Sopra il template: 14-22 ore.

**5**

PUNTI DI CUSTOMIZZAZIONE

standard, su template gia' montato

**14-22 h**

SETUP TOTALE

vs 50-80h da zero

## 01 · Sorgenti documenti

I documenti restano dove sono oggi. DocBot legge da li, non vi chiede di centralizzare. Standard supportati:

- + **Cartelle di rete:** la share aziendale dove vivono i capitolati, i listini, i manuali. Lettura notturna con riconoscimento automatico dei file modificati e aggiornamento incrementale dell'indice.
- + **SharePoint o Google Drive:** librerie documentali cloud, accesso autenticato con permessi aziendali. Visibilita' mappata al gruppo di utenti del dominio operativo.
- + **Notifica da gestionale documentale interno:** se avete un portale SharePoint o un sistema documentale dedicato (es. Alfresco, M-Files), una notifica al rilascio segnala a DocBot che il documento e' nuovo o aggiornato. Indicizzazione in tempo quasi reale.
- + **Caricamento manuale:** documenti spot inviati a un indirizzo dedicato per il vostro cliente, indicizzati entro 24 ore lavorative. Utile durante onboarding iniziale e per aggiunte spot.

I permessi del documento sorgente NON vengono ereditati per default: tutti i documenti indicizzati sono visibili a tutti gli utenti del canale chat. Se serve segmentazione per ruolo (es. listini commerciali visibili solo al sales), gestiamo via canali chat distinti, non con filtri sulla singola query.

## 02 • Come dividiamo i documenti

Per rispondere bene serve trovare la porzione di documento giusta. Una porzione troppo piccola perde contesto, troppo grande dilata la ricerca e fa scendere la precisione delle citazioni. Strategia di partenza:

---

PDF TESTUALE	Divisione per sezione (titolo + 2-4 paragrafi). Ogni citazione punta a documento + sezione.
PDF SCANSIONATO	Lettura OCR prima della divisione, poi stesso schema del PDF testuale. La qualità di lettura viene misurata per ogni porzione: sotto una soglia minima la porzione viene segnalata.
DOCX	Divisione per intestazioni native del file. Tabelle preservate nella struttura.
MARKDOWN	Divisione per intestazioni. Caso ideale per qualità delle risposte (struttura esplicita nel file sorgente).
LISTINI EXCEL	Non supportati direttamente. Da convertire in PDF o CSV. La struttura tabellare profonda ha bisogno di un trattamento dedicato che riserviamo a un secondo step.

---

Per documenti molto lunghi (capitolati 80+ pagine) usiamo una divisione gerarchica: prima un sommario del documento intero, poi le porzioni granulari. La domanda iniziale trova il sommario, da lì scende sui dettagli rilevanti. Riduce le risposte sbagliate del 30-40%.

## 03 • Vincoli del modello e tono

Le regole con cui il modello risponde sono il punto più delicato. Quattro vincoli del sistema, non disattivabili:

---

### 01 **Citazione obbligatoria**

Ogni affermazione fattuale deve avere una citazione esplicita (file + sezione). Il modello si rifiuta di rispondere se non può citare. Senza questo vincolo le allucinazioni diventano routine.

---

### 02 **"Non lo so" autorizzato**

Se l'informazione non è nelle porzioni di documento recuperate, la risposta è un "non è nei documenti che ho indicizzato" esplicito, con suggerimento di chi contattare in azienda. Niente tentativi di indovinare, niente parafrasi creative.

---

### 03 **Lingua italiana**

Risposte in italiano per default. Se il documento sorgente e' in inglese (es. manuale macchina del fornitore tedesco), la risposta resta in italiano con citazione del passaggio originale in lingua. Inglese disponibile per organizzazioni multilingua.

---

### 04 **No consigli legali, fiscali, medici**

Se la query tocca dominio regolamentato (interpretazione contrattuale, calcolo IVA, sicurezza prodotto), il modello rinvia esplicitamente a consulenza umana. Vincolo necessario per limitare responsabilita' su risposte critiche.

---

**Tono:** diretto, concreto, no fronzoli. Vi chiediamo al kickoff 5-10 esempi di Q&A "fatti bene" e "fatti male" per calibrare voice. Tipica taratura: 3-5 ore in settimana 2.

## 04 • Canale chat

Tre canali standard, scegliibili anche in combinazione:

- + **Chat web integrata:** si inserisce nella vostra intranet o portale interno. Login automatico con identita' aziendale (Microsoft 365 o Google Workspace). Audit trail per utente.
- + **Bot Telegram:** bot dedicato al vostro cliente, aggiunto al gruppo aziendale o usato in privato. Per organizzazioni che gia' usano Telegram per comunicazione interna.
- + **App Microsoft Teams:** app dedicata, distribuita ai dipendenti dal vostro IT. Login automatico con identita' aziendale. Per organizzazioni Microsoft-centric.

I canali coesistono in parallelo: gli stessi documenti rispondono su chat web e Telegram con le stesse citazioni. Storico unificato.

## 05 • Aggiornamento del corpus

I documenti cambiano. La gestione del cambio e' parte del servizio:

---

#### AGGIORNAMENTO AUTOMATICO

Lettura notturna su cartelle di rete e SharePoint: file modificati o aggiunti vengono ri-indicizzati in modo incrementale. Tempo tipico 5-30 minuti per batch.

---

#### AGGIORNAMENTO IMMEDIATO

Notifica al rilascio dal vostro gestionale documentale, oppure su richiesta scritta (entro 4 ore lavorative). Per documenti critici (es. nuovo listino commerciale rilasciato a meta' settimana).

---

---

## RICOSTRUZIONE COMPLETA

Indice azzerato e ricostruito sull'intero corpus. Richiesto 2-4 volte all'anno quando passiamo a modelli aggiornati o cambiamo strategia di divisione. Incluso fino a 4 volte all'anno.

---

## STORICO

Quando un documento viene aggiornato, le porzioni vecchie sono marcate come obsolete (non cancellate): le risposte storiche restano riproducibili per audit. Rimozione definitiva fissata a 180 giorni.

---

### COSA NON PERSONALIZZIAMO · I DUE VINCOLI CHE FANNO FIDARE IL PERSONALE

"Cita la fonte" e "se non sai, dillo" sono hardcoded. Non si tolgono. Non sono dettagli implementativi: sono il contratto fra il sistema e i vostri colleghi. Il giorno in cui si toglie il primo vincolo, DocBot inventa una risposta plausibile a una query critica e qualcuno la prende per buona; tre settimane dopo nessuno crede piu' a quello che dice il sistema. Il giorno in cui si toglie il secondo, DocBot produce best guess su query fuori dominio e si sostituisce silenziosamente al collega che invece sapeva. Le due regole sono lente da apprezzare e veloci da rompere, ed e' per questo che sono regole di sistema, non di configurazione.