



(REVIEW ARTICLE)



## Application of fast batch obtaining cited frequency of web of science with VBA

Qiu Xiaohua \*

*School of International Trade and Economics, Central University of Finance and Economics, Beijing, China.*

International Journal of Science and Research Archive, 2024, 12(01), 2568–2572

Publication history: Received on 28 April 2024; revised on 08 June 2024; accepted on 11 June 2024

Article DOI: <https://doi.org/10.30574/ijrsra.2024.12.1.1031>

### Abstract

Citation frequency is currently recognized as one of the evaluation indicators in the academic field. Therefore, research related to evaluation requires obtaining the citation frequency of literature, especially the citation frequency of Web of Science. Using programming languages like Java to batch obtain the citation frequency of Web of Science through the Web of Science API interface requires a certain programming foundation and specialized programming tools, which are generally difficult for researchers to master. This article explores a convenient and fast method to batch obtain the citation frequency of Web of Science using VBA. Using Web of Science's literature matching retrieval service, it is possible to obtain the citation frequency of literature in batches. The literature matching retrieval service is implemented in Excel using VBA code. Through testing and practice, it has been proven that batch and rapid retrieval of Web of Science citation frequency can be achieved using VBA in Excel. This article not only provides the entire runnable source code for easy expansion but also gives practical examples for reference.

**Keywords:** VBA; Web of Science; Literature Matching Retrieval Service

### 1. Introduction

The citation frequency is one of the main indicators currently used to evaluate academic papers, academic journals, researchers, and research institutions. Many evaluation metrics are based on citation frequency, such as the widely accepted journal impact factor and the h-index for evaluating researchers, among others. The practice of using citation frequency as the sole or absolute indicator for evaluation has been criticized by scholars. However, due to the relative objectivity and authenticity of citation frequency, which can basically reflect the actual behavior of researchers in using literature, it is still considered one of the most reliable and objective evaluation indicators.

Many universities in China mainly rely on citation frequency to evaluate papers and researchers, especially the personnel department in talent evaluation and the research department in paper rewards. The most widely recognized and trusted citation frequency is Thomson Reuters' Web of Science, which is also one of the most commonly used indicators among many evaluation metrics.

The most common way to obtain the citation frequency of Web of Science papers is to log in to Web of Science for queries. However, since the citation frequency of papers is updated regularly, manually querying a large number of updated citation frequencies can be very time-consuming and labor-intensive. Developing programs using the Web of Science interface can avoid this problem. However, for most people, developing programs using programming languages (such as Java, .NET, PHP, etc.) may seem a bit difficult, and the language code is also difficult to understand and master. In contrast, VBA is different. It does not require special development tools, and the language is relatively easy to understand and master. Therefore, this article will study how to use VBA code programs to quickly and batch obtain the citation frequency of Web of Science, and provide executable program code and steps for reference and learning by relevant researchers.

\* Corresponding author: Qiu Xiaohua, Email: [cufeqiuxiaohua@163.com](mailto:cufeqiuxiaohua@163.com)

## 2. Article Match Retrieval Service(AMR)

The article matching retrieval service is an API service introduced by Thomson Reuters for the Web of Science Core Collection database. Through this API, it is possible to query the real-time citation frequency of literature included in the Web of Science Core Collection. The content of this section mainly refers to Thomson Reuters' technical document: "Thomson Reuters Links Article Match Retrieval Service Version 1.5." Please note that access to this document requires authorization through a username and password, so the link cannot be provided here.

### 2.1. Apply for the Article Matching Retrieval Service

Users who have subscribed to the Web of Science Core Collection database can use this API service for free. However, before using the Document Matching Retrieval Service, users need to apply separately to Thomson Reuters to activate this service. The application address is: <http://ip-science.thomsonreuters.com/info/terms-amr/>. For institutional users, this service is IP-controlled, meaning that the Document Matching Retrieval Service can only be used on the server associated with the IP address approved for the service. Therefore, it is essential to determine the IP address for using the service before submitting the application form, as the IP address cannot be changed once the service is activated. After confirming the IP address, users can submit an online service application to Thomson Reuters. Generally, it takes one working day to activate the IP address for using the Document Matching Retrieval Service, and the relevant information for the service will be sent to the applicant's email address [1].

### 2.2. Data returned from literature matching retrieval services

The literature matching retrieval service returns result data in XML format. This service can obtain real-time bibliographic data for Chinese literature in the Web of Science, including citation frequency, unique identifiers, DOIs, link URLs, URLs citing the literature, and URLs of related literature, among other information. The specific elements of the returned data are shown in Table 1.

In addition to returning bibliographic data, the literature matching retrieval service can also obtain the trend chart URL of the five-year impact factor of the journal in which the literature is published. Through this URL, the impact factor of the journal for the past five years can be queried. The prerequisite for obtaining the journal's impact factor is that the institution must have purchased the Journal Citation Reports.

**Table 1** Data Returned by the Article Match Retrieval Service

Elements of data returned by AMR	Description
Times Cited	Citation frequency of the literature
UT	Unique identifier of the literature in Web of Science
DOI	Digital Object Identifier (DOI)
PMID	Unique identifier of the literature in the U.S. National Library of Medicine's PubMed
Source URL	URL linking to the detailed information display page of the literature in Web of Science
Citing Article URL	URL linking to the list of literature citing the literature in Web of Science
Related Records URL	URL linking to the list of literature related to the literature in Web of Science
Journal Impact Factor Trend Graph URL	URL linking to the five-year trend chart of the Journal Citation Reports (JCR) journal impact factor

### 2.3. Data requested by the literature matching retrieval service

The literature matching retrieval service is actually a type of web service. To obtain the data in Table 1, you must send an HTTPS post request to Thomson Reuters' address for handling literature matching retrieval services. The specific address for this purpose is: <https://ws.isiknowledge.com/cps/xrpc>. By sending an authorized server to this address, you can retrieve the data in Table 1.

When sending a request to the literature matching retrieval service, you need to provide the information of the requested literature in XML format. The elements included in the request are shown in Table 2. The elements in Table 2 are divided into two categories: identifier elements and normal bibliographic data elements. When sending a request, there are two ways to combine the elements:

- Including identifier elements: To ensure the accuracy of the data, a request generally includes at least one identifier element, such as DOI, UT, or PMID. Other normal bibliographic data elements are optional because usually one identifier element is enough to retrieve the required literature information.
- Not including identifier elements: When a request does not include any identifier elements, to determine a unique literature, the request must include at least four bibliographic data elements: journal title, volume, issue, and starting page or article number. This combination of four elements can uniquely determine a literature.

**Table 2** Data Sent in Requests to the Article Match Retrieval Service

Elements of the Request	Categories	Description
DOI	Identifier	Digital Object Identifier (DOI)
UT	Identifier	Unique Article Identifier in Web of Science
PMID	Identifier	Unique Identifier in the U.S. National Library of Medicine's PubMed
Article Number	Bibliographic Data	Unique Number Assigned by the Journal Publisher to the Article
ISSN	Bibliographic Data	International Standard Serial Number (ISSN)
ISBN	Bibliographic Data	International Standard Book Number (ISBN)
Author	Bibliographic Data	Author(s)
Year	Bibliographic Data	Publication Year
Article Title	Bibliographic Data	Title of the Literature
Journal Title	Bibliographic Data	Journal Title
Volume	Bibliographic Data	Volume
Issue	Bibliographic Data	Issue
Start page	Bibliographic Data	Starting Page Number
Book Title	Bibliographic Data	Title of the Book or Conference Proceedings
Book Series Title	Bibliographic Data	Title of the Series

### 3. Using Excel VBA Code to Quickly Obtain WOS Citation Frequency

Visual Basic for Applications (VBA) is a programming language provided by Microsoft to extend the functionality of its desktop software, especially Excel. This programming language is powerful yet relatively simple in syntax, making it easy to learn and widely used for improving office productivity. VBA can send HTTP requests and process XML files, making it possible to retrieve the citation frequency of literature in Web of Science using VBA. This approach has the following advantages:

- Simple and easy-to-learn syntax: VBA has a simple syntax, making it accessible even to those without a programming background. The development tools and runtime environment for VBA are simple, and you can edit and execute code directly in Excel.
- Open code for easy extension: VBA code is open, allowing you to modify it to meet personalized functional requirements.
- High program execution efficiency: The citation frequency in Web of Science is updated weekly, so it needs to be updated regularly. VBA can be executed in Excel, making it convenient and efficient to obtain the citation frequency of literature in real time.

Below is an example of VBA code in Excel to retrieve the citation frequency of literature in Web of Science:

```

Dim row_i As Integer
Dim xmlHTTP As MSXML2.XMLHTTP
Set xmlHTTP = CreateObject("msxml2.xmlhttp")
Dim xmlDom As MSXML2.DOMDocument
Set xmlDom = CreateObject("MSXML2.DOMDocument")
xmlDom.async = False
Dim xmlRequest As String
xmlRequest = "<?xml version=""1.0"" encoding=""UTF-8"" ?>" & _
    "<request xmlns=""http://www.isinet.com/xrpc42"" src=""app.id=USTB"">" & _
    "<fn name=""LinksAMR.retrieve"">" & _
    "<list>" & _
    "<map>" & _
    "</map>" & _
    "<map>" & _
    "<list name=""WOS"">" & _
    "<val>timesCited</val>" & _
    "</list>" & _
    "</map>" & _
    "<!-- LOOKUP DATA -->" & _
    "<map>"

For row_i=1 to 50
    xmlRequest= xmlRequest & _
        "<map name=""cite_id "" & row_i & """">" & _
        "<val name=""doi"">" & Worksheets(1).Cells(row_i,1).Value & "</val>" & _
        "</map>"
Next
    xmlRequest= xmlRequest & _
        "</map>" & _
        "</list>" & _
        "</fn>" & _
        "</request>"

xmlDom.LoadXML (xmlRequest)
xmlHTTP.Open "POST", "https://ws.isiknowledge.com/cps/xrpc", False
xmlHTTP.send (xmlDom.XML)
Dim xmlNodes As Object
Set xmlNodes = xmlHTTP.responseXML.SelectNodes("//map[@name=""WOS""]")

For row_i = 1 To 50
    Worksheets(1).Cells(row_i, 2).Value = xmlNodes(row_i - 1).Text
Next
Set xmlNodes = Nothing
Set xmlHTTP = Nothing
Set xmlDom = Nothing

```

In the above VBA code, the DOI is used as the unique identifier for literature queries, meaning that the citation frequency of the literature is obtained through its DOI. The code retrieves the citation frequency of DOIs in the first column of an Excel sheet, from the first row to the fiftieth row. The most crucial aspect of the VBA code is the construction of the XML request string and the parsing of the XML response. The literature matching retrieval service requires that the requested elements be in XML format, so when retrieving multiple pieces of literature in a single request, an XML request must be constructed. Similarly, the results returned by the literature matching retrieval service are also in XML format, requiring the construction of a Document Object Model (DOM) to parse the XML results and fill the parsed results into the corresponding cells of an Excel sheet. The literature matching retrieval service stipulates a maximum of 50 documents per request, so when retrieving a large number of documents, it can be done in batches. Batch retrieval is simple, requiring only the modification of the row number variable in the for loop, with no need to modify other parts of the code, thereby improving the efficiency of information retrieval. It is important to note that when running the above code, the Microsoft XML 6.0 Dynamic Library must be referenced in the Tools menu of the Excel VBA code interface.

---

#### **4. Application Practice**

In the specific practical application, the author utilized VBA code to quickly and efficiently retrieve the citation frequencies of Web of Science literature in batch, thereby improving the efficiency of literature processing. The author's application practice mainly focuses on the following two aspects:

##### **4.1. Discipline Evaluation**

Discipline evaluation is a task carried out by university libraries in response to the development planning needs of the university in recent years, and it is also one of the main ways for university libraries to provide discipline services. The main task of discipline evaluation is to evaluate and analyze the current situation of the disciplines involved in the university and the competitiveness of these disciplines in domestic and global contexts. One of the main tasks of discipline evaluation is to statistically analyze highly cited papers in the disciplines of the university. Although highly cited papers can be directly obtained from Web of Science, it can be very cumbersome to retrieve them each time due to the rapid updates in citation frequency. Using VBA can greatly improve efficiency.

##### **4.2. Paper Rewards**

Paper rewards refer to the annual rewards given by the research office of the university for academic papers published in the previous year. The library is responsible for providing paper data to the research office, especially data on papers with high citation frequencies. Universities generally reward papers with high citation frequencies published in the past five years separately. Therefore, it is necessary to update the citation frequencies of papers published by the university in the past five years. Since this is a long-term task, the library will compile the data on papers published by the university each year. Therefore, based on the data from the previous five years, the author used VBA to batch update the citation frequencies of all papers published in the past five years and then identified papers with the required high citation frequencies according to different disciplines.

---

#### **5. Conclusion**

The study successfully demonstrates that VBA in Excel can be effectively utilized to batch retrieve citation frequencies from Web of Science, providing a simpler and more accessible alternative to programming languages like Java. By leveraging the literature matching retrieval service, the research offers a practical solution for researchers who may lack advanced programming skills. The provision of runnable source code and practical examples further enhances its utility and adaptability. This study not only streamlines the process of obtaining citation data, thus supporting academic evaluations, but also paves the way for future enhancements in user-friendly data retrieval methods in the research community.

---

#### **References**

- [1] Thomson Reuters Links Article Match Retrieval Service[Internet].2023 April 20[cited 2024 May 10]. Available from [http://wokinfo.com/products\\_tools/products/related/amr/](http://wokinfo.com/products_tools/products/related/amr/)