International Journal of Information Science and Management
Vol. 15, No. 2, 2017, 135-143

Design and Implementation of Iranian Digital Object Recognizer (DOR) Phase 1: Persian Articles

Mohammad Reza Ghane
Associate Prof. in Knowledge & Information Science. Regional Information Center for Science and Technology, RICeST, Iran
Corresponding author: ghane@ricest.ac.ir

Sara Koleini
M.S. in Computer Engineering,
Senior expert staff of network engineer,
RICeST, Iran
koleini@ricest.ac.ir

Mohammad Bagher Dastgheib
Assistant Prof. in Computer Engineering
RICeST, Iran
dastgheib@ricest.ac.ir

Abstract

The Digital Object Recognizer (DOR) is a system for identifying Persian digital objects on the web. DOR is a unique alphanumeric string assigned by Regional Information Center for Science and Technology (RICeST) to any entity to identify content and provide a persistent link to its location on the Internet digital networks. The paper describes the construction and functions of the DOR system for Persian articles and then introduces and outlines the underlying architecture regarding DOR system. The study concluded that using DOR will enable Iranian journal publishers assign it to both print and electronic journal articles; finally enumerated the benefits of DOI for stakeholders in the scholarly communication process.

Keywords: Digital Object Recognizer (DOR), Persian Articles Identifier, Persian Journals Identifier, Iran.

Introduction

With the increasing amount of electronic resources in the network such as the Internet, There is a need to have a standard way to identify, retrieve and manage these resources. These include items such as electronic data file with different format such as paper, book or chapter of a book, charts, tables, etc. These items are known on the web as digital object. The term digital object refers to the data structured as a set of bits is calculated by a computer. Each set of bits includes at least one pair of values that can be used for digital object (USGS, 2014). In total it is assumed that the accumulation in one or several locations helps to an object to be accessible.

One of the basic feature of a digital object is its stored location that results object retrieval. In this regard, there is a need to establish a standard system to identify digital objects as content management infrastructure and related data and metadata has been associated with the links (DOI, 2016). In general, the digital environment requires a unique identifier for the management of resources. A Document may carry multiple identifiers. Consequently, there is no guarantee for assigning a unique code to the same object by the

various data provider services. Accordingly, generating standard code for an electronic object is essential for its uniqueness. Digital Object Identifier (DOI) as a global content identifier meets the needs of publishers and authors (Tiliute, 2016) as producers and consumers of research results. Identifiers that can be attributed to an electronic work are unchangeable and are the indication of intellectual property. Wang (2007, 164) believes that "the persistent, unique, and reliable identifier, along with the resolution system, provides a perfect solution to manage the intellectual content in a digital environment). An identifier for digital objects can be labeled on the electronic document (journal articles, books) or different parts of a document (tables, figures, book chapter, and a paragraph). It is noteworthy that higher education (HE) institutions are the main stakeholders in conducting projects and producing research results. Their findings are available both in print or electronic format for fee or free via World Wide Web. In this regard one of the most important steps for HE institutions that are active in digital environment to transmit content, is to standardize the construction of Internet-related technology to achieve the objectives.

There exists many digital objects with multiple identifiers in the digital age. But this is not true for the Persian contents. One may express that Persian journals can use DOI as an international identifier. In reply to this question the imposed sanctions on Iran - even scientific ones, are testimony to the many years of attempts completed by RICeST. Therefore the Persian digital content in the virtual environments must be managed. To meet this objective, we need an infrastructure to design and implement for the Persian electronic objects. To this end a comprehensive study on Digital Object Identifier (DOI) system was conducted. As a result the DOR system, that is a localized system for Persian digital objects, is designed and implemented. The main advantages of DOR are to identify journals contents and "provide a persistent link to its location on the Internet" (APA Style, 2016) as well as linking function to the contents are referencing in the Persian journals. It is noteworthy that RICeST is accountable for acquiring, organizing, and disseminating more than 1600 Iranian peer-reviewed journals. In addition to, RICeST publishes more than 100 journals out of 1600 and it is an opportunity to assign DOR to the articles publish by these journals. Making digital contents easy to find, link, cite, and assess is an outstanding performance of DOR.

The Structure of DOR

In its initial phase, DOR system is designed to assign DOR number to Persian journal articles. The technological infrastructure is quite promising and enable us entering a new phase in other digital document formats. DOR indicator is like a digital fingerprint and for each article define a unique numerical identifier. It makes the opportunity to identify and retrieve a paper, without any concern over its location.

Using DOR is beneficial for journals the time they take to publish articles. Publishers request to assign a digital alpha-numeric identifier (DOR) to any article of new issue of journal that causes the paper simply retrieve electronically. One of the advantages of using DOR is that the digital version of the articles, even if the journal's URL is changed or halted can be accessible. DOR assignment to journal articles is likely to accelerate content visibility. DOR identifier is constant values as such if a publisher convert its publication from one system to another or transmit the content from one publisher to another publisher, all links to

the documents that previously created still remain active. So, one of the main feature of DOR is its persistence and practicality. By clicking on DOR that is assigned to a digital object, user direct to the desired document. With these features, effective management and accurate tracking of scientific communication is possible. DOR detect objects reliably and make them available. Using DOR, there is accurate and quick information retrieval become possible on the web, and also it protecting intellectual property.

Currently RICeST acts as Registration Agency and DOR Foundation. Regarding its structure, the numbering system that is considered by DOR is as follows: The DOR syntax shall be made up of a DOR prefix and a DOR suffix separated by a forward slash. The directory indicator is "98" that is country code for IRAN. Registrant code “1000” is considered for the Regional Information Center for Science and Technology as the central registration agency. Therefore, each DOR begins with the prefix 98.1000. In order to have more security, the assigned suffix is encrypted. For each article a unique DOR will create by designed encryption algorithm, before the announcement of this unique code to the user, the uniqueness of the code in the system is confirmed. The following items are considered to assign a suffix for each DOR: 1. ISSN or eISSN, 2. publication year, 3.journal vol. and issue, 4. the start page number of each article. After that the suffix will be encoded. Accordingly, Encoding is based on UTF-8.

Before journal publication, a publisher send his request to assign DOR to each article. Publisher enter the exact article information that is needed for DOR assigning. It should be noted that after receiving the DOR for an article it is impossible to change it. DOR system is accessible through dor.ricest.ac.ir. In this system, by default, the simple search page will be available to web users as shown in figure 1. There are two search methods: simple and advanced searches. It is noteworthy to mention again that RICeST is in charge of collecting, storing and disseminating Iranian journal contents. In this regard, DOR access feature provide the contents to the users. At simple search, user enter the DOR number to find article with the assigned DOR. In the case of DOR validity, article information is displayed to the user such as Article title, Author(s), Journal title, Publication year, Vol., Issue, and DOR (as shown in figure 2). It is possible to get more information about article such as: Abstract, Keywords, location of Full text as shown in figure 3.If the publisher entered the location of full text, it will be accessible via click on the URL location by user.

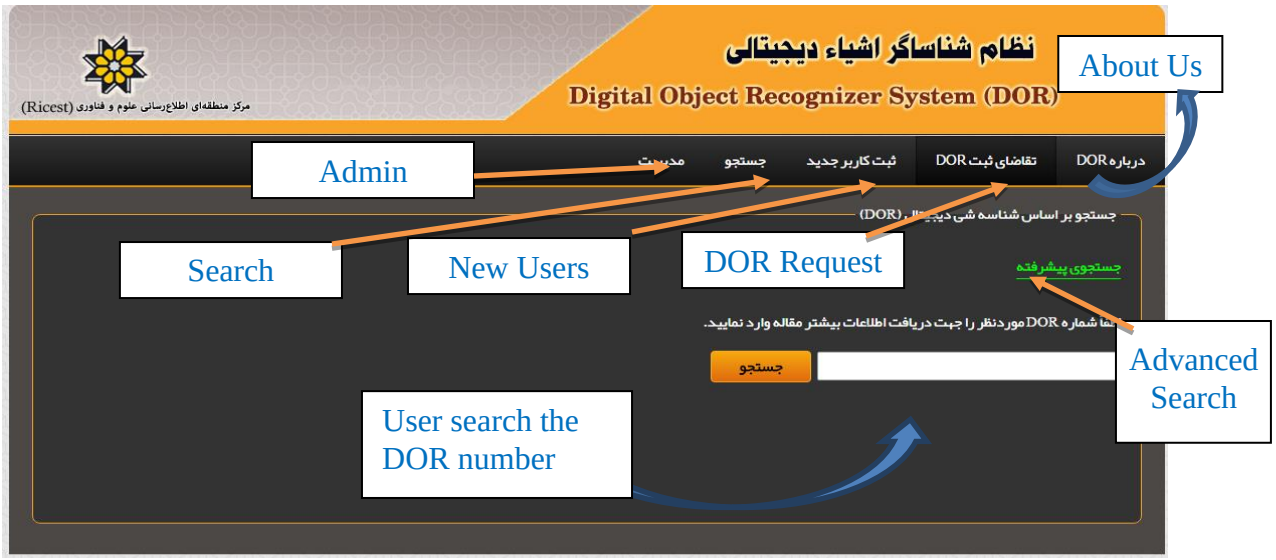

Figure 1. simple search page

In advanced search, user can retrieve the article(s) using tips such as DOR number, keyword, Author, publication year (figure 4). To summarize the following services are provided by DOR system:

1. Choosing an author, user can access to other published articles from the same author.
2. Choosing "article title" or "more information", user can access to further information such as location of files, full-text articles, abstracts and keywords
3. Selecting a keyword, related documents will be retrieved
4. Selecting the URL, provide access to full text
5. Selecting the journal title in result page, link to journal page at RICeST's "IRAN Journal" is accessible and user can get journal's table of content and article's full text.

Figure 2. Sample search result

Figure 3. Additional information for an article

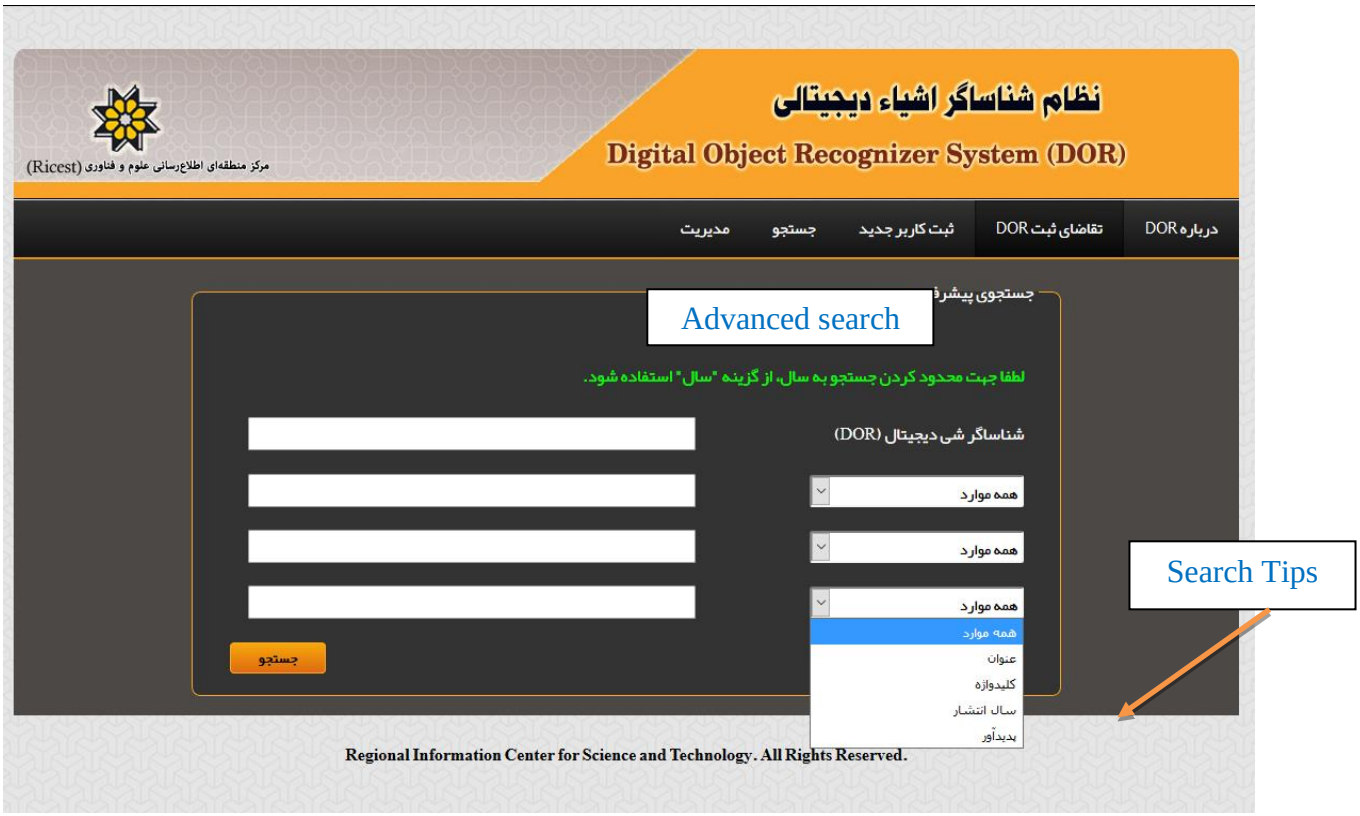

Figure 4. Advanced Search

DOR Architecture

The DOR system architecture is shown in figure 5. Accordingly, DOR system consists of four major modules. The client agent can request for a DOR code. The DOR system can accept request from web search or parametric agent from any program. The client agent requests are processed in resolution system. In this module the user requests are decoded. The object reference and its definitions are stored separately in a data repository system. This sub system has been developed to store a brief story about the object. The file of object can be accessed by the URL that is stored in the object definition.

DOR system Search engine is a complete search engine that can search the repository and DOR information by author, title, keywords and etc. The search engine can be used in web domain or can be called as a service. This ability can be used for many hosts that must cover the DOR capabilities.

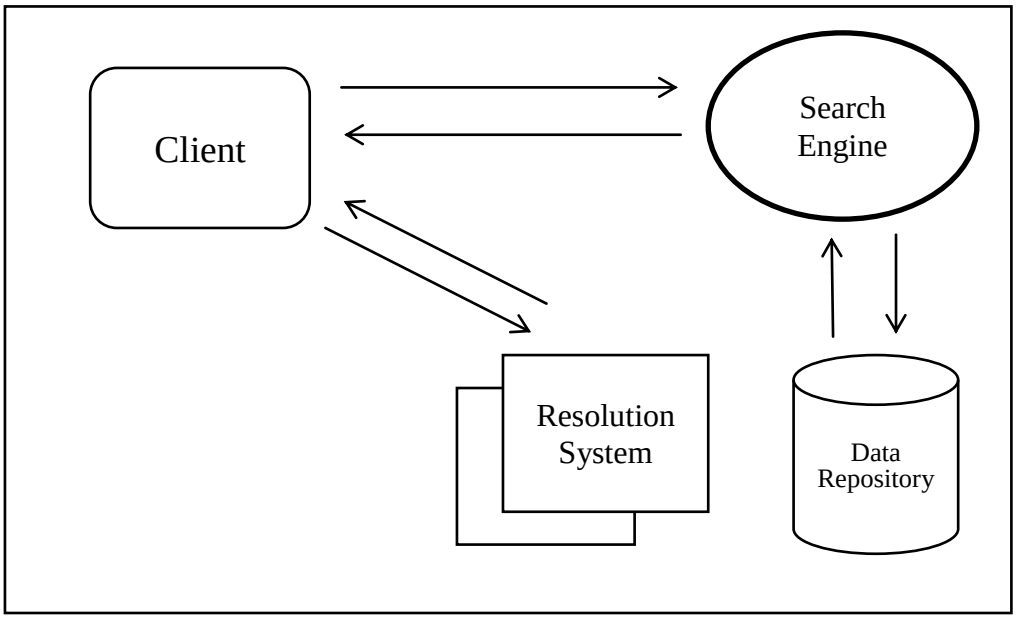

Figure 5. The DOR system architecture

The metadata catalogues and guides for DOR objects are stored in the data repository. The data repository cannot be accessed directly by a DOR number, but the resolution system and the search engine can fetch the information for a user request. The base of system is tuned to have a qualified response to a DOR request. However it is possible to fetch the information by other search strategies like title search. The no free lunch rule is applicable for DOR system; the layered system has some steps to resolute the DOR code. The modularity of the system can improve the visibility and simplicity for system maintenance. The life of systems such as DOR or DOI is more than an application. So the architecture of the system must be tuned for a long life cycle. To this end, system layers developed individually. For example the data layer is separated and can be changed and maintained with a low cost.

The DOR system has a name service that is placed in the resolution sub system. When a new object defined and gets its DOR, the naming service reserves the name of the object. The naming service can resolve the saved names quickly. Unlike DOI system with many agencies around the world, the DOR system is designed to be centralized that supports all hosts and agencies. This service has two tasks. The first one activated when a user request for resolving a DOR number comes from a client module. In this step the resolution sub system is called by a service or by search engine. The results of searching the DOR number can be showed in desired format to the user or host. The second one is the request for a new DOR for an object. In this step, the object definition is entered by the client host. After data completion the resolution system saves the object name(s) and information in the data repository. A new DOR number is generated for each object and this unique number has been returned to the user.

Due to all of journal issues and articles that have been sent to the RICeST by XML, it is possible that automatically a DOR number by "XML system request" as a client to be assigned to all of the objects.

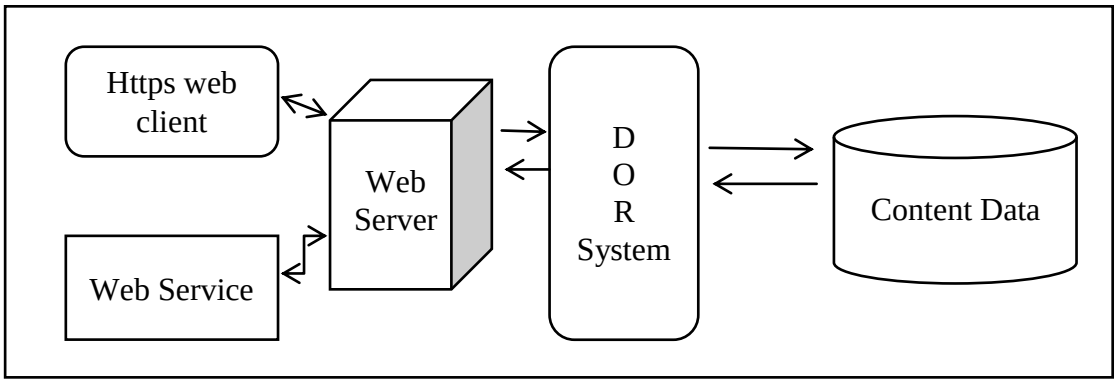

Figure 6. The DOR system and Web Engine

Figure 6 shows the architecture of DOR system and web engine. As it is shown in figure 6, web requests are classified in two types. The first one is https web requests. This type is fed to the web server from web client page. Users that use this web page URL can send their request to the server in this way. The second way is web service. In this way a computer program or a robot can communicate to the server and send a group of requests.

Discussion and Conclusion

The new era brought about electronic publishing that is in favor of scientific community. The crucial issues of the new opportunity are resources management and their accessibility (Chandrakar, 2006). Regarding these concerns, for the first time in IRAN, Digital Object Recognizer (DOR) designed and implemented by an expert team at RICeST. The DOR system define a unique number for each object in the digital environment, therefore scientific products monitoring will be done very quickly and accurately and enable researchers to retrieve the exact content at any time without any wariness about dead links. DOR determine intellectual property in the digital environment and provide intellectual property management of published works in all formats (ibid). The remarkable thing is that if the owner of a journal (publisher) change or the content move from one host to another there is no change in the DOR number and object (s) are identified with the same DOR number that is assigned. The advantages of using DOR are:

1. Saving time and cost: using standards to enable data sharing. This facilitates scientific collaboration and increase content value through the sharing. It makes easy monitoring and identify the works related to the subject of your research and results in fast and accurate works location

2. Intellectual property protection: sustainable DOR is able to determine the creators' copyright and in this regard copyrighted works are always traceable, searchable, and retrievable.

3. DOR number is an effective tool in referencing academic works and can be considered for communication and evaluation systems such as social networks and Islamic World Science Citation Center (ISC), respectively.

4. Regarding above benefit, DOR makes an opportunity to receive more citations and

provide links between scientific works in the digital environment. Therefore, Using DOR is one of the best way to increase the ranking of a digital object and enhance citations and links to research findings and support reference accuracy, as such increases visibility of published works

5. Since Persian published articles support English abstracts, non-Farsi speakers will be aware of works in their respective areas that assigned DOR

Acknowledgement

We would like to express our gratitude to RICeST's Scientific committee for approval this research project. We also thank Naghmeh Sohrabian and Farzad Rasti as members of programming team.

References

APA Style (2016). Electronic Sources and Locator Information. Retrieved from http://www.apastyle.org/manual/related/electronic-sources.pdf

Chandrakar, R. (2006). Digital object identifier system: An overview. *The Electronic Library*, 24(4), 445-452.

Digital Object Identifier, DOI (2016). Retrieved from www.doi.org (June 15, 2016).

Tiliute, D. E. (2016). Digital object identifier for a better image of scholar journals. *USV Annals of Economics & Public Administration*, 16.

USGS (2014). Digital Object Identifiers (DOI) Usage and Adoption in U. S. Geological Survey. Retrieved from https://www2.usgs.gov/datamanagement /files/DOI-DCC.pdf

Wang, J. (2007). Digital object identifiers and their use in libraries. *Serials Review*, 33(3), 161-164.