

# Text Mining

## <http://mlenodetest1.scholarsportal.info:8004>

The above link is for the MarkLogic REST-API tied to the text mining database which contains an up to date list of the journal-info.xml and toc.xml files found in the production SP Journals db. The collection group and client files are also included, but each collection has been flagged for text-mining based on it's corresponding OUR license like so:

```
<collection name="/collection/elsevier/academicpress" start-date="1993-01-01" license_tag="ScienceDirect" textmining="Yes">...</collection>
```

## Browsing entitled journals alphabetically

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/list>

| args     | options | desc  |
|----------|---------|---|
| rs:ip    |         | We need to pass the client's IP address to determine collection entitlement |
| rs:limit |         | Letter to limit by  |

### Example:

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/list?rs:ip=142.150.195.79&rs:limit=a>

## List all available subject headings

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/heading>

| args   | options                            | desc   |
|--------|------------------------------------|--|
| rs:ip  |                                    | We need to pass the client's IP address to determine if there are entitled journals found within listed subjects |
| rs:sub | leave blank for top level headings |  |

### Example:

This one lists all of the subjects:

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/heading?rs:sub=&rs:ip=142.150.195.79>

And sub-subjects:

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/heading?rs:sub=Arts%20and%20Humanities&rs:ip=142.150.195.79>

## Browsing entitled journals by subject

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/subject>

| args  | options | desc   |
|-------|---------|--|
| rs:ip |         | We need to pass the client's IP address in order to determine collection entitlement |
| rs:s1 |         | Use this to search by parent heading.  |
| rs:s2 |         | Use this for sub / child headings.   |

### Example:

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/subject?rs:s1=Chemistry&rs:ip=142.150.195.79>

or sub heading:

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/subject?rs:s2=Business%20Management&rs:ip=142.150.195.79>

## Retrieve all article URIs for a journal

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/uri>

| args | options  | desc  |
|------|----------|---|
| rs:j |          | The journal URI (issn with no hyphen)             |
| rs:s | optional | Start year  |
| rs:e | optional | End year (defaults to current year if left blank) |

### Example:

<http://mlenodetest1.scholarsportal.info:8004/v1/resources/uri?rs:j=02671905&rs:s=2009&rs:e=2010>

**Note:** The date range here can serve two purposes. If a journal's entitlements have beginning and end dates, we need to pass them here, otherwise the date range can be set according to a user's needs.

---

## Getting the journal-info or toc.xml files

This is just a standard document call. Unfortunately the documents are in xml, so they'll need to be transformed into json by the application:

<http://mlenodetest1.scholarsportal.info:8004/v1/documents?uri=/00059900/journal-info.xml>

<http://mlenodetest1.scholarsportal.info:8004/v1/documents?uri=/00185043/toc.xml>