article-downloader Documentation

Release 1.0

MIT DMSE

Jun 14, 2018

Contents

CHAPTER 1

Contents

1.1 articledownloader.articledownloader module

class articledownloader.articledownloader.**ArticleDownloader**(*els_api_key=None*, *sleep_sec=1*, *timeout_sec=30*)

get_abstract_from_doi(***keywords*)
Returns abstract as a unicode string given a DOI

Parameters

- **doi** (*str*) – DOI string for the article we want to grab metadata for
- **mode** (*str*) – Only supports 'elsevier' for now

Returns An abstract (or None on failure)

Return type unicode

get_dois_from_journal_issn(***keywords*)
Grabs a set of unique DOIs based on a journal ISSN using the CrossRef API

Parameters

- **issn** (*str*) – The ISSN of the journal
- **rows** (*str*) – the maximum number of DOIs to find
- **pub_after** (*int*) – the minimum publication year for DOIs returned
- **mailto** – mailto address for API

Returns the unique set of DOIs as a list

Return type list

get_dois_from_search(***keywords*)
Grabs a set of unique DOIs based on a search query using the CrossRef API

Parameters

- **query** (*str*) – the search string
- **rows** (*str*) – the maximum number of DOIs to find
- **mailto** – mailto address for API

Returns the unique set of DOIs as a list

Return type list

get_html_from_doi(***keywords*)
Downloads and writes an HTML article to a file, given a DOI and operating mode

Parameters

- **doi** (*str*) – DOI string for the article we want to download
- **writefile** (*file*) – file object to write to
- **mode** (*str*) – choose from {'elsevier' | 'springer' | 'acs' | 'ecs' | 'rsc' | 'nature' | 'wiley' | 'aaas' | 'emerald'}, depending on how we wish to access the file

Returns True on successful write, False otherwise

Return type bool

get_metadata_from_journal_issn(***keywords*)
Grabs metadata based on a journal ISSN using the CrossRef API

Parameters

- **issn** (*str*) – The ISSN of the journal
- **rows** (*str*) – the maximum number of DOIs to find
- **pub_after** (*int*) – the minimum publication year for DOIs returned
- **mailto** – mailto address for API

Returns the metadata for the articles according to this ISSN

Return type list

get_pdf_from_doi(***keywords*)
Downloads and writes a PDF article to a file, given a DOI and operating mode

Parameters

- **doi** (*str*) – DOI string for the article we want to download
- **writefile** (*file*) – file object to write to
- **mode** (*str*) – choose from {'crossref' | 'elsevier' | 'rsc' | 'springer' | 'ecs' | 'nature' | 'acs'}, depending on how we wish to access the file

Returns True on successful write, False otherwise

Return type bool

get_title_from_doi(***keywords*)
Returns title of an article as a unicode string given a DOI

Parameters

- **doi** (*str*) – DOI string for the article we want to grab metadata for
- **mode** (*str*) – Only supports 'crossref' for now

Returns A title (or None on failure)

Return type unicode

get_xml_from_doi(***keywords*)
Downloads and writes an HTML article to a file, given a DOI and operating mode

Parameters

- **doi** (*str*) – DOI string for the article we want to download
- **writefile** (*file*) – file object to write to
- **mode** (*str*) – choose from {'elsevier' | 'aps'}, depending on how we wish to access the file

Returns True on successful write, False otherwise

Return type bool

load_queries_from_csv(***keywords*)
Loads a list of queries from a CSV file

Parameters **csvf** (*file*) – file object containing a CSV file with one query per line

Returns a list of queries, processed to be insertable into REST API (GET) calls

Return type list

CHAPTER 2

Indices and tables

- genindex
- modindex
- search

Python Module Index

a

Index