

Search Engine for Astrology

Project final presentation

ITCS414 Information Storage and Retrieval



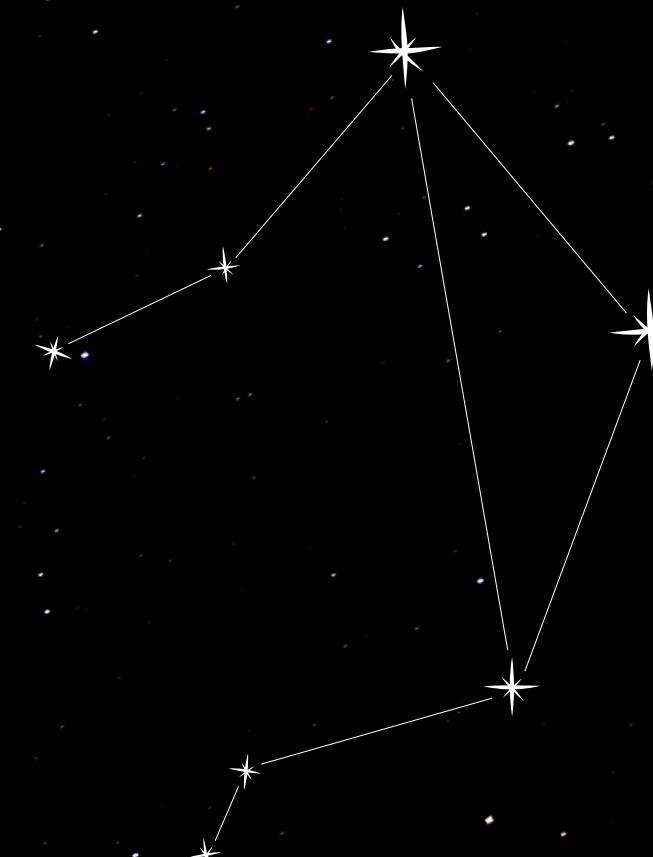
Meet The Team



6488179
Ponnápassorn
Iamborisut



6488181
Thadeeya
Duangkaew



6488210
Ravikarn
Jarungjittittawas



AstroSage 101

AstroSage 101 is an innovative astrology-focused search engine designed to provide a comprehensive and intuitive platform for astrology enthusiasts. Our motivation behind creating AstroSage 101 stems from the recognition of the increasing interest and curiosity about astrology and its potential to empower and enrich lives.



Contents

Technical difficulties

Challenges

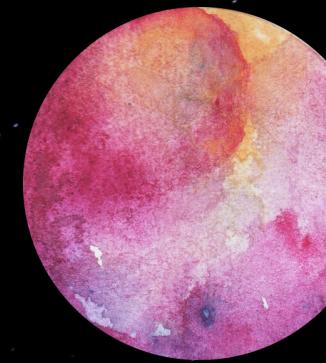
Lessons learned

Demonstration of
Astrology search system

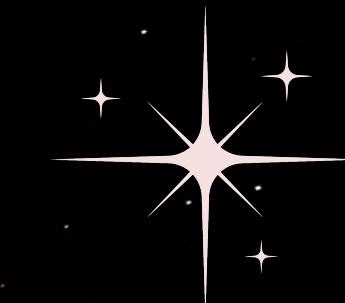


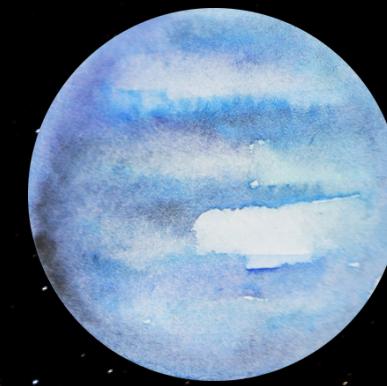


Technical Difficulties



kibana

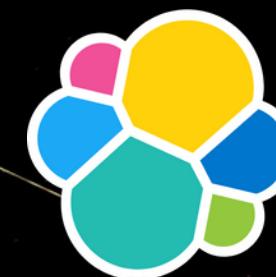




```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\mmay> cd D:\elasticsearch-8.11.1\bin
PS D:\elasticsearch-8.11.1\bin> .\elasticsearch.bat
Nov 25, 2023 11:47:54 AM sun.util.locale.provider.LocaleProviderAdapter <clinit>
WARNING: COMPAT locale provider will be removed in a future release
[2023-11-25T11:47:55,967][INFO ][o.a.l.i.v.PanamaVectorizationProvider] [LAPTOP-PT7SFDLG] Java vector incubator API enabled; uses preferredBitSize=512
[2023-11-25T11:47:56,751][INFO ][o.e.n.Node] [LAPTOP-PT7SFDLG] version[8.11.1], pid[14476], build[zip/6f9ff581fbcd658e6f69d6ce03050f060d1fd0c/2023-11-11T10:05:59.421414]
[OpenJDK 64-Bit Server VM/21.0.1/21.0.1+12-29]
[2023-11-25T11:47:56,753][INFO ][o.e.n.Node] [LAPTOP-PT7SFDLG] health status changed from [RED] to [GREEN] (reason="shards started [[.apm-source]]")
[2023-11-25T11:47:56,753][INFO ][o.e.n.Node] [LAPTOP-PT7SFDLG] -Des.networkaddress.cache.negative.ttl=10, -Djava.security.egd=file:/dev/./urandom, -headless=true, -Dfile.encoding=UTF-8, -Djna.nosys=true, -XX:+UseG1GC, -Djava.io.tmpdir=C:\Users\mmay\AppData\Local\Temp\elasticsearch, --add-modules=jdk.incubator.vector, -XX:+HeapDumpOnOutOfMemoryError, -XX:+ExitOnOutOfMemoryError, -XX:HeapDumpPath=data, -XX:ErrorFile=logs/hs_err_pid%p.log, -Xlog:gc*,gc+age=trace,safepoint:file=logs/gc.log:utctime,level,pid,tags:filecount=32,filesize=64m, -Xms12169m, -Xmx12169m, -XX:MaxDirectMemorySize=6383730688, -XX:InitiatingHeapOccupancyPercent=30, -XX:G1ReservePercent=25, -Des.distribution.type=zip, --module-path=D:\elasticsearch-8.11.1\lib, --add-modules=jdk.net, --add-modules=ALL-MODULE-PATH, -Djdk.module.main=org.elasticsearch.server]
[2023-11-25T11:48:04,908][INFO ][o.e.p.PluginsService] [LAPTOP-PT7SFDLG] loaded module [repository-url]
[2023-11-25T11:48:04,908][INFO ][o.e.p.PluginsService] [LAPTOP-PT7SFDLG] loaded module [rest-root]
[2023-11-25T11:48:04,909][INFO ][o.e.p.PluginsService] [LAPTOP-PT7SFDLG] loaded module [x-pack-core]
```



elastic

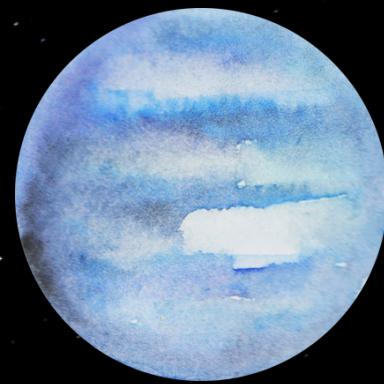
A yellow status indicates that only replicas are unassigned, while a red status implies that one or more primary shards are unassigned.



kibana

Kibana on the other hand, is designed to work only with Elasticsearch and thus does not support any other type of data source.

In order to extrapolate data from other sources, it needs to be shipped into the ELK Stack (via Filebeat or Metricbeat, then Logstash, then Elasticsearch) in order to apply Kibana to it.



```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\mmay> cd D:\kibana-8.11.1\bin
PS D:\kibana-8.11.1\bin> ./kibana.bat
Kibana is currently running with legacy OpenSSL providers enabled! For details and instructions on how to disable see https://www.elastic.co/guide/en/kibana/8.11/production.html#openssl-legacy-provider
{
  "log.level": "info",
  "@timestamp": "2023-11-25T04:50:21.562Z",
  "log": {
    "logger": "elastic-apm-node",
    "agentVersion": "4.0.0",
    "env": {
      "pid": 15488,
      "procTitle": "Windows PowerShell",
      "os": "win32 10.0.22621",
      "arch": "x64",
      "host": "LAPTOP-PT7SFDLG",
      "timezone": "UTC+0700",
      "runtime": "Node.js v18.18.2"
    },
    "config": {
      "serviceName": {
        "source": "start",
        "value": "kibana"
      },
      "commonName": "service_name",
      "serviceVersion": {
        "source": "start",
        "value": "8.11.1"
      },
      "commonName": "service_version",
      "serverUrl": {
        "source": "start",
        "value": "https://kibana-cloud-apm.apm.us-east-1.aws.found.io/"
      },
      "logLevel": {
        "source": "start",
        "value": "info"
      },
      "commonName": "log_level",
      "active": {
        "source": "start",
        "value": true
      },
      "contextPropagationOnly": {
        "source": "start",
        "value": true
      },
      "environment": {
        "source": "start",
        "value": "production"
      },
      "globalLabels": {
        "source": "start",
        "value": "[{"git_rev": "09feaf416f986b239b8e8ad95ecdda0f9d56ebec"}]"
      },
      "secretToken": {
        "source": "start",
        "value": "[REDACTED]"
      },
      "commonName": "secret_token",
      "breakdownMetrics": {
        "source": "start",
        "value": false
      },
      "captureSpanStackTraces": {
        "source": "start",
        "sourceValue": false
      },
      "centralConfig": {
        "source": "start",
        "value": false
      },
      "metricsInterval": {
        "source": "start",
        "value": 120
      },
      "sourceValue": "120s",
      "propagateTracestate": {
        "source": "start",
        "value": true
      },
      "transactionSampleRate": {
        "source": "start",
        "value": 0.1
      },
      "commonName": "transaction_sample_rate",
      "captureBody": {
        "source": "start",
        "value": "off"
      },
      "commonName": "capture_body",
      "captureHeaders": {
        "source": "start",
        "value": false
      },
      "activationMethod": "require",
      "ecs": {
        "version": "1.6.0"
      },
      "message": "Elastic APM Node.js Agent v4.0.0"
    }
  }
}
[2023-11-25T11:50:33.978+07:00][INFO] [root] Kibana is starting
[2023-11-25T11:50:34.000+07:00][INFO] [[node]] Kibana process configured with roles: [background_tasks, ui]
[2023-11-25T11:51:58.834+07:00][INFO] [[plugins-service]] Plugin "cloudChat" is disabled.
[2023-11-25T11:51:58.840+07:00][INFO] [[plugins-service]] Plugin "cloudExperiments" is disabled.
[2023-11-25T11:51:58.840+07:00][INFO] [[plugins-service]] Plugin "cloudFullStory" is disabled.
[2023-11-25T11:51:58.841+07:00][INFO] [[plugins-service]] Plugin "cloudGainsight" is disabled.
[2023-11-25T11:51:58.943+07:00][INFO] [[plugins-service]] Plugin "profilingDataAccess" is disabled.
```

Challenges

1

Scoring and Ranking Logic

Ensure that relevant search results are ranked correctly.

search_app.py

```
# match all terms
elif keyword == "":
    body = {
        'size': page_size,
        'from': page_size * (page_no - 1),
        'query': {'match_all': {}}, # Match all
        'sort': [
            {
                'created': {
                    'order': 'desc' # Sort by
                }
            }
        ]
    }
```

```
# One word query or partial match
body = {
    'size': page_size,
    'from': page_size * (page_no - 1),
    'query': {
        'multi_match': {
            'query': keyword,
            'fields': ['name', 'description'],
            'fuzziness': 'AUTO' # Enable partial
        }
    },
    'sort': [
        {
            '_score': {
                'order': 'desc' # Sort by relevance
            }
        }
    ]
}
```

Console Search Profiler Grok Debugger Painless Lab BETA

History Settings Variables Help 200 - OK 376 ms

```
1
2 GET articles/_search?q=name:zodiac
3
4
5
6
7
8
9
10
11
12
13
14 },
15 "max_score": 5.404893,
16 "hits": [
17 {
18   "_index": "articles",
19   "_id": "qW07AIwBWQn-UlwWwOxv",
20   "_score": 5.404893,
21   "_source": {
22     "name": "The Fibonacci Zodiac",
23     "author": "Russ von Ohlhausen",
24     "picture_src": "https://www.astro.com/im/in/ivccn_ohlhausen_hp.jpg",
25     "tags": [
26       "Fibonacci",
27       "Zodiac"
28     ],
29     "description": "What is 'astrology'? What does it mean to each person who studies it, practices it, profanes it or denounces it? You might think this would be a simple query that is easily answered but as many of us in the field of astrology know, there is no uniform or simple way to address this question. Is it art, is it science, is it a form of spiritual discipline or is it a trailhead to understanding the pathway to our higher potentials?",
30     "is_active": false,
31     "created": "2020/01/27"
32   }
33 },
34 {
35   "_index": "articles",
36   "_id": "JG07AIwBWQn-UlwWwOxu",
37   "_score": 2.909696,
38   "_source": {
39     "name": "The Divine Zodiac - A Window into God's Mind or Simply Humanity's Greatest Work of Art?",
```

Challenges

2

Highlighting Logic

To make it work smoothly in a variety of search situations and managing HTML markup can be tricky.

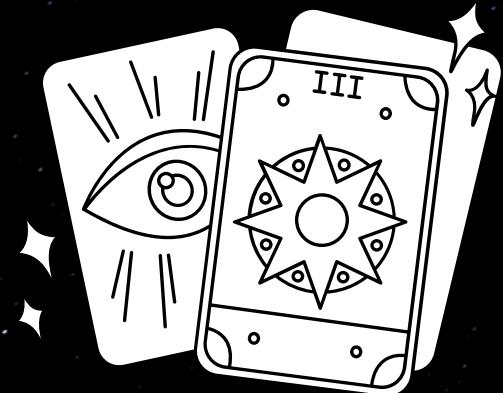
```
@app.template_filter('highlight')
def highlight_filter(s, keyword):
    return re.sub(re.escape(keyword), '<strong class="highlight">' + keyword + '</strong>', s, flags=re.IGNORECASE)
```

```
# Enable highlighting
body['highlight'] = {
    'pre_tags': ['<strong class="highlight">'],
    'post_tags': ['</strong>/.highlight'],
    'fields': {
        'name': {},
        'description': {}
    }
}
```

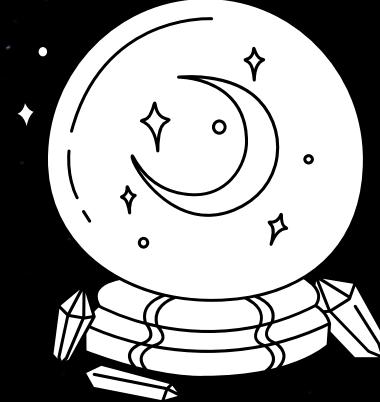
search_app.py

```
def highlight_field(doc, field, filter_type):
    highlighted_field = doc.get('highlight', {}).get(field, [doc['_source'][field]][0])

    # Check the filter type and apply highlighting accordingly
    if filter_type == 'author' and field == 'author':
        return '<strong class="highlight">' + highlighted_field + '</strong>'
    elif filter_type == 'created' and field == 'created':
        return '<strong class="highlight">' + highlighted_field + '</strong>'
    else:
        return highlighted_field
```



Lessons learned



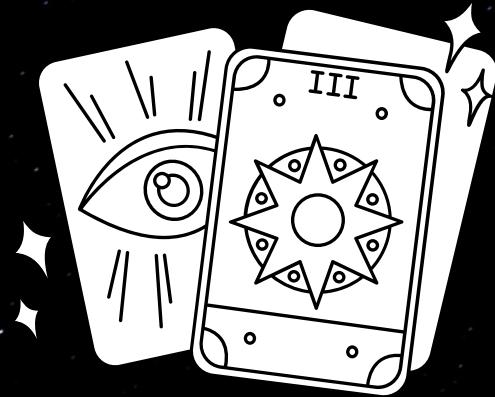
Querying Elasticsearch

Flask and Elasticsearch Integration

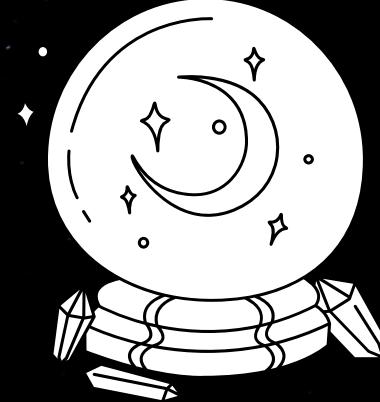
Sorting and Ranking

Comprehending Elasticsearch Index





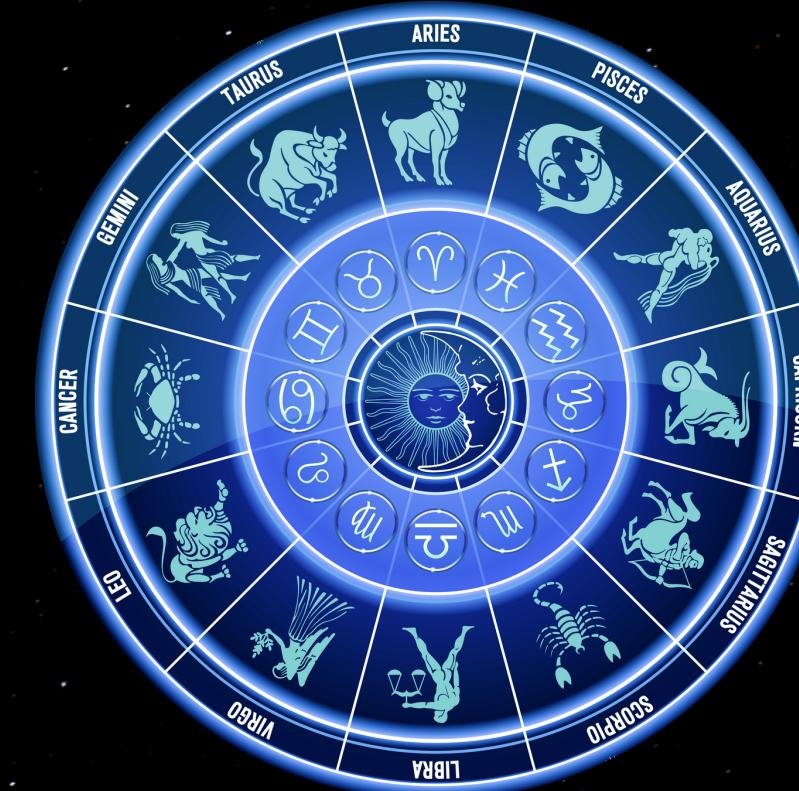
Lessons learned



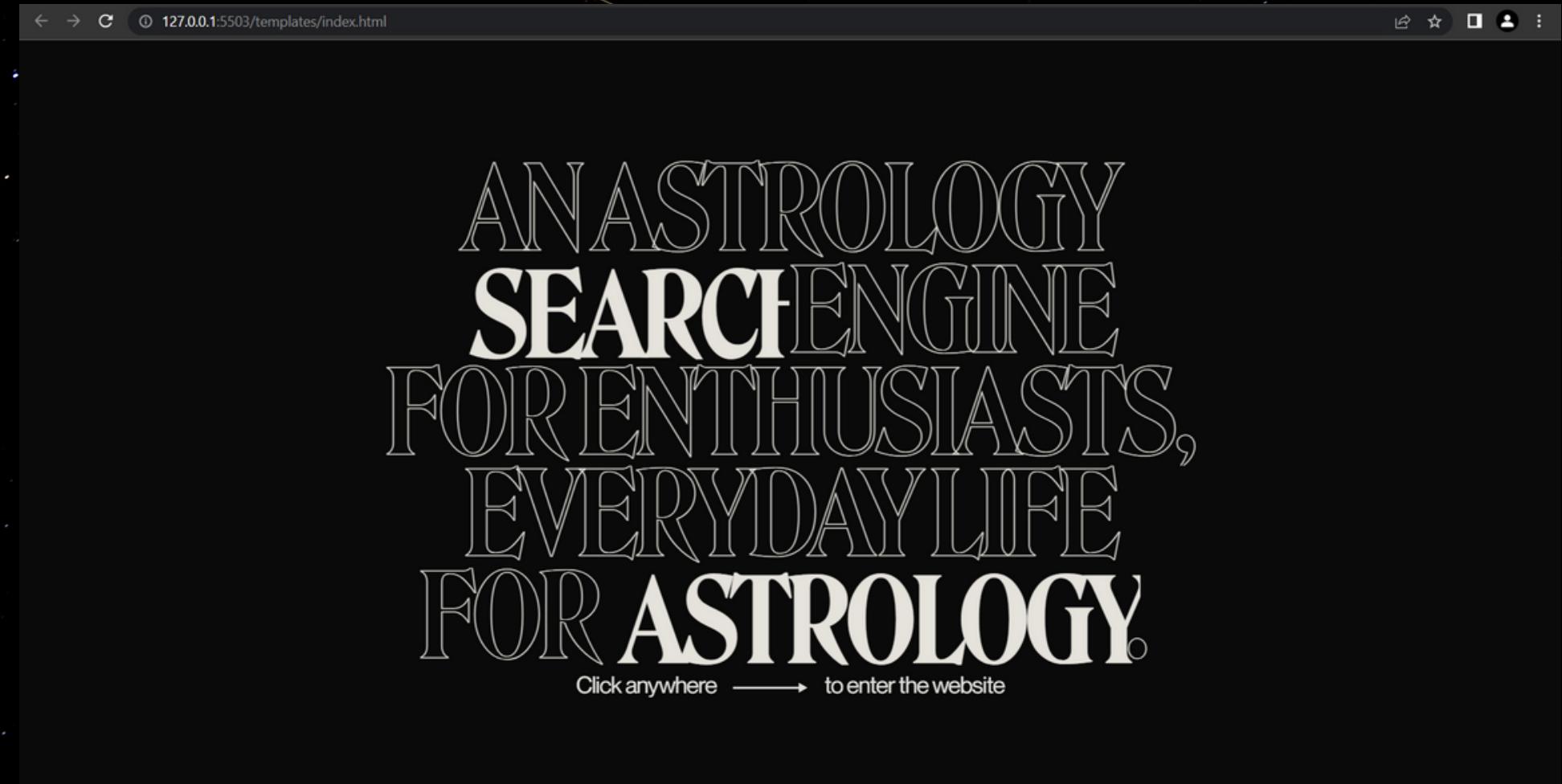
Information Retrieval Principles

Query Parsing and Analysis

Testing



Demonstration of Astrology search system



Thank you

For your attention and If you have any question
feel free to ask.

