Development of a Versatile, Full-Featured Search Functionality for Indico

Penelope Constanta

CHEP 2019

4 November 2019
The Collaboration

- Fermilab
  - Penelope Constanta

- BNL
  - Ofer Rind
  - Jose Caballero Bejar

- CERN
  - Pedro Ferreira
  - Adrian Mönnich
  - Pablo Panero
  - Carina Rafaela De Oliveira Antunes
  - Aristofanis Chionis Koufakos
Overview

• **Indico is:**
  – an open-source event management system, popular in HEP community
  – extensible through its plugin architecture (PayPal, video conferencing, search etc.)

• **Indico v2.x:**
  – has many improvements throughout the system
  – lacks search capabilities, outside the CERN eco-system that uses SharePoint

• **Search plugin necessity:**
  – CERN is moving away from SharePoint by the end of this year, to the new invenio based *CERN Search μservice*, necessitating the development of an indico interface
  – Fermi and BNL user communities requested a full functional search before deploying the new indico version

• **Fermi-BNL-CERN collaboration to build the search plugins:**
  – Utilizing the new *CERN Search μService* and make it available to the community
**Indico Search**

- **Indico v0.98 – v1.2:**
  - search utilizes invenio (v1.1) as its search engine sending its metadata in XML format
  - search results are formatted and displayed appropriately by indico
  - Framework can be used outside CERN’s environment

- **Indico v1.9 – v2.2:**
  - search sends search metadata to SharePoint by re-purposing the existing invenio plugin code
    - Metadata formatting does not take advantage of the new python packages (SQLAlchemy, marshmallow, etc.)
  - search results are displayed by SharePoint (indico simply displays the SharePoint page)
  - Framework cannot be used outside CERN’s environment

- **Collaboration plugin development for next version of indico v2.2.x:**
  - search utilizes invenio’s (v3) *CERN Search Api* component and Elasticsearch as its search engine, sending its metadata in JSON format and taking advantage of SQLAlchemy and marshmallow
  - search results are formatted and displayed appropriately by indico
  - framework is developed so that it can be used outside CERN’s environment
Indico Code Architecture

Indico Core System

Plugin Subsystem

Search Plugins

livesync
(core search engine population plugin)

livesync
Agent #1

livesync
Agent #2

Search
(core search result display plugin)

search
Agent #1

search
Agent #2

…

…

Missing Plugins to be implemented by the collaboration

Requires UI modifications
Indico Search System Architecture

Sends indico object metadata to https://cernsearch_api/

Agent that sends search strings, receives & displays search results (https://cernsearch_api/)

- Indico
  - livesync Agent #1
  - search Agent #1

CERN Search μService

Elasticsearch Server
Implementation Challenges

- Indico v2.x moved away from the ZOPE database to PostgreSQL and almost the entire indico code was re-written and restructured
  - Any familiarity with the previous versions’ code is not useful
  - Plugin development is seemingly easier but at the end one needs to understand all the internals of the new indico plugin system as well as the interface with the base plugins and the core indico code, along with the numerous new python packages

- CERN Search µservice is very new and documentation is targeted for CERN’s internal use
  - Deployment through docker-compose prove to be more challenging as the µservice is targeted for CERN’s internal use.

- FNAL and BNL developers worked for a fraction of their time on the indico project and were not familiar with the used python packages.
CERN’s documentation is excellent for installing/upgrading and setting up indico!

Installation
- Just follow CERN’s indico 2.x installation
- For our development purposes we installed the developer’s version:
  - https://docs.getindico.io/en/latest/installation/development/

Enable search plugins – Configuration
- All required steps are at:
Deployment of CERN Search μservice

- CERN provided the docker-compose.yml that creates:
  - The cern_search_rest_api as an invenio component
  - The PostgreSQL application
    • not required if connecting to an existing DB
  - The Elasticsearch (ES) application
    • not required if connecting to an existing ES installation
  - The ES kibana application
    • not required if connecting to an existing ES installation
  - The tika server to parse PDF, pptx, LaTex etc. files, needs to be added, if not connecting to an existing tika server.
  - It also initializes the invenio DB and uploads the ES mappings
Implementation Status

• livesync agent for CERN Search µservice
  – First version development almost completed
  – Not fully tested, awaiting the cern-search-api deployment

• Indico search User Interface
  – First version development completed, requires minor modifications
  – Search results UI:
    • provides filtering capabilities for Speakers and Affiliations
    • uses different tabs for events, contributions, attachments, notes
    • displayed page controls
  – Tested with mock data

• search agent for CERN Search µservice
  – Last stages of development
  – Not fully tested, awaiting the cern-search-api deployment
Indico livesync Agent plugin Configuration

LiveSync_Json is the agent for the CERN search µservice
Indico Search User Interface

Welcome to Indico. The Indico tool allows you to manage complex conferences, workshops and meetings. To start browsing, please select a category below.
Future Development

• All plugins developed by the collaboration will be integrated into indico and CERN will take ownership.

• Further development may include:
  – Improved resilience and recovery for the livesync agent
  – Extensions to search UI, if needed
  – Improved developer documentation and deployment for non-CERN environments

If you can find this talk on CERN’s indico site, using indico search, in 2020 then this collaboration was successful!