Elasticsearch as a new search engine for Invenio and more

Johnny Mariéthoz, RERO - Library Network of Western Switzerland,
Johnny.Mariethoz@rero.ch

Abstract

[Elasticsearch][1] (ES) is a recent powerful search engine based on [Lucene][2], as [SOLR][3], but with a set of unique and interesting features such as RESTful APIs.

Following an open strategy, [Invenio software][4] is increasingly integrating external libraries and offering the possibility of plugging in external search engines. We take advantage of this and propose the integration of ES as a search engine alternative.

We show how ES has been integrated in Invenio not only as a full-text search engine but also for faceting, filtering, highlighting search results, specific navigation tools and even log analysis. This has been implemented in [RERO][5]'s digital library [RERO DOC][6], which is demonstrated to illustrate the potential of this approach both in terms of new user interface features and advanced information extraction capabilities.

Audience

The presentation shows new navigation features and how they are implemented. Thus, it fits a large audience including both end-users and IT professionals with a technical background, namely developers and repository managers.

Background

Interest Group Presentations focused on Invenio.

Presentation content

This presentation covers:

- demonstrations of an Invenio instance using elasticsearch (ES)
  - a presentation of the new navigation features in the user interface
  - digitized press documents illustrating the use of advanced ES features
- a general presentation of ES
- basic python ES examples to illustrate its usability
- the integration of ES in Invenio
  - index design
  - data mapping
  - advanced ES features particularly useful for Invenio
- deployment tricks for a proper ES cluster
- other possible use cases for ES, such as statistics

This presentation doesn't cover:

- the source code analysis

Conclusion

Elasticsearch shows its capability to cover important Invenio needs and more, by offering new features such as hierarchical facets. Its ease of use, RESTful API and complete Python library makes it a good candidate to become the main search engine of Invenio. We can also imagine using it for many other purposes such as statistics, system monitoring, static page searching (e.g. help pages), etc.

References

1. [elasticsearch](#)
2. [Lucene](#)
3. [SOLR](#)
4. [Invenio](#)
5. [RERO](#)
6. [RERO DOC](#)