1

SCIENTIFIC JOURNAL ARTICLES METADATA HARVESTER

Leon Andretti Abdillah [1)]

[1)] Sistem Informasi, Fakultas Ilmu Komputer Universitas Bina Darma
Jl. Ahmad Yani No. 12, Palembang 30264 Indonesia
email : leon.abdillah@yahoo.com

ABSTRACT

Scientific journals are very important in recording the finding from researchers around the world. The recent media to disseminate scientific jurnals is PDF. On scheme to find the scientific journals over the internet is via metadata. Metadata stores information about article's summary. Embedding metadata into PDF of scientific article will grant the consistency of metadata readness. Harvesting the metadata from scientific journal is very interesting field at the moment. This paper will discuss about scientific journal articles metadata harvesters involving XMP. The harvester will extract information about the summarization of the article(s) based on common used metadata fields and author enriches them with some usefull fields.

Key words
Scientific journal article, metadata, harvester, PDF, XMP

1. Introduction

Every day, each publisher and/or author(s) compete to publish their new papers, scientific literatures, through the cloud. Scientific literatures published in both manuscript format and available electronically. The basic task is to make these documents searchable and retrievable [1]. In the field of digital library, scientific workers always search a lot of scientific documents at the domain of their researches [2]. They will work for new idea, invention, etc based on the previous research or to deal with the current and future challenges.

The electronic representation of scientific documents may include journals, technical reports, program documentation, laboratory notebooks etc [3]. The most popular of scientific documents is scientific journals article of a particular field or topic.

An article from scientific journal, commonly dominated by word(s) or text(s), but to clarify the discussion then it could be added by several forms such as black-and-white line(s), chart(s), diagram(s), equation(s), formula(s), graphic(s), illustration(s), photograph(s), picture(s), table(s), etc. Scientific journals have a formal structure that has to be understood by all those who read it (authors, readers and editors) in order to be useful [4], especially for editors who will interact directly to the article. Editor will check whether the submitted article meet the requirements and editorial policy of the journal [5], these activities will keep quality of the journal in supplying new knowledge to the world science. Scholarly or scientific papers usually have certain pieces of metadata (usually assigned by authors) describing the topics and the main ideas of the contents [6].

The term *metadata* has been increasingly adopted and co-opted by more diverse audiences, the definition of what constitutes metadata has grown in scope to include almost anything that describes anything else [7]. Metadata are literally or technically 'data about data' or information about information or information that makes data useful. More over metadata as data whose primary purpose is to describe, define and/or annotate other data that accompanies it [8]. The structured data of Metadata describes the characteristics of a resource. It shares many similar characteristics to the cataloguing that takes place in libraries, museums and archives. The term "meta" derives from the Greek word denoting a nature of a higher order or more fundamental kind. A metadata record consists of a number of pre-defined elements representing specific attributes of a resource, and each element can have one or more values [9]. It is an extensive and expanding subject that is prevalent in many environments [10]. They provide information on such aspects as the 'who, what, where and when' of data and can be considered from the perspective of both the data producer and the data consumer. For the producer, metadata are used to document data in order to inform prospective users of their characteristics, while for the consumer, metadata are used to both discover data and assess their appropriateness for particular needs – their so-called 'fitness for purpose'. Providing metadata is the responsibility of each data provider with the quality of the metadata a significant problem [11]. Figure 1 shows some of metadata schemes.

In term of search, metadata is very useful key for search engine to recognize as the guide about what information should be provided to the users and it also determines the level of success of a search.The most