Advanced Engineering Technology Applications

The system of automated formatting conversion for bibliometric data

Lyazzat Atymtayeva[1], *Arailym Abdrakhym*[2] *and Shynar Toktar*[3]

[1]Department of Computer Science, Suleyman Demirel University,Kaskelen, Kazakhstan
[2]Department of Computer Science, Suleyman Demirel University,Kaskelen, Kazakhstan
[3]Department of Computer Science, Suleyman Demirel University,Kaskelen, Kazakhstan

Received: ..., Revised: ..., Accepted: ...
Published online: ...

Abstract: This paper describes a system for advanced search of bibliometric data. The result contains a portfolio of the author's publications, and the user can export it as a CSV or PDF file. The main idea of the project is to reduce the time for data collection from various sources of famous scientific bibliometric databases.

Keywords: Bibliometric Data, Web applications, API, data collection

1 INTRODUCTION

The complexity of collecting and analyzing bibliometric data has recently become the task of numerous studies due to the rapid growth in the amount of data, the variability of their format, the large dispersion and heterogeneity of data sources.

At the same time, as a rule, each database that collects and processes bibliometric data, such as Scopus [1], Web of Science [2], EBSCO [3], IEEE Xplore [4], Elsevier [5] and others, has its own format, processing methods , storage and data display.

For scientists engaged in research, the main source of obtaining new research data, promising directions, methods and methodologies are the research publications. On the other hand, any organizations, such as universities, research centers, etc., are interested in increasing their rating in research sector due to the growth of the number and quality of publications in scientific journals, which are also a source of data for bibliometric databases.

There are university ranking systems, such as the Shanghai Organization for World University Ranking (ARWU)[6], QS World University ranking [7], etc., which determine the university ranking based on several criteria. One of the criteria is the ranking of publications of university scientists. In the most cases this index defines the preferences of applicants (potential students) which are in the process of the place of study selection. Therefore, for academic and research organizations, one of the priority areas of their strategy is considering the quantity and quality of publications of their scientists and motivating them to increase the number of papers in high quality peer-review journals. This is also a strategy for administrative organizations such as the Ministry of Education and Science.

It becomes more difficult to monitor and consider the publications of scientists, since the number of bibliometric databases grows. The data collection experience in organizations usually provides the manual processing by searching data for each database and manually compiling a list of publications.

This process is time consuming since it is necessary to correct format errors, find for missing information, search in different sources and databases, adapt it for language translations, decipher abbreviations and other problems. If the number of publications of scientists exceeds 100, their processing takes about 50-60 minutes. More difficult is the task of identifying experts in a particular field of knowledge for the effective work of scientific dissertation councils, expert commissions for the project selection and other similar committees.

Creating a single automated format for all publications of a scientist from different sources in one place is an actual task. This reduces the time for collecting and processing data, eliminates manual work and allows us to automatically update the current state of the scientist publications as they become available, upgrade the information, and simplify data analysis.

* Corresponding author e-mail: lyazzat.atymtayeva@sdu.edu.kz

This paper is devoted to the description of a system for a single format of publications from different sources, developed on modern programming tools and displaying a convenient user interface.

2 SYSTEM STRUCTURE AND DESCRIPTION

The developed system represents a client-server application on the base of a built-in MVC architecture with the focus on the front-end and back-end parts.

2.1 Front End System Design

For the developing system one of the priority tasks was the development of a convenient user interface (user- oriented design). Since the main users of the system are scientists and administrators of scientific organizations, the design development strategy was based on a minimalistic and task-oriented design.

Web application usability focused on the three main criteria: efficiency, benefit ability and optimal result. At the same time, the formation of the user's mental model can be based on the answers to the following questions:

1) Can users find what they are looking for? (Efficiency criterion)

2) Do they spend a lot of time searching and submitting a request? (benefit ability criterion)

3) Is it possible to save search results in PDF and CSV format in a short time? (Optimal results criterion)

The design of the application considered the following main sources, as the most used and highly rated among scientists. But this list can be expanded.

- Google Scholar[8]
- Scopus[1]
- IEEE Xplore Digital Library[9]
- Web of Science [2]
- DOAJ open global trusted [10]
- National Library of Medicine Pubmed.gov [11]
- EBSCO [3]

The scheme of interaction between the user and the system can be simplified through the use-case diagram (see Fig. 1).

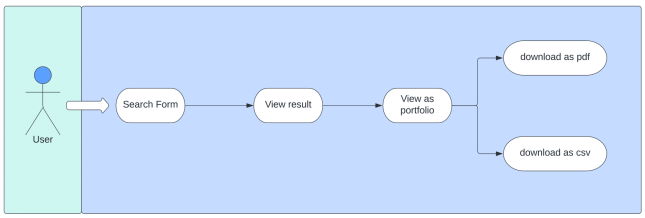


Fig. 1: Use case diagram

We consider the following design elements for the search process:

- The main page of the site for searching and additional pages for using the site.
- Publishing of all list of papers collected from the author's external sites in the form of a table.
- Customizing the result required files. The user can search for an author by his full name or ID, view his or her portfolio, and download it as a PDF or CSV (see Figure 1).

2.2 The assembling of the system server part (back-end)

For the backend, we used the Django Rest framework to create endpoints using a functional view. The system provides two types of searches: by first and last name and by ScopusID (for those who want to get a list only from the Scopus source). The surname and name of the author are entered in English since the world bibliographic systems use English [8].

2.3 Data storage

Since the system uses up-to-date data at the current moment from different sources, there is no need to statically store the data. This decision was made based on the following advantages: Because every time we have a new request, the User does not miss the opportunity to get an updated result. In this case, the query results do not need to be saved.

2.4 Export processes

The main advantage of the developed web application is the ability to export the result. We have chosen two types of export: export to PDF and export to excel. We used the LATEX system tool to create the PDF. So, we could create a simple and standard Latex Template. To visualize the list of publications, we used the APA style [12] in the following format: Authors (Year of publication), Title, Publisher, page range. The Django application django-render pdf we used for the rendering the template and list of publications [13].

To simplify the process of formatting the list of results, we have also added the Export to Excel option. With Django HttpResponse we could create a csv object and with the CSV module we may add the data (see Figure 2).

```
def exporttocsv(request):
    with open("exp.pickle", "rb") as fp:  # Unpickling
        PubList = pickle.load(fp)
    response = HttpResponse(content_type='text/csv')
    response['Content-Disposition'] = 'attachment; filename="csvpublication.csv"'
    response.write(u'\ufeff'.encode('utf8'))
    writer = csv.writer(response, delimiter=';')

    writer.writerow(['Site', 'Title', 'Authors', 'Type', 'HtmlLink',  'Publisher', 'Year', 'Pages'])
    Res = PubList[1][1:]
```

Fig. 2: Code of export to csv

3 ANALYSIS OF OTHER WEB SYSTEMS API's OPPORTUNITIES

To analyze and create our system front end design, we considered seven similar websites and explored their APIs.

1) Google Academy [7]. Here we have taken the opportunity to search for keywords or authors or non- fiction in one place.

2) PabMed system [11]. It is intended only for searching papers.

3)IEEE Xplore[4]. We explored the possibilities of searching for books, conferences, courses, journals, standards, authors, search for citations.

4) Scopus[1]. Here we took the advantage of the ability to search by author name and ORCID.

5) EBSCO[3]. We analyzed the search for articles, for books and magazines.

6) Web of Science [2]. Here we have identified the possibilities of searching by ISSN and the keywords.

7) DOAJ [10]. The system allows us searching by journals, articles, title, abstract, topic, author.

In all systems, the site-wide search form is located on the main page. Advanced Search can be provided by author name and ID.

4 RESULTS AND TECHNOLOGIES

As a result of the system development, the web application performs an advanced search using the following user interface (see Fig. 3)

Examples of search results are shown in Figure 4.

Figure 5 shows the result of converting to pdf format.

During development, we used the following technologies:

1) Figma as an accessible online service for creating a basic site design scheme and general visualization.

2) Bootstrap to create a responsive website for mobile devices.

3) GitHub as a project and code management system, as well as a social media platform for developers.

4) Trello for project and team management.

5) Postman as a platform for creating, using and testing APIs

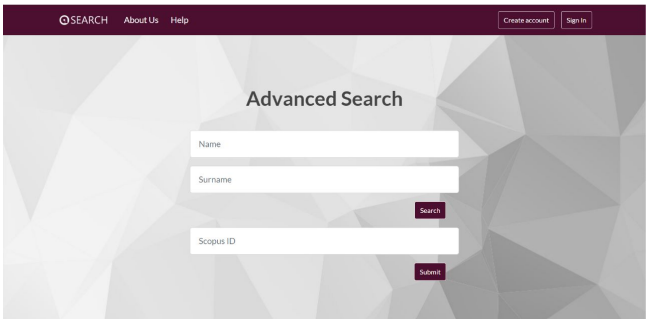


Fig. 3: Screen page of Advanced Search

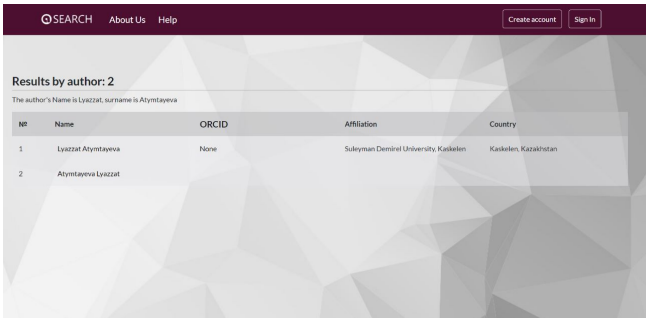
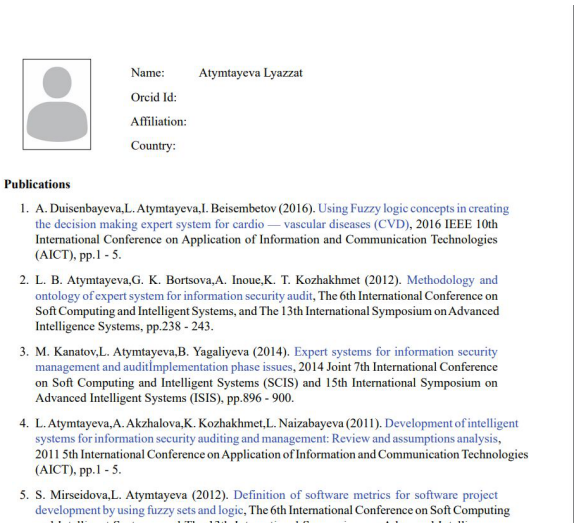


Fig. 4: The screen page of search results


Name: Atymtayeva Lyazzat
Orcid Id:
Affiliation:
Country:

Publications

1. A. Duisenbayeva,L. Atymtayeva,I. Beisembetov (2016). Using Fuzzy logic concepts in creating the decision making expert system for cardio — vascular diseases (CVD), 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), pp.1 - 5.
2. L. B. Atymtayeva,G. K. Bortsova,A. Inoue,K. T. Kozhakhmet (2012). Methodology and ontology of expert system for information security audit, The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems, pp.238 - 243.
3. M. Kanatov,L. Atymtayeva,B. Yagaliyeva (2014). Expert systems for information security management and auditImplementation phase issues, 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS), pp.896 - 900.
4. L. Atymtayeva,A. Akzhalova,K. Kozhakhmet,L. Naizabayeva (2011). Development of intelligent systems for information security auditing and management: Review and assumptions analysis, 2011 5th International Conference on Application of Information and Communication Technologies (AICT), pp.1 - 5.
5. S. Mirseidova,L. Atymtayeva (2012). Definition of software metrics for software project development by using fuzzy sets and logic, The 6th International Conference on Soft Computing

Fig. 5: Screen Page of file PDF transformed

5 DISCUSSION

The developed system provides many opportunities to facilitate the search and processing of bibliometric data. However, there are some technical problems that hinder the process of accessing data sources. One of them is access to open APIs (Applied Programming Interface). Many websites, including the database systems listed above (SCOPUS, Google, etc.), have open access to API functions. These public functions allow us searching for data by using special parsing tools.

By analyzing the bibliometric database systems in Kazakhstan, we encountered the problem in lacking such open API functions. In spite of it, the analysis could be provided by using the scrapping tool. However, any changes in the sources, such as the domain address of the site, a change in the style of data collection, any movements can affect the functionality of the system and make the search impossible.

One of the proposed this problem solutions can be creation of free APIs, at least for scientific publications in Kazakhstan, so that we could easily monitor and determine trends in the development of scientific areas.

6 CONCLUSION

The developed system for the analysis of bibliometric data and the formation of a unified format for scientist publications could be considered as useful, at least. It allows the search process simplifying and automating the authors' publications accounting. Universities and scientific organizations could be the main beneficiaries of this system.

Another topical problem that the extended version of this system could solve is the search for experts in a particular field, the formation of their portfolio and recommendations for potential research. This is relevant for expert councils, dissertation commissions, and other similar committees.

References

[1] Scopus. url: https://www.scopus.com/home.uri
[2] Web of Science. url: https://clarivate.com/webofsciencegroup/solutions/web-of-science/
[3] EBSCO. url: https://www.ebsco.com
[4] IEEE Xplore. url: https://ieeexplore.ieee.org/Xplore/home.jsp
[5] Elsevier Developer Portal. url: https://dev.elsevier.com
[6] Shanghai university ranking ARWU, url: https://www.shanghairanking.com/rankings/arwu/2022
[7] QS World University Rankings, url:https://www.topuniversities.com/qs-world-university-rankings
[8] Google Scholar, url:https://scholar.google.com/
[9] IEEE Xplore Digital Library. url: https://www.ieee.org/publications/subscriptions/products/mdl/ieeexplore-access.html
[10] DOAJ open global trusted. url: https://doaj.org/
[11] Pubmed.gov National Library of University, url: https://pubmed.ncbi.nlm.nih.gov/
[12] Massey University, url: https://owll.massey.ac.nz/referencing/referencing-styles.php
[13] Pipy.org. url: https://pypi.org/project/django-renderpdf/