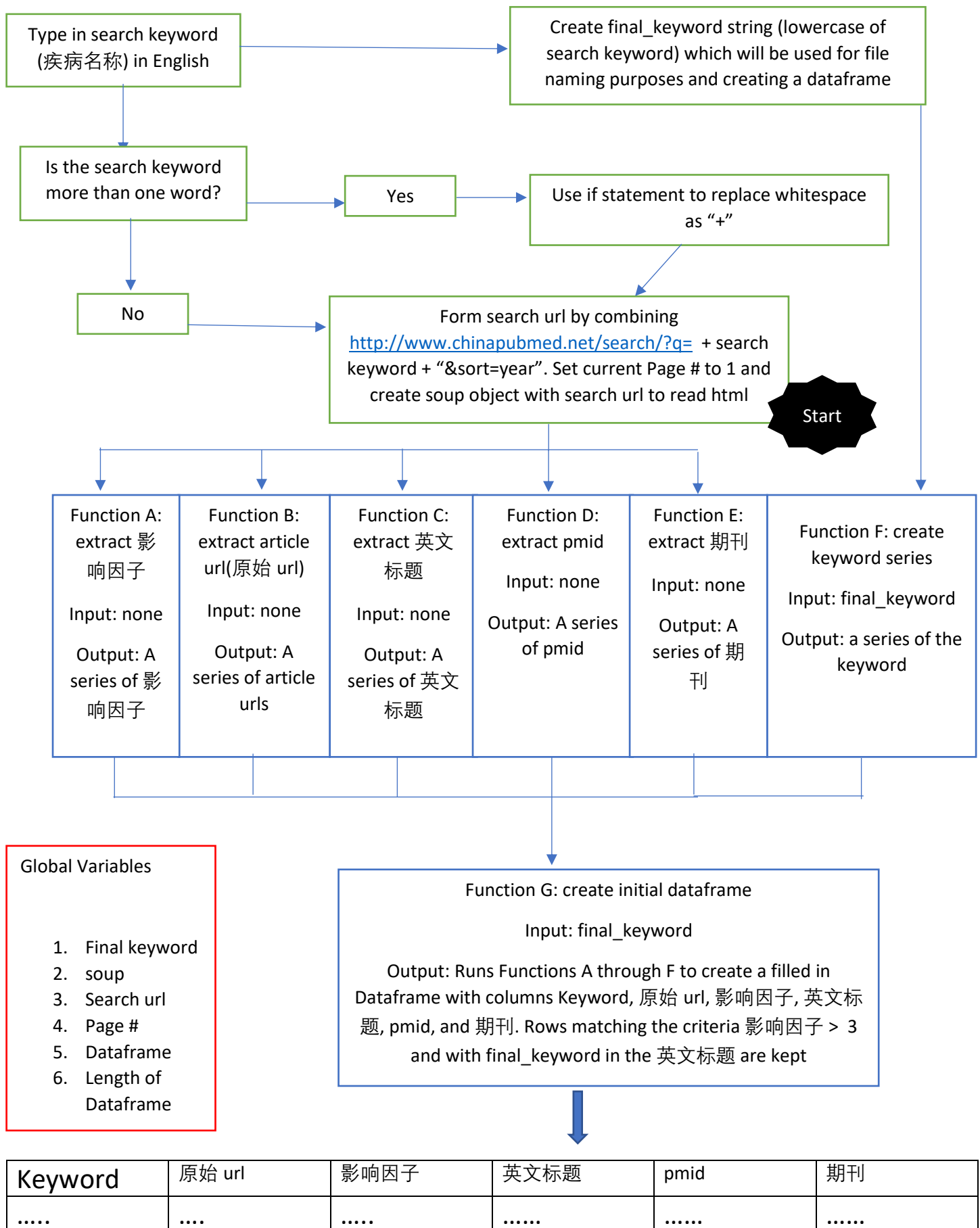
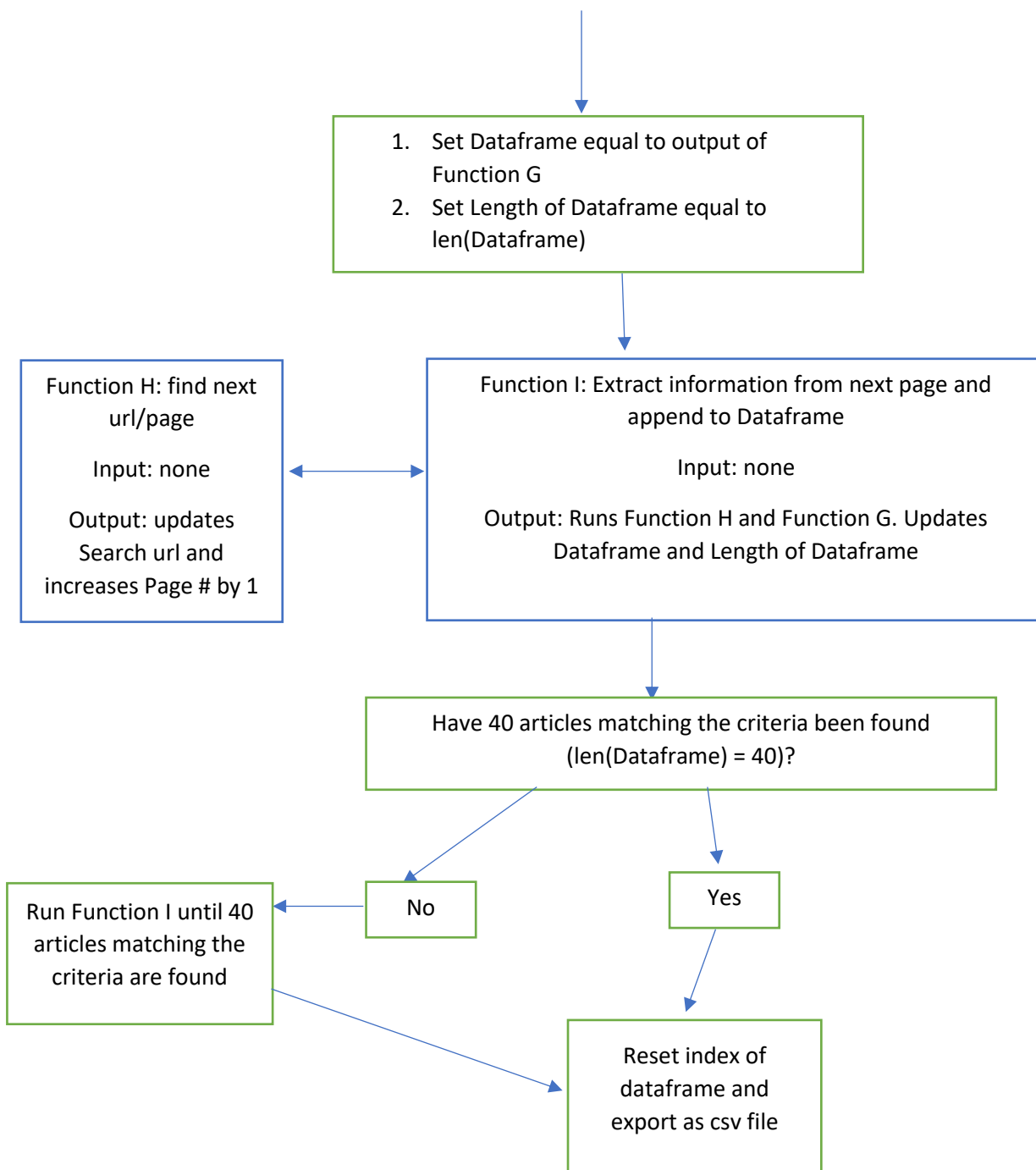


Web Spider Project Framework

1. Project Part A (Input: search keyword; Output: csv file)





Language	IDE	Libraries used
Python	Python 3 Jupyter Notebooks	1. Bs4 2. Requests 3. Pandas 4. re
Note: Path of csv file is currently set to same folder in which the python 3 script is run. This may be adjusted accordingly.		

Test Run (Web Crawler Project Part A.ipynb)

Search keyword = Breast cancer

Initial Search url = <http://www.chinapubmed.net/search/?q=Breast+cancer&sort=year>

中文PubMed

出版年份

20100101 — 20101231

影响因子

0 — 15

期刊(10)

Breast Cancer Res Treat8495

Cancer Res5218

Cancer4269

PLoS ONE4206

Br J Cancer3648

J Clin Oncol3631

Int J Cancer3291

Ann Oncol3032

Oncotarget2924

Anticancer Res2899

搜索矩阵

百度搜索 (中英文)

必应搜索 (中英文)

搜学术 (中英文)

搜Ontology (英文)

中、英文均可搜索

pmid

中英文搜索

高级搜索

热门搜索: lncRNA 外泌体 m6A circRNA

找到308734记录 | 不满意, 请使用左下搜索矩阵

排序: 相关性 发表日期 影响因子

Identifying novel sphingosine kinase 1 inhibitors as therapeutics against breast cancer mammary cancer. The aim of present work is to identify a potent lead compound against breast cancer
关键词: MD simulation MMPBSA calculations SphK1 breast cancer molecular docking
期刊: J Enzyme Inhib Med Chem 影响因子: 4.027 发表日期: 20201201 PMID: 31752564

New nanoprobe for breast cancer cell imaging based on low-density lipoprotein
Many malignant cancers have an increased demand for lipoprotein due to the requirement for lipids
关键词: MRI Nano breast cancer imaging low density lipoprotein
期刊: Artif Cells Nanomed Biotechnol 影响因子: 0.0 发表日期: 20201201 PMID: 31852275

Construction of an immune-related genes nomogram for the preoperative prediction of axillary lymph node metastasis in triple-negative breast cancer
Immune system disorder is associated with metastasis of triple-negative breast cancers (TNBCs)
关键词: ALN metastasis TNBC immune-related genes signature nomogram
期刊: Artif Cells Nanomed Biotechnol 影响因子: 0.0 发表日期: 20201201 PMID: 31858816

Enhanced cytotoxicity of a redox-sensitive hyaluronic acid-based nanomedicine toward different oncocytes via various internalization mechanisms
the antineoplastic efficacy of PTX for breast cancer and melanoma treatment and, thus, deserve f...
关键词: Hyaluronic acid anticarcinoma internalization mechanism vitamin E succinate
期刊: Drug Deliv 影响因子: 3.829 发表日期: 20201201 PMID: 31894722

A potentiated cooperation of carbonic anhydrase IX and histone deacetylase inhibitors against cancer of breast, colorectal and melanoma cancer cells. At the molecular level, this therapeutic regimen
关键词: Carbonic anhydrase IX SLC-0111 SAHA combined therapy histone acetylation
期刊: J Enzyme Inhib Med Chem 影响因子: 4.027 发表日期: 20201201 PMID: 31865754

	A	B	C	D	E	F
1	Keyword	原始url	影响因子	英文标题	pmid	期刊
2	breast cancer	http://www.chinapubmed.net/31752564	4.027	Identifying novel sphingosine kinase 1 inhibitors as therapeutics against breast cancer	31752564	J Enzyme Inhib Med Chem
3	breast cancer	http://www.chinapubmed.net/32003229	3.829	Hydrous icaritin nanorods with excellent stability improves the in vitro and in vivo activity against breast cancer	32003229	Drug Deliv
4	breast cancer	http://www.chinapubmed.net/32308054	3.829	Novel nano-pomegranates based on astragalus polysaccharides for targeting ERα-positive breast cancer and multidrug resistance	32308054	Drug Deliv
5	breast cancer	http://www.chinapubmed.net/32489130	3.829	Preparation of hydroxy genkwanin nanosuspensions and their enhanced antitumor efficacy against breast cancer	32489130	Drug Deliv
6	breast cancer	http://www.chinapubmed.net/32533397	3.677	ASO Author Reflections: Breast Cancer Detection of Baseline Screening MRI in High-Risk Women Who Are Not in the Highest Risk Groups	32533397	Ann Surg Oncol
7	breast cancer	http://www.chinapubmed.net/32886273	3.6	Nomograms for prediction of overall and cancer-specific survival in young breast cancer	32886273	Breast Cancer Res Treat
8	breast cancer	http://www.chinapubmed.net/32886287	3.677	ASO Author Reflections: Immediate Lymphatic Reconstruction: A Proactive Approach to Breast Cancer-Related Lymphedema	32886287	Ann Surg Oncol
9	breast cancer	http://www.chinapubmed.net/32883651	4.187	Using Xigua Video as a Source of Information on Breast Cancer: Content Analysis	32883651	J Med Internet Res
10	breast cancer	http://www.chinapubmed.net/32884019	3.788	Cell-free DNA concentration in patients with clinical or mammographic suspicion of breast cancer	32884019	Sci Rep
11	breast cancer	http://www.chinapubmed.net/32883270	3.091	Patient and physician factors associated with Oncotype DX and adjuvant chemotherapy utilization for breast cancer patients in New Hampshire, 2010-2016	32883270	BMC Cancer
12	breast cancer	http://www.chinapubmed.net/32883235	3.091	The efficacy of etanercept as anti-breast cancer treatment is attenuated by residing macrophages	32883235	BMC Cancer
13	breast cancer	http://www.chinapubmed.net/32883233	3.091	Long non-coding RNA SNHG3 promotes breast cancer cell proliferation and metastasis by binding	32883233	BMC Cancer
14	breast cancer	http://www.chinapubmed.net/32885617	3.418	Intra-CSF topotecan in treatment of breast cancer patients with leptomeningeal metastases	32885617	Cancer Med
15	breast cancer	http://www.chinapubmed.net/32885226	11.397	Corrigendum: Mechanisms of Therapy Resistance in Patient-Derived Xenograft Models of BRCA1-Deficient Breast Cancer	32885226	J Natl Cancer Inst
16	breast cancer	http://www.chinapubmed.net/32880602	24.191	With Hormone Receptor-Positive, ERBB2-Negative Metastatic Breast Cancer: A Randomized Clinical Trial	32880602	JAMA Oncol
17	breast cancer	http://www.chinapubmed.net/32880825	3.403	The interaction between MALAT1 target, miR-143-3p, and RALGAP2 is affected by functional SNP rs3827693 in breast cancer	32880825	Hum Cell
18	breast cancer	http://www.chinapubmed.net/32880442	13.984	Three-Dimensional Human Liver-Chip Emulating Premetastatic Niche Formation by Breast Cancer-Derived Extracellular Vesicles	32880442	ACS Nano
19	breast cancer	http://www.chinapubmed.net/32876911	3.6	treatment of HER2-positive early breast cancer (The NEOPETRA Study)	32876911	Breast Cancer Res Treat
20	breast cancer	http://www.chinapubmed.net/32875678	11.048	Efficient Polysulfide-Based Nanotheranostics for Triple-Negative Breast Cancer: Ratiometric	32875678	Small
21	breast cancer	http://www.chinapubmed.net/32876910	3.6	Risk factors for inflammatory and non-inflammatory breast cancer in North Africa	32876910	Breast Cancer Res Treat
22	breast cancer	http://www.chinapubmed.net/32880016	3.6	Development and validation of prognostic gene signature for basal-like breast cancer and high-grade serous ovarian cancer	32880016	Breast Cancer Res Treat
23	breast cancer	http://www.chinapubmed.net/32544879	4.626	The impact of COVID-19 on and recommendations for breast cancer care: the Singapore experience	32544879	Endocr Relat Cancer
24	breast cancer	http://www.chinapubmed.net/32878898	4.916	Development of point-of-care nanobiosensors for breast cancers diagnosis	32878898	Talanta
25	breast cancer	http://www.chinapubmed.net/32883002	5.492	Targeting Signaling Pathways in Inflammatory Breast Cancer	32883002	Cancers (Basel)
26	breast cancer	http://www.chinapubmed.net/32875534	3.788	Compared DNA and RNA quality of breast cancer biobanking samples after long-term storage protocols in - 80 °C and liquid nitrogen	32875534	Br J Cancer
27	breast cancer	http://www.chinapubmed.net/32880908	5.443	Maintaining breast cancer care in the face of COVID-19	32880908	Br J Surg
28	breast cancer	http://www.chinapubmed.net/32883002	4.132	CDK4/6 Inhibitors in Breast Cancer Treatment: Potential Interactions with Drug, Gene, and Pathophysiological Conditions	32883002	Int J Mol Sci
29	breast cancer	http://www.chinapubmed.net/32883003	5.492	Counteracting Chemoresistance with Metformin in Breast Cancers: Targeting Cancer Stem Cells	32883003	Cancers (Basel)
30	breast cancer	http://www.chinapubmed.net/32883003	5.492	CDK4/6 Inhibitors in Hormone Receptor-Positive Metastatic Breast Cancer: Current Practice and Knowledge	32882980	Cancers (Basel)
31	breast cancer	http://www.chinapubmed.net/32882850	4.132	Molecular Investigation on a Triple Negative Breast Cancer Xenograft Model Exposed to Proton Beams	32882850	Int J Mol Sci
32	breast cancer	http://www.chinapubmed.net/32555534	5.695	Distinct temporal trends in breast cancer incidence from 1997 to 2016 by molecular subtypes: a population-based study of Scottish cancer registry data	32555534	Br J Cancer
33	breast cancer	http://www.chinapubmed.net/32875480	3.6	African breast cancer patients with and without concurrent HIV infection	32875480	Breast Cancer Res Treat
34	breast cancer	http://www.chinapubmed.net/32870818	11.724	HIFs, angiogenesis, and metabolism: elusive enemies in breast cancer	32870818	J Clin Invest
35	breast cancer	http://www.chinapubmed.net/32875465	3.677	with or without Postmastectomy Radiation Therapy for Breast Cancer	32875465	Ann Surg Oncol
36	breast cancer	http://www.chinapubmed.net/32875404	3.833	Precision Medicine for Breast Cancer Utilizing Circulating Tumor DNA: It Is in the Blood	32875404	Curr Treat Options Oncol
37	breast cancer	http://www.chinapubmed.net/32873298	4.002	High expression of MKK3 is associated with worse clinical outcomes in African American breast cancer patients	32873298	J Transl Med
38	breast cancer	http://www.chinapubmed.net/32546834	5.695	A dynamic web-based decision aid to improve informed choice in organised breast cancer screening. A pragmatic randomised trial in Italy	32546834	Br J Cancer
39	breast cancer	http://www.chinapubmed.net/32882267	3.528	Effects of metformin on adipose-derived stromal cell (ADSC) - Breast cancer cell lines interaction	32882267	Life Sci
40	breast cancer	http://www.chinapubmed.net/32882354	5.317	The role of circulating tumor cells in breast cancer and implications for radiation treatment decisions	32882354	Int J Radiat Oncol Biol Phys
41	breast cancer	http://www.chinapubmed.net/32882276	4.198	The R251Q mutation of LSD1 promotes invasion and migration of luminal breast cancer cells	32882276	Int J Biol Macromol
42	breast cancer	http://www.chinapubmed.net/32882665	3.687	, revealing potential targets and new biomarkers for basal-like breast cancer prognosis	32882665	Int Immunopharmacol

2. Project Part B (Input: csv file, Output: csv file)

Language	IDE	Libraries used/to be used
Python	Python 3 Jupyter Notebooks	1. Bs4 2. Requests 3. Pandas 4. Re 5. Numpy
Note: Path of csv file is currently set to same folder in which the python 3 script is run. This may be adjusted accordingly.		

Structure:

1. Create function to read csv file from Part A as dataframe1 and return dataframe1
2. Create functions to extract the following information
 - a. Function A: 关键词
 - b. Function A.2: 关键词当前网页的网址
 - c. Function B: 中文标题
 - d. Function C: 中文正文
 - e. Function D: 作者
 - f. Function D.2: 作者当前网页的网址
 - g. Function E: 英文正文
 - h. Function F: 发表日期
 - i. Function G: 官网
 - j. Function H: 免费下载链接
 - k. Function I: 全文下载链接
3. Create a function (J) to create and fill new columns in dataframe (new columns = 关键词, 中文标题, 中文正文, 作者, 英文正文, 发表日期, 官网, 免费下载链接, 全文下载链接)
4. Use appropriate functions from Project Part A to crawl through up to five pages of urls extracted from Functions A.2 and D.2 Create a new dataframe2 for this extracted information.
5. Repeat function J to extract information for missing columns for this new dataframe2.
6. Append dataframe2 to dataframe1 to make dataframe_final. Reorganise order of columns if necessary. Reset index of dataframe_final and export as csv. Desired format is as shown:

Keyword	原始 url	关键词	中文标题	中文正文	英文标题	作者	英文正文	pm id	期刊	影响因子	发表日期	官网	免费下载链接	全文下载链接
---------	--------	-----	------	------	------	----	------	-------	----	------	------	----	--------	--------

[illegible]