

Documents Component Ontology (DoCO): Laguna University Accreditation Tool

Marlon L. Atanacio ¹, Chrisna L. Fucio ², Alejandro V. Matute Jr. ³

Abstract: The document layers' representation, arrangement, and discourse in machine-readable survey forms are pivotal for both users and machines in organizing and understanding the documents. In this endeavor, the researchers instituted DoCO, the Document Components Ontology for the Accreditation of Laguna University. This is the classification that gives an all-purpose organized file elements' vocabulary to chronicle units of a document. As a result of the overall assessment of the respondents and clients in terms of Functional Suitability, Performance Efficiency, Compatibility, Usability, Reliability, Security, Maintainability, and Portability which got 4.12 as the overall weighted mean, which is verbally interpreted as "Good." The respondents perceived the developed project as effective because it could give quick, valid, and fixed information. This can be utilized for the accreditation purposes of the institution.

Keywords: DoCO, Document Components, Ontology, document elements

1. Introduction

In today's generation, technology is everywhere. Having a data management system that is used to collect, store, and retrieve the documents of an organization or an institution is in demand. According to the International Organization for Standardization (ISO) [1], an operating system helps the institution manage the interrelated units of its line to realize its purposes. These objectives include byproduct or assistance quality, management system, environmental service, health and safety in the area, and many others. In terms of document filing, storing, and archiving, all of the transactions that generate data are eventually stored in the database, which is efficient as well as convenient to use. Data archiving [2] is the practice of identifying inactive data and moving it from production systems to long-term storage systems.

Stedman and Vaughan [3] discussed that data management is the ingesting, storing, organizing, and maintaining of the information gathered by an institution, and this aims to make sure that the data

¹ Department of Information Technology, Laguna University, Sta. Cruz Laguna, Philippines
Email: lon.atanacio@gmail.com

² Department of Information Technology, Laguna University, Sta. Cruz Laguna, Philippines
Email: clfucio20@gmail.com

³ Department of Information Technology, Laguna University, Sta. Cruz Laguna, Philippines
Email: alejandroventiromatute@gmail.com

Received [June 17, 2022]; Revised [November 20, 2022]; Accepted [December 2, 2022]



incorporate systems that are accurate, available, and accessible. Scanning documents has been progressively used by businesses and institutions to reduce costs, increase productivity, and improve access to information. Paper and other media can be converted into digital images.

Laguna University (LU) is one of the fastest-growing and best-known public universities in Santa Cruz, Laguna and started its operation in 2006. The institution currently uses manual methodology in handling and storing documents for accreditation. Laguna University underwent certification from the Association of Local Colleges and Universities Commission on Accreditation (ALCU-COA) [4]. This is the certification unit of educational entities in the local administration with the Securities and Exchange Commission (SEC) registry on December 9, 2003. ALCU-COA is a pioneer member of the National Network of Quality Assurance Agencies (NNQAA) with SEC registry on January 8, 2004, as well as the Accrediting Agency of Chartered Colleges and Universities in the Philippines (AACUP), the certification and recognition unit of state educational institutions, and a member of the Coordinating Council for Accreditation of accrediting units for government and nongovernment Higher Education Institutions (HEIs) established on January 19, 2009. Based on the website of ALCU-COA [4], “accreditation” was started in 1957 through the Philippine Association of Accredited Schools and Universities (PAASCU). It was purposely established for nongovernment academic entities. The creation of ALCU began in Local Universities and Colleges (LUCs) to meet society’s academic and demographic demands. As time went by, the Association of Local Colleges and Universities was established, and Dr. Benjamin G. Tayabas served as its first president [4].

Laguna University (LU) encountered problems in terms of managing documents; there were lots of papers to keep, consolidate, record, and also organize. The teaching and non-teaching personnel were consuming a lot of time on this task. Since they are doing the manual task, the documents of LU have data redundancy, leading to a great possibility of causing further complexity in the accreditation process. In line with these, the researchers have developed and designed a Laguna University Management System for Accreditation that can organize documents and store them in databases for future use, as they can be easily stored, uploaded, downloaded, retrieved, archived, and printed when needed. It helps the LU attain improvements in their documents for accreditation, which is inconvenient, to become more efficient, not only for the faculty and staff but especially for the accreditors.

The rest of this study is organized as follows: in Section 2, the significant studies about representations illustrating the document components are explained; in Section 3, a review of DoCO is presented, outlining the bases and the descriptions of the arrangement of documents concerning both class designs and possessing language patterns; in Section 4, the utilization of DoCO is emphasized concerning the explication and retrieval of the components; and lastly, further improvements are explained for the utilization hereafter in Section 5.

2. Literature Review

Previously, a few studies have introduced versions such as Resource Description Framework Schema (RDFS) vocabularies and Web Ontology Language (OWL) ontologies to depict specific areas and bibliographic assets’ metadata, like Dublin Core Terms (DCTerms), Publishing Requirements for Industry Standard Metadata (PRISM), and Bibliographic Ontology (BIBO). The illustration of Semantic Publishing and Referencing (SPAR) ontologies is one of the purposes to represent the entirety of the distributing area.

The first form of SPAR ontologies has seven (7) different designs, which are as follows:

1. The FRBR-aligned Bibliographic Ontology (FaBiO) [5] describes entities that are released for publication or feasibly publishable (*e.g.*, journal publications, conference studies, books) and have a list of sources;
2. The Citation Typing Ontology (CiTO) [5] provides a representation of the factual and rhetorical kinds of citations;
3. The Bibliographic Reference Ontology (BiRO) [6] distinguishes bibliographic lists, sources, and their collections;
4. The Citation Counting and Context Characterization Ontology (C4O) [6] allows to save a specific number of in-text citations internationally in a specific time;
5. The Publishing Roles Ontology (PRO) [7] characterizes the functions of entities in publication. They are writers, revisers, validators, presses, or catalogers;
6. The Publishing Status Ontology (PSO) [7] determines the condition of the documents in a publication, such as outline, presented, for revision, and many others;
7. The Publishing Workflow Ontology (PWO) [8] describes the phase in the sequence connected to the publication of a document.

The aforementioned ontologies and DoCO should make the first form of SPAR ontologies. In addition, the form has been newly expanded with four (4) additional ontologies. They are the following:

1. Scholarly Contributions and Roles Ontology (SCoRO) [9] describes the benefactions and functions of an individual regarding publications (like the function of article illustrator and/or guarantor);
2. Funding Research Administration and Projects Ontology (FRAPO) [10] deals with the managerial data of the studies, examples, allocations, solicitations, partnerships, and many others;
3. DataCite Ontology [11] allows the metadata qualities of the DataCite Metadata Schema Specification [12] (*i.e.*, a list of metadata qualities for the precise recognition of cited and retrieved resources), which are determined in RDF;
4. Bibliometric Data Ontology (BiDO) [13] illustrates the numerical and categorical bibliometric information (*i.e.*, journal impact factor, author h-index, and categories describing research careers) in Resource Description Framework (RDF).

In identifying particular parts as absolutely logical or simply auxiliary, the text is not categorized as being meticulously primary (that is, taking merely linguistic capability), because that brings language too, throughout the same body of words.

The record markup dialects distinguish a citation as a pure fundamental piece without citation to its illustrative capability:

- “A *paragraph* is typically a run of phrasing content that forms a block of text with one or more sentences” [14];
- “*Paragraphs in DocBook may contain almost all text in lines and most block elements*” [15].

Those aforementioned descriptions highlight those passages’ suggested meanings as experienced by the readers. It helps them categorize a passage from a text.

To combine the document elements' completely structural qualifications and also their completely rhetorical suggested meaning, the DoCO was developed. The formation is explained as follows: The DoCO emergence was done through researching various institution files and the publishers' patterns, which were categorized into two (2) points of view: the class and the language, and assessed via previous studies on file designs [10][11][14]. DoCO introduced the Pattern Ontology that illustrates the class designs [10] and the Discourse Element Ontology that determines those language elements. More so, this explains the blended courses that present structural and rhetorical content. An illustration presents the construction as well as the DoCO types. The succeeding portion explains its structural design postulation as described in [10] and the language elements, which are commonly visible in scholarly journals, that highlight the DoCO theoretical supports. It also presents some of the DoCO document components connected to the scientific journal definitions.

To view how structure can be broken into nuclear parts, there are samples of printed archives. The researchers introduced a few fundamental samples that are sufficient to demonstrate what most clientele demand, categorized by two (2) basic perspectives:

- Orthogonality – Every design requires to have a distinctive cause, placing a particular contexture.
- Specificity – Every design can be utilized in particular areas (*e.g.*, within other designs).

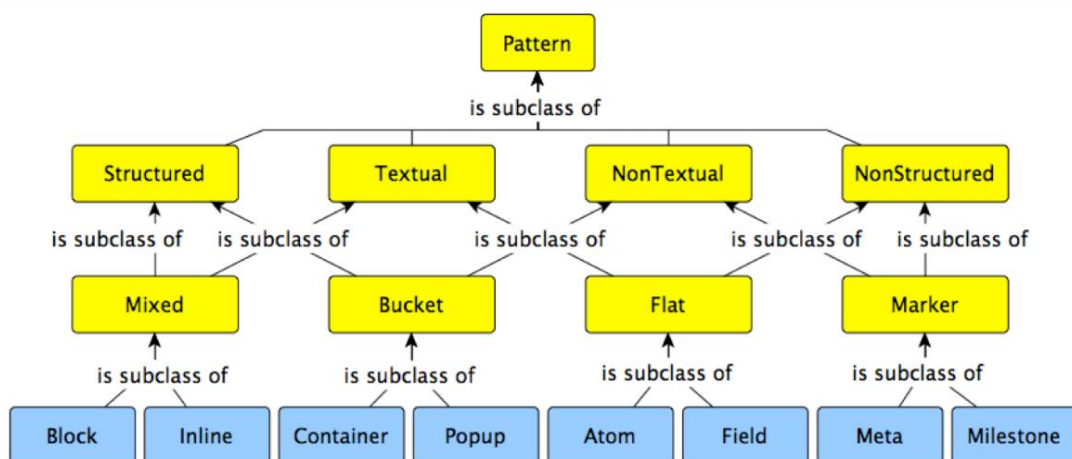


Figure 1. A Graffoo Illustration

The textual file guides were entirely determined in FRAPO [10]. They are the Pattern Ontology, outlined in Figure 1. The designs are explained and characterized by two (2) various qualities: Textual or NonTextual and Structured or NonStructured. These fundamental qualities are merged to attain four (4) various disjoint types explaining entities that (A) have text and substructures (po: Mixed), (B) have substructures but do not possess text (po: Bucket), (C) have text but do not have substructures (po: Flat), (D) do not possess text, nor substructures (po: Marker). The type is a superclass of two (2) subclasses that inclusively describe the eight (8) actual guides characterized by the passage constructions. A unique instance is the pattern “po: Container”, divided into three (3) particularized subunits.

3. Methodology

2.1 Research Design

The study aims to assess the DoCO structures in upgrading the supervision of files for accreditation at Laguna University. This research employed the descriptive design by utilizing a rubric concerning

the ISO/IEC 25010 that examines the software standard as evaluated by the specialists through experimentation. The product quality model defined in ISO/IEC 25010 comprises eight quality characteristics, namely: (1) Functional Suitability, (2) Performance Efficiency, (3) Compatibility, (4) Usability, (5) Reliability, (6) Security, (7) Maintainability, and (8) Portability [1].

The researchers utilized two (2) methodologies: Developmental and Descriptive. Descriptive research aims to precisely and methodically describe a population, situation, or phenomenon. This methodology is an appropriate choice when the research is aiming to identify characteristics, frequencies, trends, and categories [16]. This was utilized to collect information from the respondents for establishing the design. Developmental research is a particular way of addressing the basic questions of why and how to teach what to whom. It includes a cyclical process of small-scale in-depth development and evaluation, at a content-specific level, of exemplary teaching-learning sequences [17]. The Modified Waterfall Model was applied, for it is following the process of implementation to devise the system. Also, the Planning, Conducting, and Reporting phases are indispensable to this endeavor.

2.2 Research Methods

To determine the acceptability level of a Laguna University Documents Component Ontology, an evaluation questionnaire was administered. A software evaluation is a type of assessment that seeks to determine if the software or a combination of software programs is the best possible fit for the needs of a given client [18]. The software evaluation based on ISO/IEC 25010:2011 was used, which served as the basis to evaluate the functionality and quality characteristics of the developed system to determine if it meets the desired objectives and requirements. For this evaluation, the total respondents comprised 57 teaching personnel, non-teaching personnel, the head of the Quality Management Office (QMO) at Laguna University, and an IT Expert from the industry.

Information Technology (IT) specialists from Laguna University - Department of Information Technology conducted the assessment utilizing the 5-point Likert scale, where five (5) is the maximum and one (1) is the minimum as shown in Table 1.

Table 1. Five Likert Scale

| Scope | Scale | Analysis |
|-----------|-------|----------------|
| 4.51-5.0 | 5 | Outstanding |
| 3.51-4.50 | 4 | Good |
| 2.51-3.50 | 3 | Satisfactory |
| 1.51-2.50 | 2 | Poor |
| 0-1.50 | 1 | Unsatisfactory |

2.3 Conceptual Framework

In Document Retrieval, depicted in Figure 2, users forward a concern to the structure to recover significant files. It aims to regulate the vocabulary utilized in the file to prevent language confusion. The green author uses the term “*river*” in the green file. The red author utilizes the term “*watercourse*” in his file to reference the same thought. This document component ontology connected the words “*river*” and “*watercourse*” to the same concept by utilizing a synonym link. This concept is contained in the green

and red document indexes. Indexes have a description of the document’s content. Thus, document indexes and user queries utilize the same vocabulary; hence, comparisons can be made through the information retrieval system.

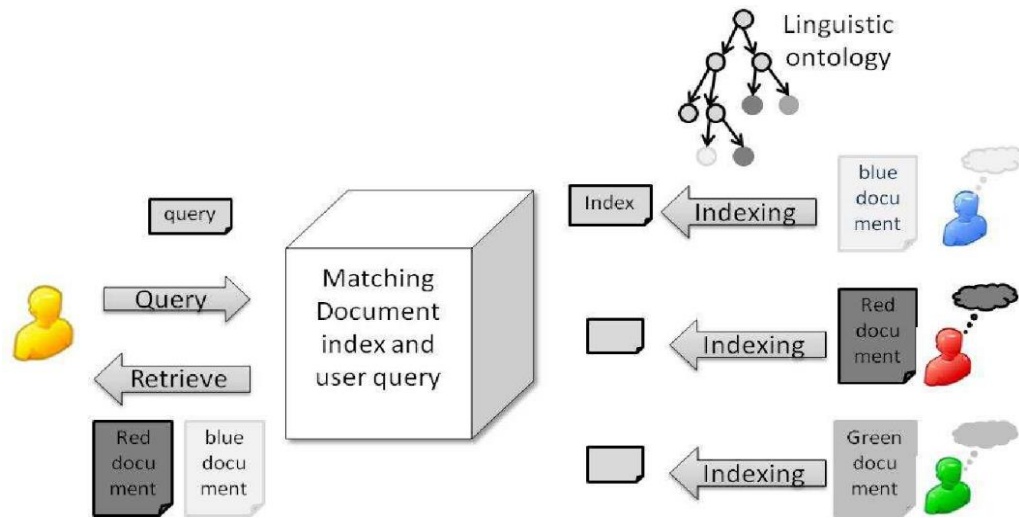


Figure 2. Recovery of Document

4. Results and Discussion

Table 2. Software Evaluation Summary

| Criteria | Weighted Mean | Verbal Interpretation |
|------------------------------|---------------|-----------------------|
| Functional Suitability | 4.63 | Outstanding |
| Performance Efficiency | 3.11 | Satisfactory |
| Compatibility | 4.14 | Good |
| Usability | 4.52 | Outstanding |
| Reliability | 4.04 | Good |
| Security | 4.36 | Good |
| Maintainability | 4.05 | Good |
| Portability | 4.10 | Good |
| Overall Weighted Mean | 4.12 | Good |

The distribution of responses was arranged using tables for a clear presentation of data. The data were analyzed and interpreted accurately and accordingly. Table 2 presents each criterion used in the assessment of the developed study. The weighted mean formula was utilized to identify the descriptive equivalent of the dimension.

Table 2 shows the overall assessment of the respondents' understanding of the Laguna University Documents Component Ontology. The clients assessed the project employing the ISO/IEC 25010 in terms of Functional Suitability, Performance Efficiency, Compatibility, Usability, Reliability, Security, Maintainability, and Portability which got a value of 4.12 as the overall weighted mean, which was verbally interpreted as "Good." The respondents perceived the developed project as effective because it could give quick, valid, and fixed information.

5. Conclusions and Recommendations

On the basis of the foregoing findings, the following conclusions and recommendations were drawn:

1. For future researchers, designing a more efficient ontology algorithm for document classification will be helpful to the development of the system.
2. Implementation of Natural Language Processing algorithms on ontology concepts is highly recommended.
3. Researchers could make a mobile application that would enable the system to run on mobile devices. Being able to run the Laguna University Documents Component Ontology for Accreditation on mobile devices would greatly increase the portability of the system.
4. This developed system is "Outstanding" in all dimensions of the ISO/IEC 25010 design as assessed by the respondents.
5. Application of Matching Document Index and User Query has developed the document indexing needed for accreditation.
6. Laguna University Documents Component Ontology for Accreditation is a helpful system to aid the accreditation of the institution.

References

- [1] ISO 2500, "ISO/IEC 25010", www.iso25000.com/index.php/en/iso-25000-standards/iso-25010 (Accessed January 25, 2021).
- [2] Druva, "Data archiving", www.druva.com/glossary/what-is-data-archiving-definition-and-related-faqs/ (Accessed January 25, 2021).
- [3] C. Stedman and J. Vaughan, "What is data management and why is it important?", www.techtarget.com/searchdatamanagement/definition/data-management (Accessed January 10, 2021).
- [4] ALCUCOA, www.alcucoa.org.ph (Accessed January 10, 2021)
- [5] S. Peroni and D. Shotton, "FaBiO and CiTO: ontologies for describing bibliographic resources and citations", *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 17, December 2012, pp. 33-43, doi: 10.1016/j.websem.2012.08.001.
- [6] A. Di Iorio, A.G. Nuzzolese, S. Peroni, D. Shotton, and F. Vitali, "Describing bibliographic references in RDF", in *Proc. of 4th Workshop on Semantic Publishing, CEUR Workshop Proceedings*, vol. 1155, 2014, www.ceur-ws.org/Vol-1155/paper-05.pdf (Accessed January 23, 2015).
- [7] S. Peroni, D. Shotton, and F. Vitali, "Scholarly publishing and linked data: describing roles, statuses, temporal and contextual extents", in *Proc. of the 8th International Conference on Semantic Systems*, September 2012, pp. 9-16, doi: 10.1145/2362499.2362502.
- [8] A. Gangemi, S. Peroni, D. Shotton, and F. Vitali, "A pattern-based ontology for describing publishing workflows", in *Proc. of the 5th International Workshop on Ontology and Semantic Web Patterns, CEUR*

- Workshop Proceedings, vol. 1302, October 2014, www.ceur-ws.org/Vol-1302/paper1.pdf (Accessed January 23, 2015).
- [9] D. Shotton and S. Peroni, “*SCoRO, the Scholarly Contributions and Roles Ontology*”, www.sparontologies.github.io/scoro/current/scoro.html (Accessed December 23, 2021).
- [10] D. Shotton, “*FRAPO, the Funding, Research Administration and Projects Ontology*”, www.sparontologies.github.io/frapo/current/frapo.html (Accessed December 23, 2021).
- [11] D. Shotton and S. Peroni, “*The DataCite Ontology*”, www.sparontologies.github.io/datacite/current/datacite.html (Accessed December 23, 2021).
- [12] DataCite Schema, “*DataCite Metadata Schema*”, www.schema.datacite.org/ (Accessed December 23, 2021).
- [13] F. Osborne, S. Peroni, and E. Motta, “*Clustering citation distributions for semantic categorization and citation prediction*”, in Proc. of the 4th Workshop on Linked Science (LISC 2014), CEUR Workshop Proceedings, vol. 1282, 2014, pp. 24-35.
- [14] I. Hickson, “*HTML5: a vocabulary and associated APIs for HTML and XHTML*”, W3C Candidate Recommendation 29 April 2014, World Wide Web Consortium, 2011, www.w3.org/TR/html5/ (Accessed January 23, 2015).
- [15] N. Walsh, “*DocBook 5: The Definitive Guide*”, Sebastopol, CA, USA, O’Really Media, 2010, ISBN: 0596805029.
- [16] S. McCombes, “*Descriptive Research: Definition, Types, Methods & Examples*”, www.scribbr.com/methodology/descriptive-research/ (Accessed May 15, 2019)
- [17] K. Klaassen and K. Kortland, “*Developmental Research*”, in Encyclopedia of Science Education, R. Gunstone, Eds., Dordrecht, Netherlands: Springer, 2013, pp. 1-4, doi: 10.1007/978-94-007-6165-0_155-1
- [18] M. Tatum, “*What Is Software Evaluation?*”, www.easytechjunkie.com/what-is-software-evaluation.htm (Accessed December 23, 2021).