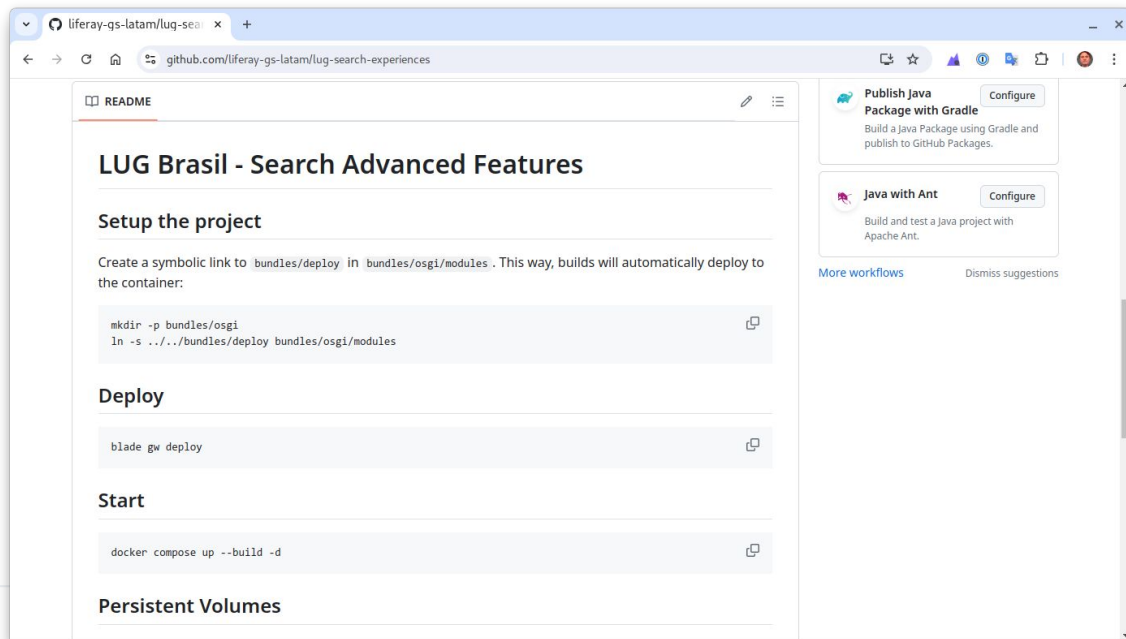


Domine a Busca:

Explorando recursos avançados

Workspace

github: /liferay-gs-latam/lug-search-experiences



Overview do Workspace de Trabalho

- Docker compose based
- Persistent Volumes
- Liferay configuration with environment variables
- Keep portal state (with database and document library dumps)
- Integração com Liferay workspace (blade)

Busca

Busca



Visão geral funcionalidades de busca (CE)

- Full-text search
- Indexing of all content types (blogs, documents, web content, etc.)
- Highly configurable search pages
- Automatic index synchronization for added, updated, and deleted content
- Search result filtering by roles and permissions
- Search bar suggestions
- Search result configuration (e.g., filter and sort)
- Faceted search
- Enable or disable advanced search syntax (e.g., AND/OR/NOT, wildcards, and more)
- Semantic search
- Federated Search

Visão geral funcionalidades de busca (DXP)

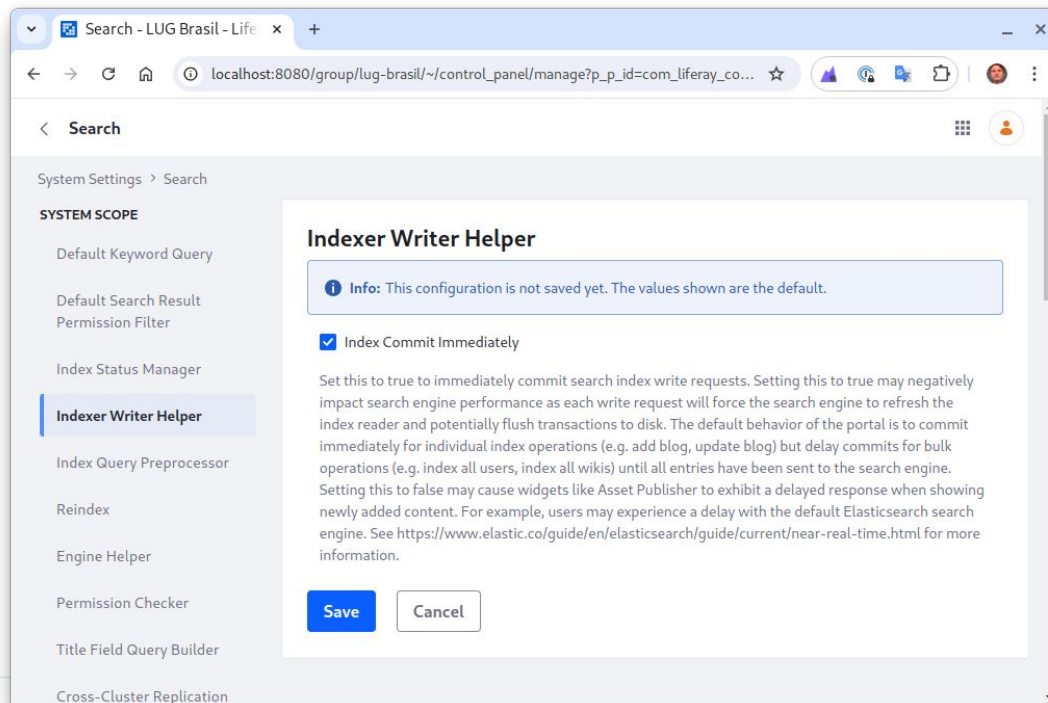
- Search Blueprints
- Synonyms
- Result Rankings

Tuning e Configurações



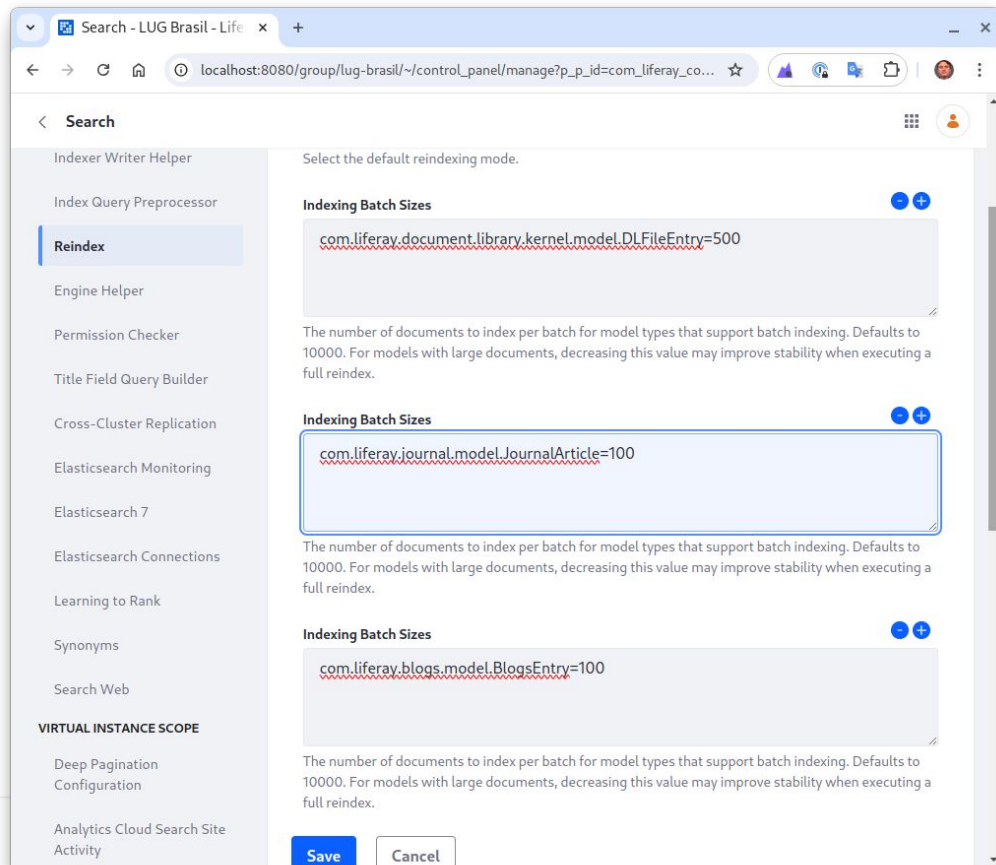
Tuning e Configurações

- Indexer Writer Helper



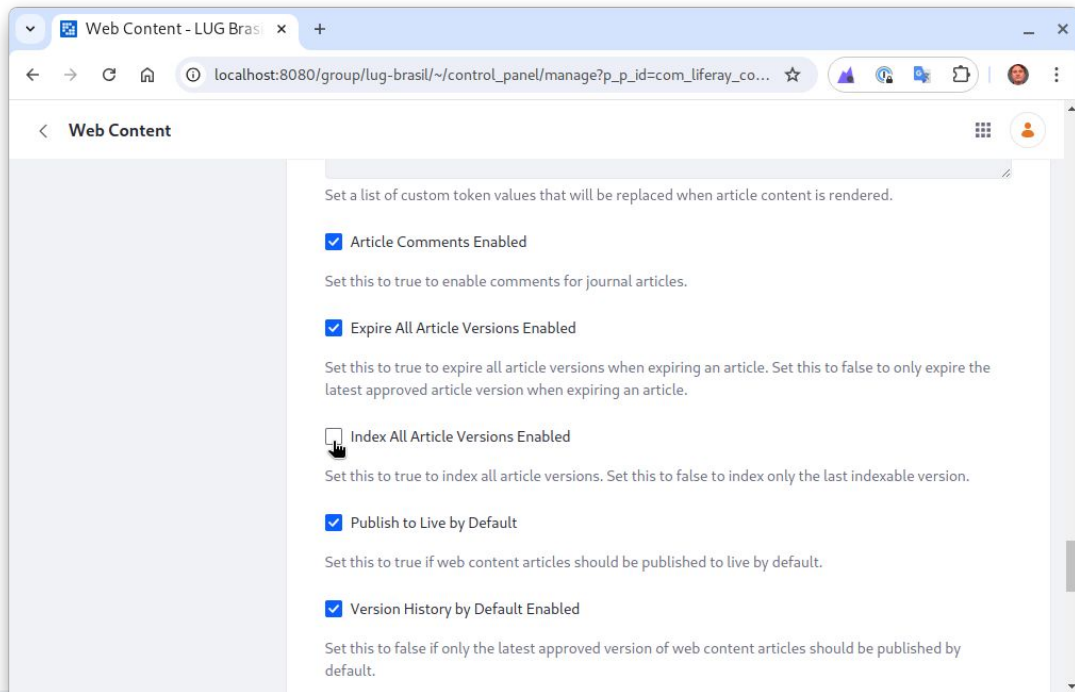
Tuning e Configurações

- Reindex Batch Size



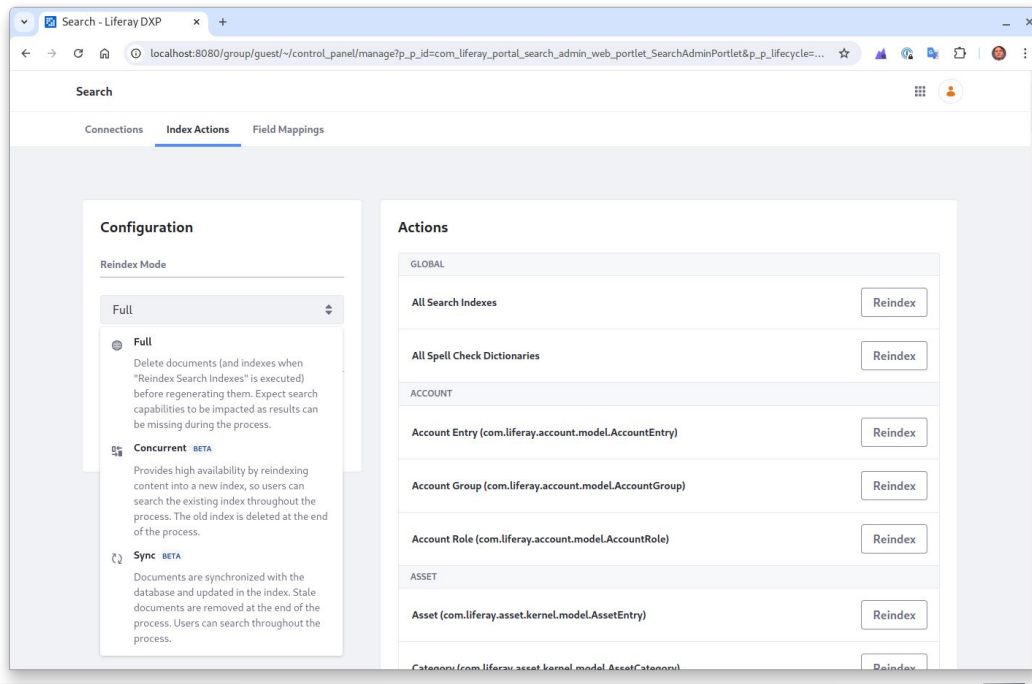
Tuning e Configurações

- Disable Index all articles versions



Tuning e Configurações

- Novas formas de reindexação



Explorando a página de busca



Explorando Blueprints e Result Rankings



Busca em um Elasticsearch externo



https://dumps.wikimedia.org/



<https://github.com/AlonEirew/wikipedia-to-elastic>

The screenshot shows the GitHub repository page for `AlonEirew/wikipedia-to-elastic`. The page includes a README section with a description of the project, a list of project features, and a sidebar with statistics.

README Apache-2.0 license

DOI 10.5281/zenodo.3239509

Wikipedia to ElasticSearch

This project generates an ElasticSearch, or file index from Wikipedia (xml dumps). The process will analyze, extract and store Wikipedia article text and several distinct Wikipedia attributes and relations (detailed below).

Project Features:

- Export Wikipedia in different languages {*English, French, Spanish, German, Chinese*}
- Export other Wikimedia resources: {*Wikipedia, Wikinews, Wikidata*}
- Support storing to either an Elastic index or file system (json files)
- Support the extraction of Wikipedia article text clean of markdown and html tags
- Integrated with Intel [NLP Architect](#)

Packages
No packages published

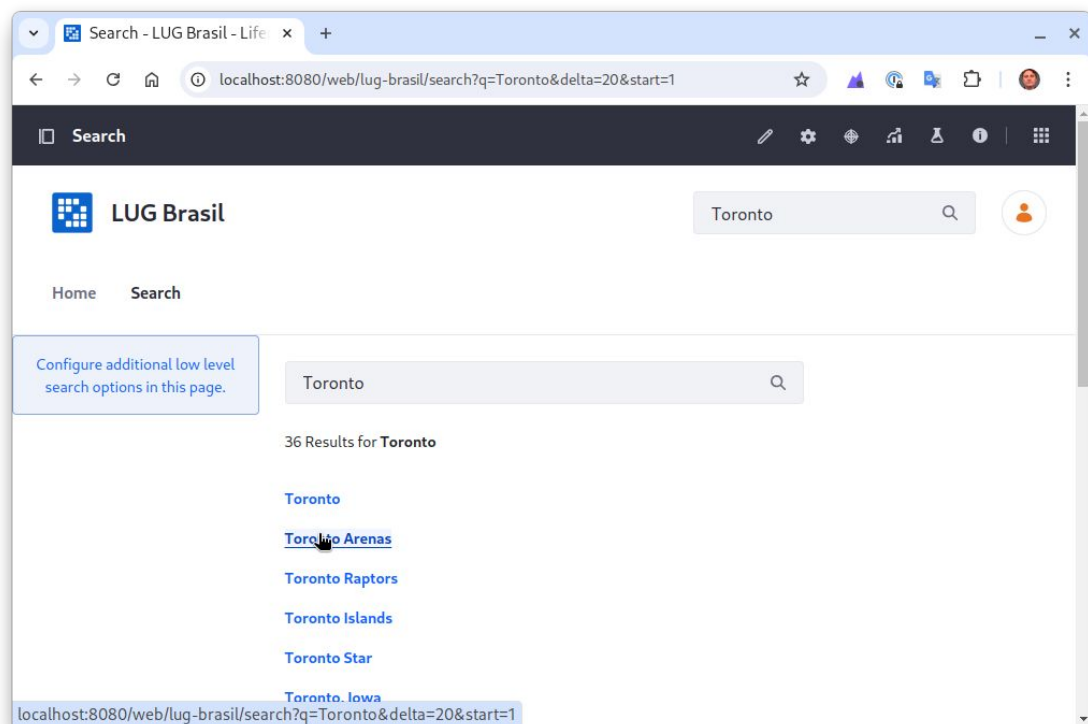
Contributors 2

- AlonEirew** Alon Eirew
- minhptx** Minh Pham

Languages

Language	Percentage
Java	98.2%
Shell	1.6%
Dockerfile	0.2%

LUG Brasil | Explorando recursos avançados da busca



Desafio: Alterar o Widget Template para construir o link do resultado para a página atual do Wikipedia.

Link para uma página do Wikipedia:

`http://en.wikipedia.org/?curid=<<ID-DA-PAGINA>>`

LUG Brasil | Explorando recursos avançados da busca

```
8      "hits": {
9        "hits": [
10         {
14           "_explanation": {
16             "details": [{
23               "value": 10.924561
24             },
25             {
26               "value": 10.924561
27             },
28             {
29               "value": 10.924561
30             }
31           ],
32           "_source": {
33             "title": "Toronto",
34             "parsedParagraphs": [
35               "Toronto is the most populous city in Canada and the capital cit",
36               "Limited commercial and passenger service to nearby destinations",
37               "Within a few hours' drive, Hamilton's John C. Munro Internation",
38               "The grid of major city streets was laid out by a concession roa",
39               "There are several municipal expressways and provincial highways"
40             ]
41           },
42           "_id": "64646",
43           "_score": 10.924561,
44           "_shard": "[enwiki_v3][0]",
45           "_node": "ir6W0WIXTqWxONFqQ-h1Hw"
46         },
47         {
48           "_ignored": ["parsedParagraphs.keyword"],
49           "_index": "enwiki_v3",
```

<#list entries as entry>

entry.getFieldDisplayContexts()

name, value

name, value

...

</#list>

LUG Brasil | Explorando recursos avançados da busca

<https://github.com/andrefabbro/wikipedia-to-elastic/commits/master/>

Commit

Add new attribute named wikipediaPageId to source

andrefabbro committed 4 hours ago

Showing 4 changed files with 20 additions and 5 deletions.

src/main/java/wiki

- data
 - WikipediaParsedPage.java
 - WikipediaParsedPageBuilder.java
- persistency
 - ElasticAPI.java
- utils/parsers
 - WikipediaSTAXParser.java

src/main/java/wiki/data/WikipediaParsedPageBuilder.java

```
@@ -13,6 +13,7 @@ public class WikipediaParsedPageBuilder {
13 13     private List<String> parsedParagraphs;
14 14     private String redirectTitle;
15 15     private WikipediaParsedPageRelations relations;
16 +    private String wikipediaPageId;
16 16
17 17     public WikipediaParsedPageBuilder setTitle(String title) {
17 18         this.title = title;
18 19
19 19
20 20
21 21
22 22
23 23
24 24
25 25
26 26
27 27
28 28
29 29
30 30
31 31
32 32
33 33
34 34
35 35
36 36
37 37
38 38
39 39
40 40
41 41
42 42
43 43     this.parsedParagraphs = parsedParagraphs;
44 44     return this;
45 45 }
46 46
47 +
48 +     public WikipediaParsedPageBuilder setWikipediaPageId(long id) {
49 +         this.wikipediaPageId = String.valueOf(id);
50 +         return this;
51 +     }
52 52
```

LUG Brasil | Explorando recursos avançados da busca

```
8      "hits": {
9        "hits": [
10         {
14           "_explanation": {
16             "details": [{
73               },
74               "value": 10.924561
75             }],
76             "value": 10.924561
77           }],
78           "value": 10.924561
79         },
80         "_source": {
81           "title": "Toronto",
82           "parsedParagraphs": [
83             "Toronto is the most populous city in Canada and the capital cit
84             "Limited commercial and passenger service to nearby destinations
85             "Within a few hours' drive, Hamilton's John C. Munro Internation
86             "The grid of major city streets was laid out by a concession roa
87             "There are several municipal expressways and provincial highways
88           ],
89           "wikipediaPageId": "64646"
90         },
91         "_id": "64646",
92         "_score": 10.924561,
93         "_shard": "[enwiki_v3][0]",
94         "_node": "ir6W0WIXTqWxONFqQ-h1Hw"
95       },
96     ],
97     "_ignored": ["parsedParagraphs.keyword"],
```

LUG Brasil | Explorando recursos avançados da busca

```
1  <#function getFieldValueByName entry fieldName>
2      <#if entry.getFieldDisplayContexts()?has_content>
3          <#list entry.getFieldDisplayContexts() as fieldDisplayContext>
4              <#if fieldDisplayContext.getName() == fieldName>
5                  <#return fieldDisplayContext.getValuesToString()?trim>
6              </#if>
8          </#list>
9      </#if>
10     <#return "">
11 </#function>
```

```
<#list entries as entry>
    <#assign wikipediaPageId = getFieldValueByName(entry, "wikipediaPageId")>
    <#if wikipediaPageId?has_content>
        <#assign title = getFieldValueByName(entry, "title")>
        <li class="align-items-center d-flex">
            <div class="c-mb-2 c-mt-2">
                <a class="link-primary single-link" href="http://en.wikipedia.org/?curid=${wikipediaPageId}" target="_blank">
                    ${title}
                </a>
            </div>
        </li>
    </#if>
</#list>
```

Busca em um outro índice no mesmo Elasticsearch



Busca em mais de um índice ao mesmo tempo

Busca Federada

