

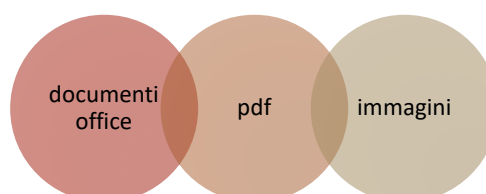
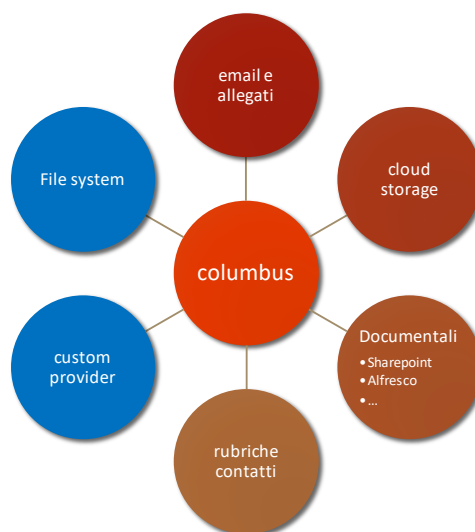
Overview

Sistema di indicizzazione e consultazione di contenuti personali e aziendali.

- Consente di ricercare, classificare e consultare indipendentemente dal formato di origine e dalla provenienza del contenuto.
- È in grado di estrarre conoscenza strutturata in modo automatico.
- Non richiede alcuna operazione di classificazione o annotazione manuale.
- Consente di ricercare con il linguaggio naturale.

Caratteristiche funzionali

- I contenuti possono provenire da contenitori generici (caselle di posta, servizi di archiviazione come Dropbox, GDrive o OneDrive) o (tramite appositi adattatori) da contenitori aziendali come i portali documentali;
- L'autenticazione al sistema è basata procedure standard (OAuth 2.0) potendo utilizzare tutti i provider compatibili (Google, Twitter, Facebook, Microsoft ecc.);
- È possibile ricercare contestualmente combinando la modalità full-text con criteri più strutturati;
- L'estrazione dei dati strutturati sui quali effettuare le ricerche è sostanzialmente a carico del sistema e non dell'utente;
- I risultati della ricerca sono raffinabili attraverso un sistema di faceting che permette di filtrare velocemente i contenuti trovati;
- I contenuti aggregati (come le mail che contengono attachments o archivi compressi) vengono analizzati interamente ed indicizzati singolarmente mantenendo il legame originariamente;
- I documenti scansionati vengono trattati con tecniche di OCR ed indicizzati;
- È possibile abilitare procedure di estrazione automatica dei dati strutturati di dominio attraverso servizi di IA;
- È possibile interrogare il patrimonio utilizzando query in linguaggio naturale (al momento solo in italiano);
- Tutti i contenuti, a prescindere dal formato originale, sono consultabili in modalità paginata con accesso diretto alla singola pagina e tempi invarianti rispetto alla dimensione del contenuto originale: ad esempio è possibile accedere alla singola pagina contenente l'informazione cercata dove è stata trovata direttamente senza doverlo scaricare interamente;
- La soluzione consente di condividere porzioni del patrimonio informativo indicizzato con altri utenti;
- I tempi di risposta del sistema sono calibrati per fornire risultati, sempre, entro i 3-5 secondi;



- Sono supportati tutti i formati standard (Pdf, HTML, Doc, Docx, Odt);
- È possibile indicizzare contenuti specifici e formati di documenti non contenuti nel set di base, sviluppando apposite estensioni software;
- È possibile integrare anche solo i servizi REST API all'interno di altre applicazioni esistenti;
- Il sistema è disponibile con UI in inglese e italiano. Le capacità di indicizzazione sono disponibili in circa 30 lingue.

Caratteristiche tecniche

La piattaforma è costituita da tre macro-blocchi:

- 1) **Servizi di indicizzazione e archiviazione:** provvedono alla creazione e alla manutenzione delle informazioni indicizzate nonché alla generazione dei formati consultabili dei patrimoni documentali analizzati;
- 2) **Servizi di pubblicazione e fruizione:** costituiscono il layer dei servizi pubblicati in formato REST a servizio di tutti clients;
- 3) **Applicazioni client:** includono tutte le applicazioni utilizzate dagli utenti: i clients desktop (Windows) e mobile (Android, iOS, Windows 10), nonché quelle web di amministrazione e gestione del proprio account.

Architettura

I blocchi 1) e 2) sono progettati per un'architettura cloud, e quindi per la scalabilità. Sono quindi nativamente multi-nodo e multi-thread e pensati per un'attivazione dinamica dei singoli nodi, in modo da gestire opportunamente SLA e carichi dinamici di lavoro.

Il blocco 3) è pensato per trarre beneficio dalle singole piattaforme su cui le applicazioni sono sviluppate, che sono quindi tutte native.

L'utilizzo di tecnologie standard (sia nell'erogazione che nei servizi) consente di utilizzare anche solo il livello dei servizi, potendolo integrare all'interno di applicazioni esistenti che sfruttarne le potenzialità per cercare e recuperare le informazioni provvedendo poi autonomamente alla presentazione delle stesse.

Tecnologie di riferimento

Le tecnologie utilizzate sono le seguenti:

- Storage ed indicizzazione
 - o ElasticSearch/Kibana/Logstash
 - o MongoDB (sostituibile con altri storage)
 - o Azure Table Storage
 - o CosmosDB
- Pipeline di indicizzazione e media-processing
 - o Correntemente .NET 4.7.1, in futuro .NET Core, eseguibile nativamente su piattaforme Linux)
- REST API
 - o .NET WebApi 5, in futuro .NET Core, eseguibile nativamente su piattaforme Linux)
- Applicazioni client
 - o Windows: WPF, .NET 4.7.1
 - o Windows 10 / Android / iOS: Xamarin