

UNDERGRADUATE THESIS

DWIPA ONTOLOGY V : ENRICHMENT AND POPULATION
OF OBJECTS ON MULTI-MODAL ONTOLOGY FOR THE
TOURISM DOMAIN ON DWIPA ONTOLOGY IV



UNIVERSITAS
BAKRIE

MUHAMMAD SULTHAN AFGHANI W.

1172001022

**INFORMATICS STUDY PROGRAM
ENGINEERING AND COMPUTER SCIENCE FACULTY
BAKRIE UNIVERSITY
JAKARTA
2022**

UNDERGRADUATE THESIS

DWIPA ONTOLOGY V : ENRICHMENT AND POPULATION
OF OBJECTS ON MULTI-MODAL ONTOLOGY FOR THE
TOURISM DOMAIN ON DWIPA ONTOLOGY IV

Submitted as partial fulfillment of the requirements for the degree of
Bachelor of Computer in Informatics Study Program



MUHAMMAD SULTHAN AFGHANI W.

1172001022

INFORMATICS STUDY PROGRAM
ENGINEERING AND COMPUTER SCIENCE FACULTY
BAKRIE UNIVERSITY

JAKARTA

2022

STATEMENT OF ORIGINALITY

The material in this Undergraduate Thesis is the result of my own work,
and all sources are quoted and cited properly.

Name : Muhammad Sulthan Afghani W.

NIM : 1172001022

Signature :



Date : August 8th, 2022

STATEMENT OF APPROVAL

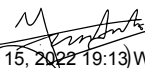
This undergraduate thesis is prepared and submitted by:

Name : Muhammad Sulthan Afghani W.
NIM : 1172001022
Study Program : Informatics
Faculty : Engineering and Computer Science
Title : Enrichment and Population of Objects on Multi-Modal Ontology
for the Tourism Domain on DWIPA Ontology IV

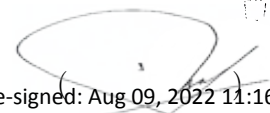
has been approved by the Board of Examiners and accepted as a partial fulfillment of the requirements to obtain a Bachelor of Computer in Informatics Study Program, Faculty of Engineering and Computer Science, Universitas Bakrie.

BOARD OF EXAMINERS

Supervisor : Guson Prasamuarso Kuntarto, S.T., M.Sc.


(Aug 15, 2022 19:13) WIB

Examiner I : Berkah Iman Santoso, S.T, M.T.I


e-signed: Aug 09, 2022 11:16 WIB

Examiner II : Iwan Adhicandra, S.T., MSc, MIEEEE, MACM

()

Authorized in : Jakarta

Date : August 8th, 2022

Acknowledgements

Praise to Allah SWT for His blessing and mercy so that the thesis report can be completed. This thesis report titled *DWIPA ONTOLOGY V : ENRICHMENT AND POPULATION OF OBJECTS ON MULTI-MODAL ONTOLOGY FOR THE TOURISM DOMAIN ON DWIPA ONTOLOGY IV* is submitted to fulfill one of the requirements in accomplishing the Bachelor Degree of Computer at the Informatics Study Program, Universitas Bakrie.

In the preparation of this report is not only result from unilateral work, but also need help from many people. Without the help from some of contributing people, this report cannot be finished well and on time. Many thanks and appreciation is expressed to:

1. Prof. Ir. Sofia W. Alisjahbana, Ph.D as the Rector of Bakrie University.
2. Mr. Esa Haruman, Ph.D as the Dean of FACULTY OF ENGINEERING AND COMPUTER SCIENCE.
3. Mr. Iwan Adhicandra, S.T., MSc, MIEEEE, MACM as the Head of Study of INFORMATICS STUDY PROGRAM.
4. Mr. Guson P. Kuntarto, S.T., M.Sc., as both my thesis and academic supervisor.
5. Mr. Berkah Iman Santoso, S.T, M.T.I, as discussant in my proposal seminar who has given suggestions for the improvement of research content.
6. My mother and father, and everyone else from my family, who have been exceedingly patient and supportive throughout the writing process.
7. My friends over at discord, with whom i went through countless sleepless nights with just rambling about life, wouldn't have stayed sane without you guys.
8. Fadil, my friend and a data scientist who played a huge part in guiding me through the data collection and image classification process in this research.
9. My DnD crew, weekend at home has never been more fun than with you guys.

Jakarta, August 8th, 2022



Muhammad Sultnan Arghani W.

DECLARATION OF PARTIAL COPYRIGHT LICENSE

As student of Universitas Bakrie, I hereby:

Nama : Muhammad Sulthan Afghani W.
NIM : 1172001022
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer
Judul : Enrichment and Population of Objects on Multi-Modal Ontology
for the Tourism Domain on DWIPA Ontology IV

The author, whose copyright is declared on the title page of this word, agree and grant Bakrie University a **Non-exclusive Royalty-free Right** for educational and academic endeavour with this Undergraduate Thesis title:

**Enrichment and Population of Objects on Multi-Modal Ontology for the
Tourism Domain on DWIPA Ontology IV**

With the granted permission to use this material, Bakrie University is allowed to keep or make digital copy, communicate, and publish this Undergraduate Thesis by providing full acknowledgement of the copyright and the source of the material. It is understood that copying and publicizing of this word for financial gain shall not be allowed without written permission. The original Partial Copyright License attesting to these terms, and signed by this author, may be found is the original bound copy of this work, retained in The Bakrie University Archive.

Authorized in : Jakarta

Date : August 8th, 2022

Jakarta, August 8th, 2022



Muhammad Sulthan Afghani Wijayanto

Abstrak

Ontologi DWIPA IV adalah ontologi penyedia informasi domain pariwisata nasional. Informasi yang disimpan mencakup berbagai akomodasi dan amenities yang terkait tempat turisme. Namun, ontologi yang ada hanya menyimpan informasi berbasis teks. Untuk memperluas parameter pencarian, direncanakan untuk memperkaya dan mempopulasi ontologi yang sudah ada dengan data multi-modal. Data yang akan dipopulasi direncanakan akan berbentuk gambar yang dianotasi. Pertama, riset ini bertujuan menciptakan basis data untuk gambar turisme yang berbasis di Indonesia. Kedua, ontologi akan diperkaya untuk mengakomodasi data multi-modal. Ketiga, gambar yang dikumpulkan akan dianotasi dan dipopulasi ke ontologi yang sudah diperkaya, sekarang diberi nama DWIPA V. Terakhir, hasil dari proses enrichment and population akan dievaluasi menggunakan OntoQA schema metrics. Diharapkan penelitian ini dapat menghasilkan ontologi berbasis multi-modal dengan domain pariwisata nasional yang dapat dijadikan basis data dari DWIPA semantic web dan mobile application. Hasil dari penelitian ini adalah ontologi yang sudah dipopulasi menjadi 3279 *instances*, 214 gambar, dan 2 class baru : *Islands* dan *Archipelago*. Hasil dari evaluasi ontologi adalah 819.75 di *Attribute Richness*, 79.98 di *Average Population*, 9,25 di *Inheritance Richness*, dan 0,47 di *Relationship Richness*.

Kata kunci : Ontologi DWIPA; ontology learning; ontology enrichment; ontology population; image annotation; multi-modal data; ontology evaluation; ontoQA; schema metrics.

Abstract

The DWIPA IV Ontology provides information about the national tourism in Indonesia. The information stored includes various accommodations and amenities regarding the tourism spot. However, the current ontology only stores text-based information. To broaden the search parameter, it is planned to enrich and populate the existing ontology with multi-modal data. The populated data is planned to be in the form of annotated images. Firstly, this research aims to create a database for tourism images based in Indonesia. Secondly, the ontology will be enriched to accommodate multi-modal data. Thirdly, collected images will be populated to the enriched ontology, now dubbed DWIPA V. Finally, the result of the enrichment and population will be evaluated using the OntoQA schema metrics. This research is expected to produce a multi-modal ontology for the national tourism domain that can serve as a basis for DWIPA semantic web and mobile application. At the end of the research, the ontology is populated to 3279 instances, on top of 214 images. 2 new subclasses have been added, those being the *Islands* and *Archipelago* subclass. The result of the evaluation is that the ontology currently has the scores of 819.75 on Attribute Richness, 79.98 on Average Population, 9.25 on Inheritance Richness, and 0.47 on Relationship Richness.

Keywords : DWIPA ontology; ontology learning; ontology enrichment; ontology population; multi-modal data; ontology evaluation; ontoQA; schema metrics.

Contents

Abstrak	vi
Abstract	vii
List of Figures	xi
List of Tables	xii
1 Research Backgrounds	1
1.1 Backgrounds	1
1.2 Problem Formulation	2
1.3 Research Goals	2
1.4 Scope of Research	2
1.5 Benefits and Contribution of the Research	2
2 Literature Review	4
2.1 E-Tourism	4
2.2 Ontology	4
2.3 Ontology Editor	5
2.4 Ontology Enrichment	5
2.5 Ontology Population	6
2.6 Multi-modal data	7
2.7 Ontology Evaluation	7
2.7.1 Schema Metrics	7
2.7.2 Instance Metrics	8
2.8 Multimodal Integration	9

2.9	Web Scraping	9
2.9.1	Web Scraper Tools	9
2.9.2	Selenium Webdriver	9
2.10	Image Classification	10
2.10.1	Random Forest Classifier	10
2.11	Related Works	11
3	Research Methods	12
3.1	Research Stage	12
3.2	Research Framework	12
3.3	Data Collection	14
3.4	Image Classification	15
3.5	Ontology Enrichment and Population	16
3.6	Research Instrumentation	19
4	Results and Implementation	20
4.1	Motivation	20
4.2	Data Collection	21
4.2.1	Text Data Collection	21
4.2.2	Image Data Collection	23
4.3	Image Classification	24
4.3.1	Data Augmentation	25
4.3.2	Classifier Training	26
4.4	Enrichment	28
4.4.1	Taxonomy Construction	29
4.4.2	Semantic Relation Extraction	30
4.5	Ontology Population	32
4.6	Evaluation	32
4.6.1	Ontology Quality Analysis (OntoQA)	32
4.7	Class and Subclass Table	35
4.8	Discussions	36

5 Conclusion and Future Works	38
5.1 Conclusion	38
5.2 Future Works	38
Bibliography	40
Appendix	44

List of Figures

Ontology Enrichment Process	5
Ontology Population Process	6
Pseudocode for Random Forest	10
Research Stage	12
Research Diagram	13
Result of scraping	22
Test data metrics of dataset with 120 images and 500 estimators	28
Test data metrics table	28
The Islands subclass	29
The Archipelago subclass	29
Semantic Relations between classes	30
A populated entity with its images	32

List of Tables

Contribution plan for research titled Enrichment and Population of Objects on Multi-Modal Ontology for the Tourism Domain on DWIPA Ontology IV	3
Ontology Editor Comparison	5
Examples of the dataset that will be used in this research	18
Class subclass DWIPA	35
Summarized comparison between DWIPA IV and DWIPA V	37
Raja Ampat Accommodation Dataset	46
Bromo-Tengger-Semeru Accommodation Dataset	47
Bukit Lawang Accommodation Dataset	48
Wakatobi Accommodation Dataset	49
Lake Toba Accommodation Dataset	50
Mandalika Accommodation Dataset	51

Listings

4.1	Code for scraping text-based data	21
4.2	Code for finding address and phone number	22
4.3	Thumbnail scraper	23
4.4	Search and Download function	24
4.5	Image augmentation for rotation transformation	25
4.6	Image augmentation for horizontal flip transformation	26
4.7	load_image_files function	26
4.8	Dataset prediction	27