



How we used and extended Liferay Headless API for a cognitive search solution



Daniele Calderini
Data Scientist



Sommario



What is expected from research today



Cognitive Search



Openk9



Integration with Lifeary



Demo time

What is expected from research today

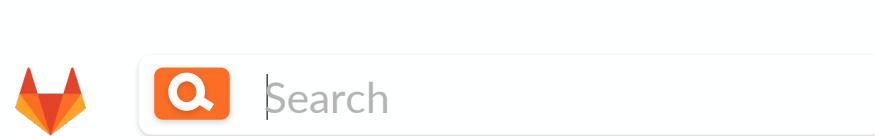
Information distributed

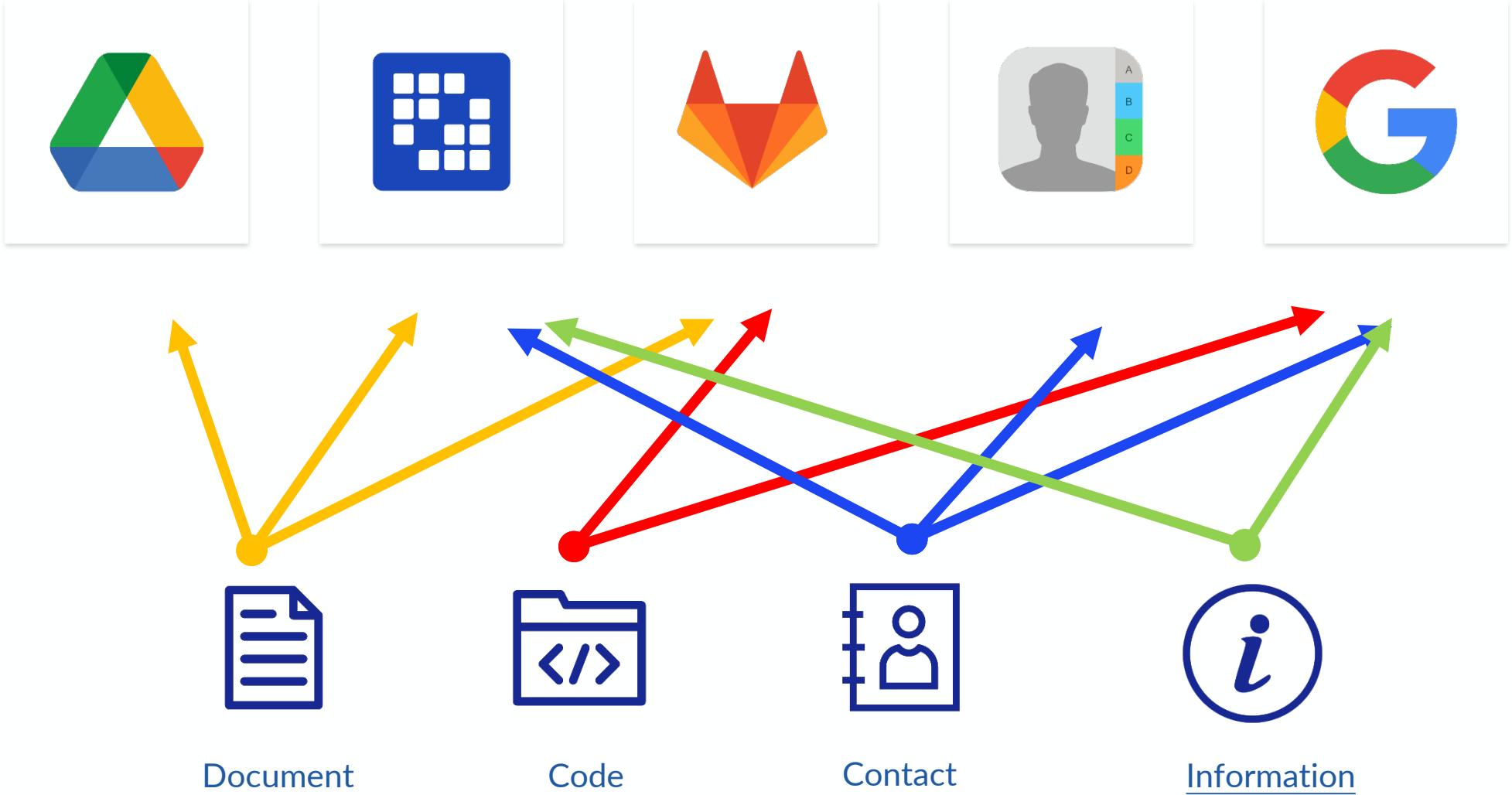


Content from multiple sources

Data comes from multiple sources

- Mail server
- Web portal
- Crm
- File System
- Etc...





Structured and unstructured data



= Unstructured data



= Structured data

Technical Specifications Vacancy Management(e1)

Written by:	Mario Rossi Developer	_____	31/03/2020
Verified by:	Francesco Bianchi Developer	_____	date
Approved by:	Paolo Verdi Developer	_____	date
Approved by:	Daniele Longo Project Manager	_____	date

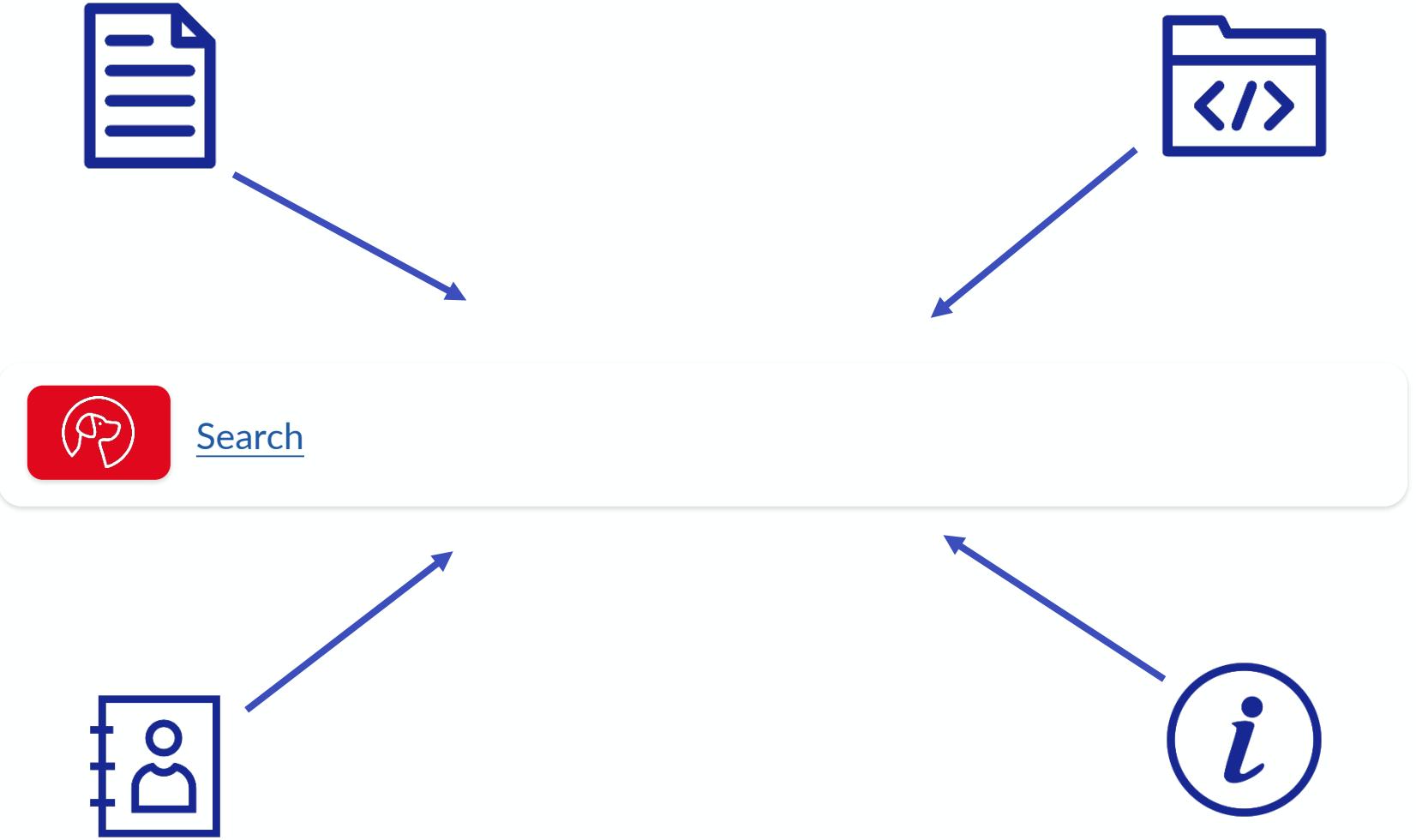
Project Code	: AA43526
Document Code	: dst-ssr20c31.docx
Version	: 00.01 del 31/03/2020 (Rev. Mod. 19D30 – M1-GP01)
Archived	: 010987-XXX Evolutive Calendario Turni /02 Realizzazione/03 Specifiche Tecniche
Distribution	: la lista di distribuzione è reperibile nel capitolo 2 del presente documento.

Firma per Ricevuta
ACME

(data)

Firma per Accettazione
ACME

(data)



Cognitive Search



Advanced search system that uses ML features to enrich the contents and support the user in search

ML Features for enrichment



Named Entity
Recognition



Text
Classification



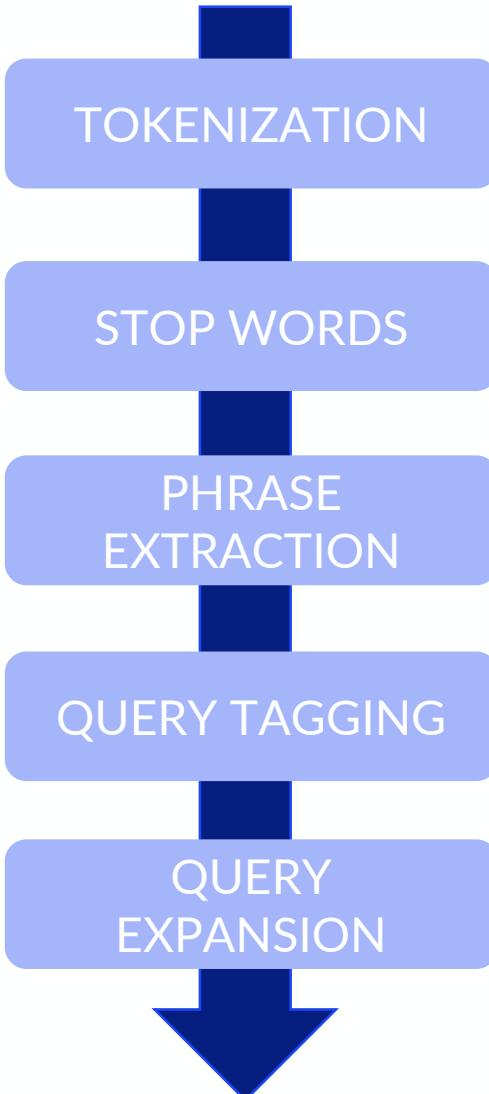
Knowledge
Extraction



Topic
Modelling

The screenshot illustrates a machine learning application for document enrichment. On the left, a sidebar lists entities such as 'smc treviso', 'specifiche funzionali', 'redatto da: enrico bortolan', 'stefano pesqualini', 'consumate applicative', 'data verificata da: paolo valeri', 'data approvata da: domenico costa', 'responsabile tecnico prodotto', 'supervisore tecnico', 'approvato da: paolo valeri', 'direttore tecnico', and 'firma per ricevuta'. On the right, a search interface shows results for 'type:document sisal', listing three documents: 'SAL_Progetto_SisalBetting.docx', 'SAL_Progetto_SisalBetting.docx', and 'SpecificheTecniche_SisalBetting.docx'. The interface includes tabs for 'DOCUMENTI', 'CONTATTI', 'CONVERSAZIONI', and 'APPLICAZIONI'.

Query Understanding



Pdf documents of Mike Brown developer of Acme Inc

Openk9

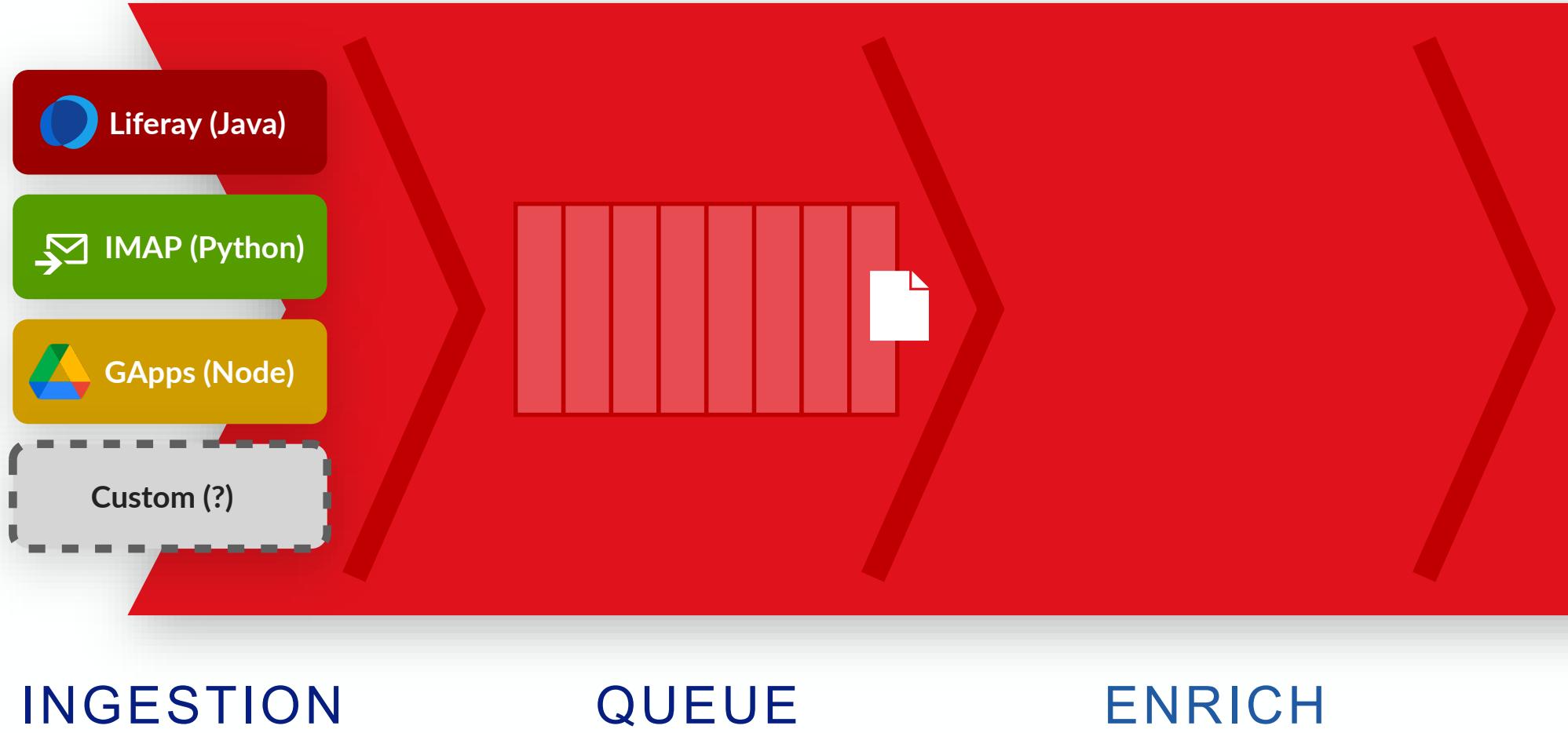
Openk9 Site: <https://www.openk9.io>

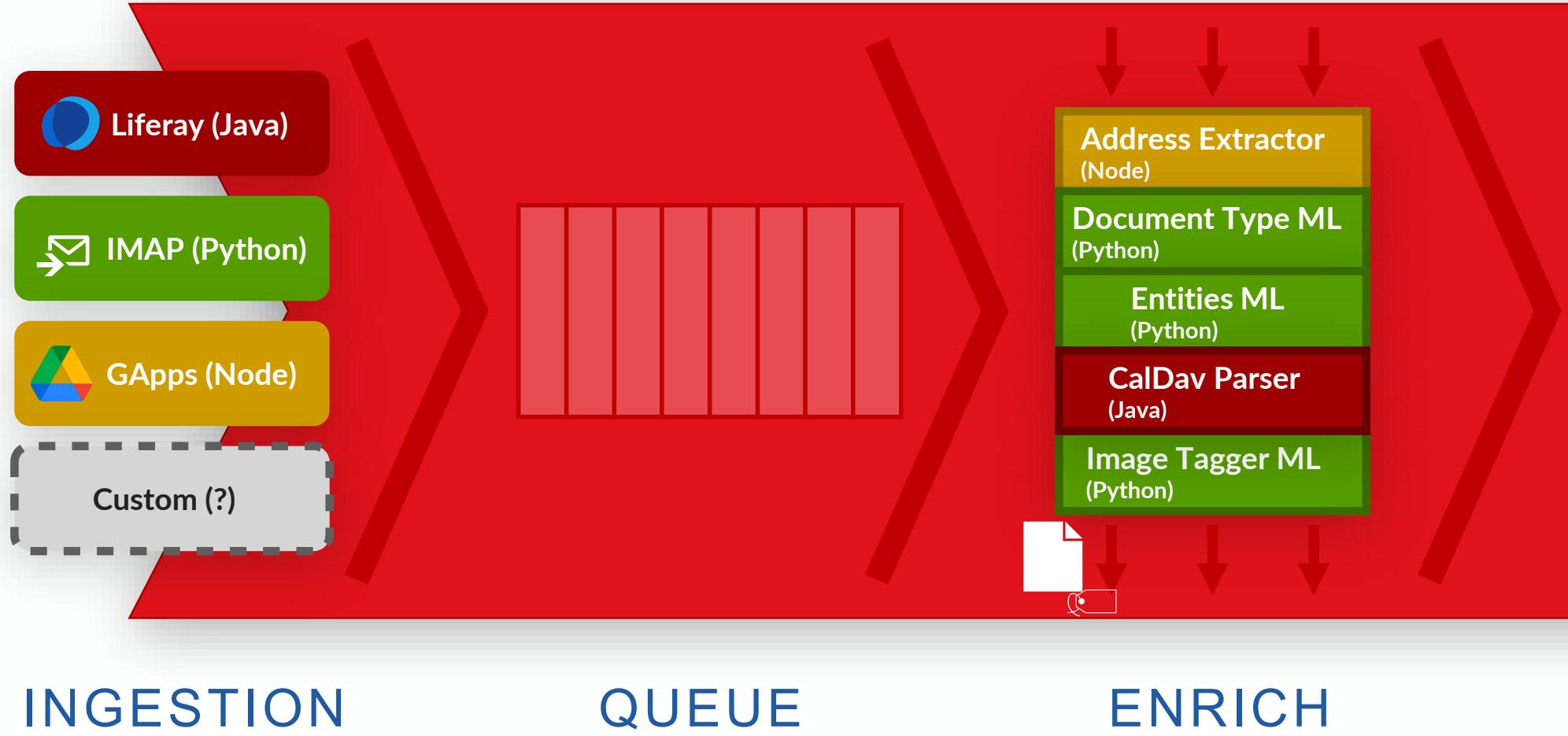
Openk9 Github Repository: <https://github.com/smclab/openk9>

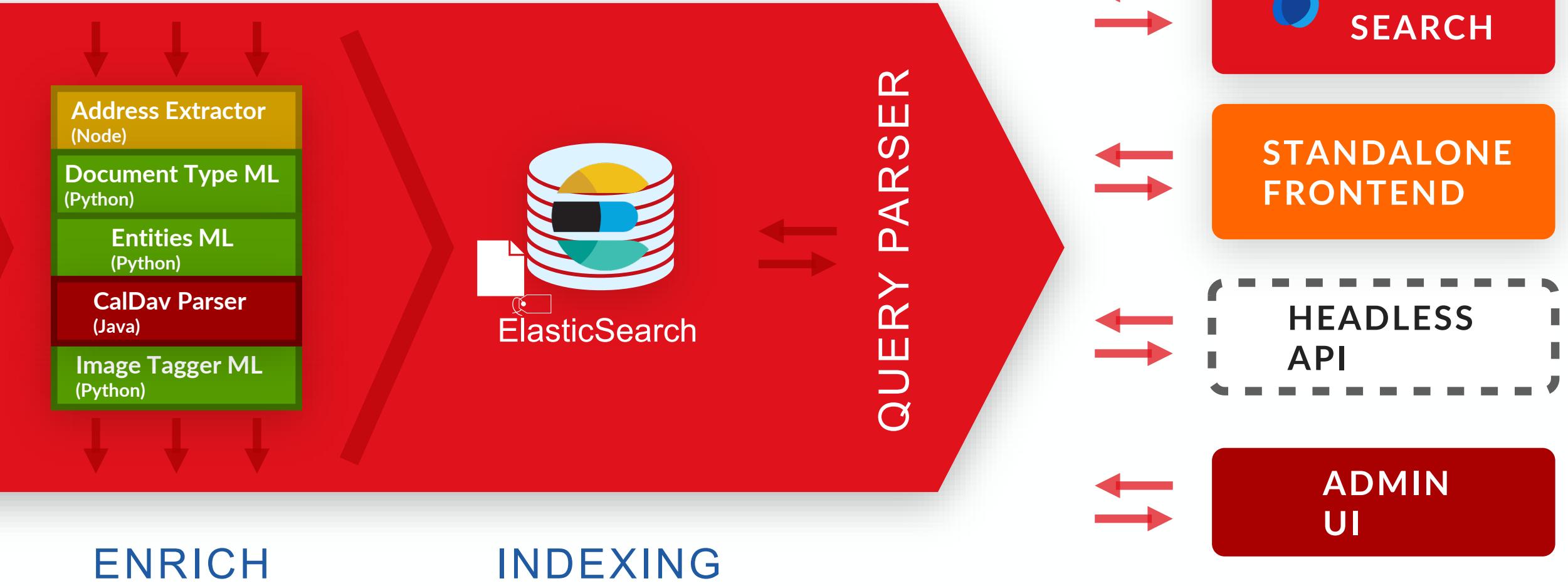
The Intelligent Open Source Engine







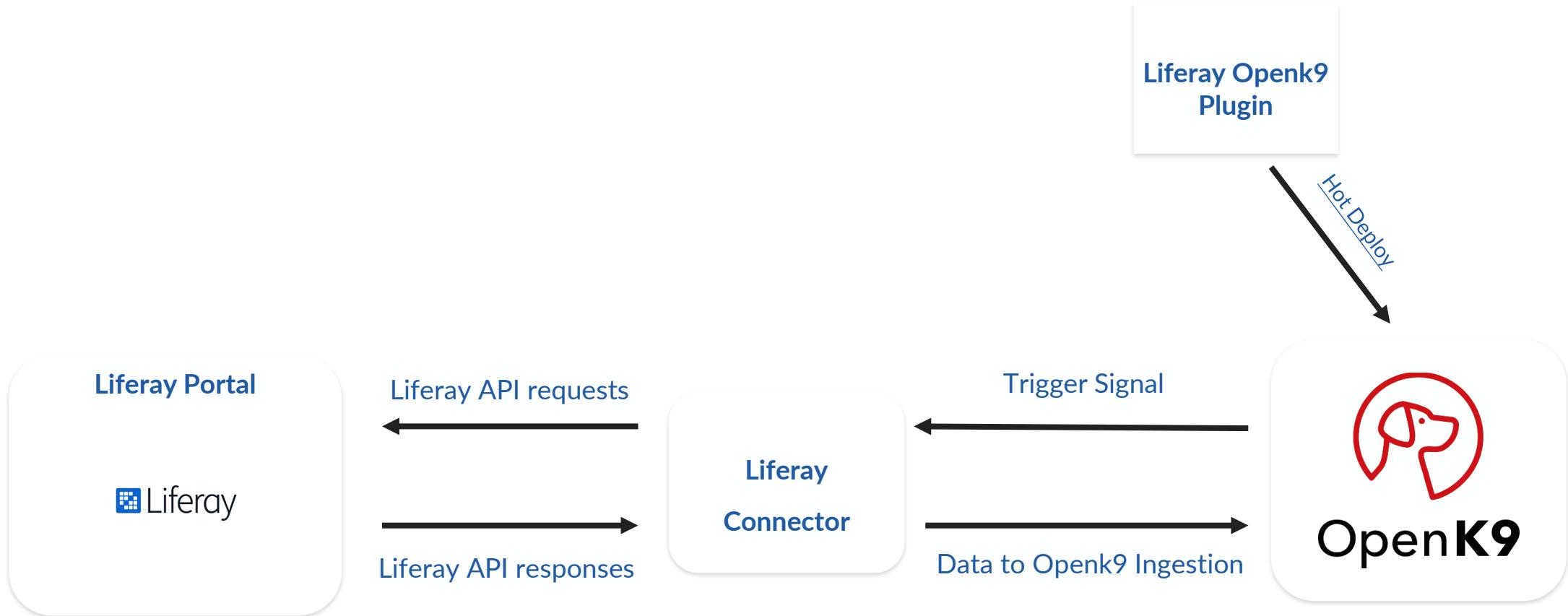




Integration with Liferay



OpenK9 - Liferay Integration





Liferay Connector

This is a simple Rest service that implements logic to:

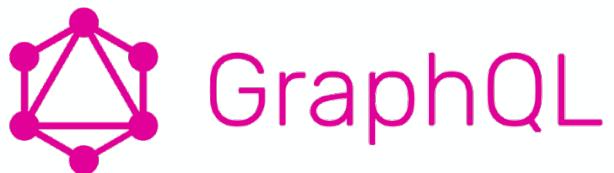
- Extract data from external data source
- Manipulates data to put in the exact form
- Sends these to Openk9 Ingestion (respecting the defined interface)

How to define connector for Openk9: <https://www.openk9.io/docs/create-external-data-parser>

Python implementation: <https://github.com/smclab/liferay-datasource/tree/main/liferay-parser>

How to integrate with Liferay

We start trying Liferay Headless APIs



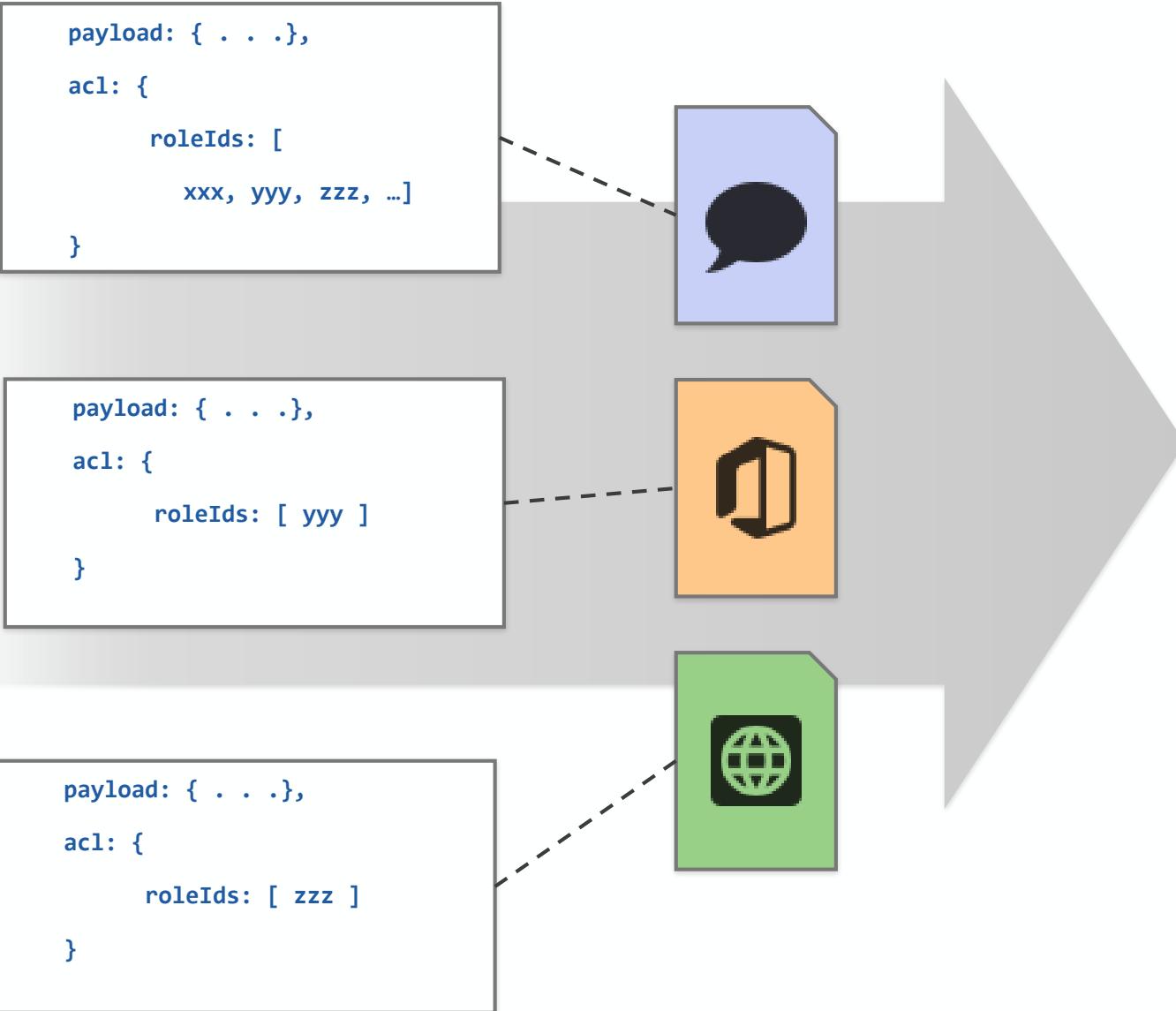
But we got some issues in get some important information for us:

- Roles
- Parsed Content
- ...

Liferay Api Extension: <https://github.com/smclab/liferay-api-extension>

Acl Handling (ingestion)

Liferay DXP

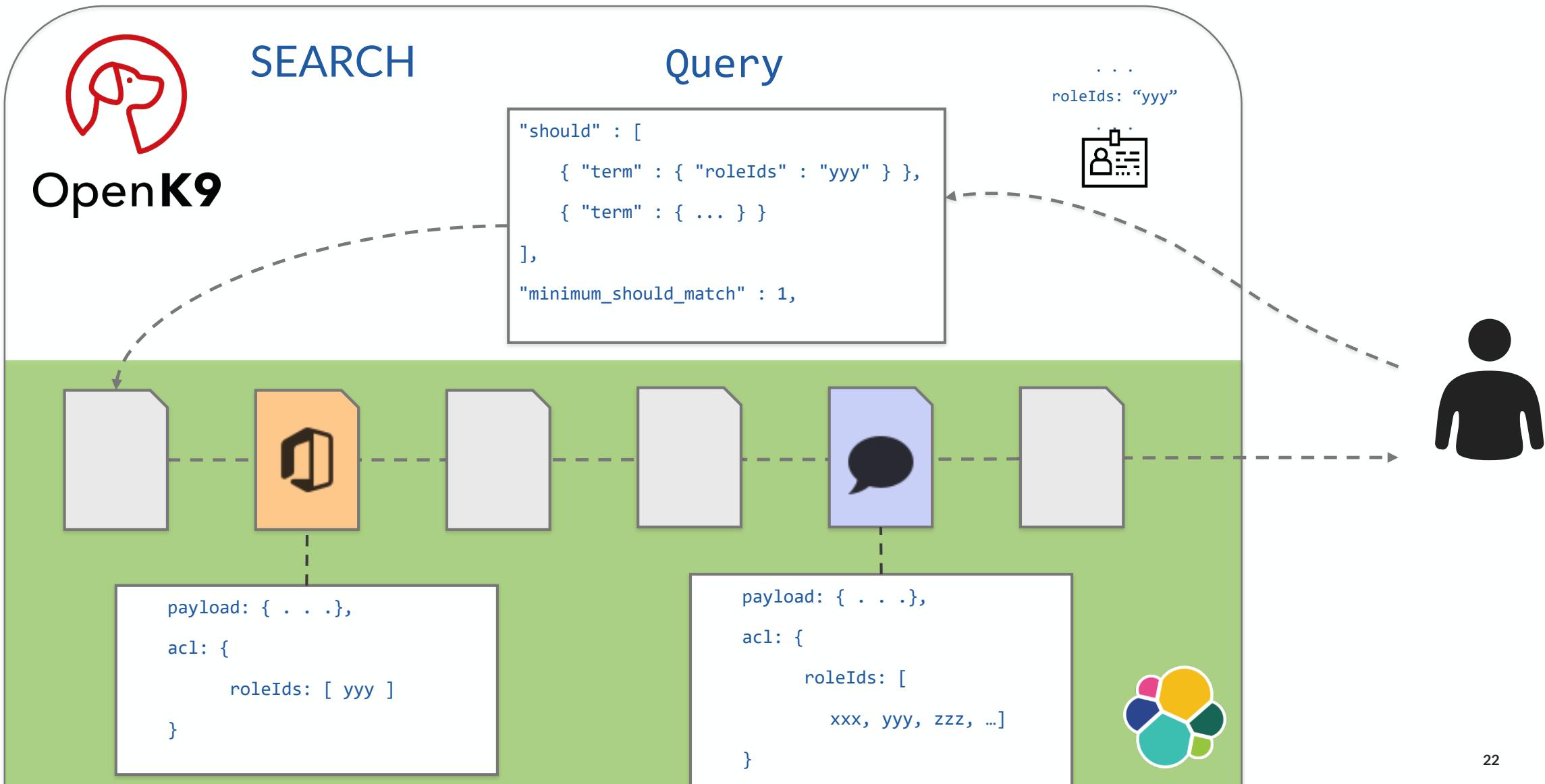


INGESTION



OpenK9

Acl Handling (ingestion)





Liferay OpenK9 Plugin

It consist of:

- Plugin Driver: defines plugin and connection to external connector
- Document Type Definition: defines data mapping
- Enrich Processors: defines enrich activities

Java Plugin Documentation: <https://www.openk9.io/docs/create-plugin-using-java>

Json Plugin Documentation: <https://www.openk9.io/docs/create-plugin-using-json>

Liferay Java Plugin: <https://github.com/smclab/liferay-datasource/tree/main/plugin>



Contatti

Daniele Caldarini

daniele.caldarini@smc.it

GitHub <https://github.com/danielecalda>

Linkedin linkedin.com/in/daniele-caldarini-101b69b4

Thank you 😊